What Strikes the Strings of Your Heart? - Multi-Label Dimensionality Reduction for Music Emotion Analysis via Brain Imaging

Journal:	*Transactions on Autonomous Mental Development*
Manuscript ID:	TAMD-2014-0052
Manuscript Type:	Multimodal Modeling and Analysis Informed by Brain Imaging
Date Submitted by the Author:	11-Nov-2014
Complete List of Authors:	Liu, Yang; The Hong Kong Polytechnic University, Department of Computing Liu, Yan; The Hong Kong Polytechnic University, Department of Computing Wang, Chaoguang; The Hong Kong Polytechnic University, School of Design Wang, Xiaohong; The Hong Kong Polytechnic University, Department of Computing Zhou, Peiyuan; The Hong Kong Polytechnic University, Department of Computing Yu, Gino; The Hong Kong Polytechnic University, School of Design Chan, Keith C.C.; The Hong Kong Polytechnic University, Department of Computing
Keywords:	biologically inspired feature extraction, emotions and development

What Strikes the Strings of Your Heart? – Multi-Label Dimensionality Reduction for Music Emotion Analysis via Brain Imaging

Yang Liu, Yan Liu, Chaoguang Wang, Xiaohong Wang, Peiyuan Zhou, Gino Yu, and Keith C.C. Chan

Abstract—After twenty years extensive study in psychology, some musical factors have been identified that can evoke certain kinds of emotions. However, the underlying mechanism of the relationship between music and emotion remains unanswered. This paper intends to find the genuine correlates of music emotion by exploring a systematic and quantitative framework. The task is formulated as a dimensionality reduction problem, which seeks the complete and compact feature set with intrinsic correlates for the given objectives. Since a song generally elicits more than one emotions, we explore dimensionality reduction techniques for multi-label classification. One challenging problem is that the hard label cannot represent the extent of the emotion and it is also difficult to ask the subjects to quantize their feelings. This work tries utilizing EEG signal to solve this challenge. A learning scheme called EEG-based emotion smoothing (E^2S) and a bilinear multi-emotion similarity preserving embedding (BME-SPE) algorithm are proposed. We validate the effectiveness of the proposed framework on standard dataset CAL-500. Several influential correlates have been identified and the classification via those correlates has achieved good performance. We build a Chinese music dataset according to the identified correlates and find that the music from different culture may share similar emotions.

Index Terms—Bilinear Multi-Emotion Similarity Preserving Embedding, Brain Imaging, ElectroEncephaloGraphy (EEG), EEG-based Emotion Smoothing, Multi-Label Dimensionality Reduction, Music Emotion Analysis

1 Introduction

Music, laxly explained as "organized sound" [78], exists in every culture and plays a prominent role in our everyday lives [85]. For many people, listening to music indispensably accompanies their routine activities such as eating, studying, walking, driving, and so on [35]. A recent study even showed that people now spend more time listening to music than watching TV/movies or reading books [65].

Why is music so prevalent? Besides the purpose of entertainment, the ability of arousing powerful emotions might be a more important reason behind most people's engagement with music [36], [92]. Passionate songs could heat up our hearts while the sad mood would be evoked when listening to sorrowful songs. Just as the Russian writer Leo Tolstoy said: "Music is the shorthand of emotion". Such an amazing ability of music has fascinated not only the general public but also the researchers from different fields throughout the ages [17].

As defined in Drever's psychology dictionary [14], emotion is "a mental state of excitement or perturbation, marked by a strong feeling, and usually an impulse towards a definite form of behavior". In order to understand such a complex concept of emotion and distinguish it from other psychological states such as reflex, motive, and attitude, Ekman presented a classical categorical model of emotions based on the human facial expressions, which divides emotion into six basic classes: anger, happiness, surprise, disgust, sadness, and fear [18]. Some similar categorical models have also been proposed individually [32], [62].

For music emotions, researchers proposed several models specifically. A hierarchical model called Geneva emotional music scale (GEMS-45) is designed by professionals, which includes 40 labels such as moved, sad, soothed, and heroic [92]. Turnbull et al. collected a number of user-generated annotations that describe 500 Western popular music tracks, and generated 18 easily understood emotion labels[1] such as happy, sad, calming, and arousing [75].

Unlike the categorical models that represent the musical emotions using a number of classes, another kind of emotion models, called dimensional models, describe the emotions in a Cartesian space with valence and arousal as two dimensions. Here valence means how positive or negative the affect appraisal is and arousal means how high or low the physiological

- Y. Liu, Y. Liu, X. Wang, P. Zhou, and K.C.C. Chan are with the Department of Computing, The Hong Kong Polytechnic University, Hong Kong SAR, P. R. China. E-mail: {csygliu, csyliu, csxhwang, cspyzhou, cskcchan}@comp.polyu.edu.hk
- C. Wang and G. Yu are with the School of Design, The Hong Kong Polytechnic University, Hong Kong SAR, P. R. China. E-mail: {whikgd, phusikoi}@gmail.com

1. Note that CAL500 dataset actually has 174 labels including genres, instruments, vocal characteristics, emotions, acoustic characteristics, and song usages. Since our objective is to analyze the relationship between music and emotion, we only utilize the labels related to emotions.

reaction is [67], [71]. For instance, happiness is an emotion of positive valence and high arousal, while sadness is an emotion of negative valence and low arousal.

Since emotion is mainly a psychological concept, many studies on music and emotion to date are psychological [13], [35], [36]. Some researchers aimed to validate the existence of relationship between music and emotion. The Greek philosopher Aristotle described the emotional effects of different musical modes. In his Politics book VIII [52], he stated that the Mixolydian mode makes people sad and grave; the relaxed mode enfeebles the mind; while the Phrygian inspires enthusiasm. By combining the Gestalt Theory and theories of Peirce and Dewey, Meyer aimed to validate the existence of emotion in music [57]. Krumhansl performed experiments to support the point of view that music itself has inherent, unchangeable qualities that will incite in a listener a specific emotional response [43]. Blood and Zatorre demonstrated that the music can evoke emotions by activating the “pleasure centers” in human brain [5].

Besides validating of the existence of relationship between music and emotion, some researchers worked on finding out how the music conveys or evokes emotions. Two founders of the Gestalt psychology, von Ehrenfels and Wertheimer, asked how a melody retains its identity when all pitch or duration values changed but relations preserved [45]. Hevner conducted several psychological experiments to study the affective value of four features of music: the major and minor modes, the rising and falling of the melodic line, the firm or flowing motion in the rhythm, and the simplicity or complexity of the harmony [29], [30]. Brickman et al. suggested the ability to manipulate specific aspects of music to influence musical preference and emotional response [6]. Cooke pointed out that all composers whose music has a tonal basis have used the same, or closely similar, melodic phrases, harmonies, and rhythms to express and evoke the same emotions [10]. He also proposed the basic expressive functions of all twelve notes of scale. Luck stated that music often elicits emotion through emotional associations to specific chord progressions [56].

With the rapid development of computer resources, many tasks in musical emotion analysis, such as music emotional content annotation [74], music recommendation [84], emotion-based music retrieval [82], music emotion detection [46], music emotion recognition [89], and emotion-based music generation [79], have been explored from the perspective of computational modeling. Most of the computational methods for musical emotion analysis are based on the categorical models and dimensional models proposed in psychology [37], [38], [54], [77]. In order to learn the relationship between the feature space and the categorical or dimensional emotion space, many popular machine learning approaches have already been employed to train the model, such as k-nearest neighbor [90], support vector machines [66], Gaussian mixture models [50], neural networks [20], and boosting [55].

Recent advances in brain imaging techniques, such as electroencephalography (EEG) [22] and functional magnetic resonance imaging (fMRI) [3], have enabled the researchers to explore the human brain activities during music listening [19], [40], [48]. By recording and analyzing the ongoing brain responses, some brain-guided computational models have been developed for music emotion analysis [15], [48]. These methods utilized the brain signals to help to link the music and evoked emotions, and thus narrowed the gaps between the low-level music features and the high-level emotion states.

Although tremendous strides forward have already been made in music emotion analysis from different perspectives, what intrinsic element of music and how it arouses a specific emotion response in the listener is still far from well-understood [88]. In order to provide a systematical and quantitative way to analyze the relationship between the music and evoked emotions, we design a computational framework based on brain imaging in this paper. After the ongoing brain activities of subjects during music listening are recorded via EEG, a learning scheme called EEG-based emotion smoothing (E^2S) is proposed to refine the user-provided emotion labels. Then a multi-label dimensionality reduction algorithm dubbed bilinear multi-emotion similarity preserving embedding (BME-SPE) is developed to uncover the intrinsic relationship between the music signals and the EEG-adjusted emotion labels.

The rest of this paper is organized as follows. In Section 2, we briefly review the related work on EEG-based music emotion analysis and multi-label dimensionality reduction. Section 3 introduces the proposed framework with the analysis of computational cost. In Section 4, a series of experiments are conducted on a standard Western music dataset CAL-500 and a self-collected Chinese music dataset to evaluate the performance of the proposed framework. We conclude the paper and discuss the future work in Section 5.

2 Related Work

2.1 EEG-based Music Emotion Analysis

The neural mechanisms involved in music emotion understanding remain as active research topics in the neuroscience community [39], [40]. As one of the most widely used brain imaging technologies that records the brains electrical activities using electrodes attached to the scalp, Electroencephalography (EEG) has been proven to provide informative characteristics in responses to the emotional states [9], [60], and thus has been applied to many music emotion analysis tasks.

In [68], Schmidt and Trainor examined whether the pattern of regional EEG activity distinguished emotions induced by musical excerpts via the regional brain activation/emotion models. Altenmüller et al. recorded the EEG activation patterns in order to investigate the neurobiological mechanisms accompanying emotional valence judgements during listening to complex auditory stimuli [1]. Baumgartner et al. studied the influence of visual and musical stimuli on brain processing by using highly arousing pictures and classical musical excerpts together to evoke the three basic emotions of happiness, sadness and fear [4]. In [48], Lin et al. investigated the connections between emotional states and brain activity recorded by EEG. They used machine learning algorithms to categorize EEG dynamics according to subject self-reported emotional states during music listening. Kroupi et al. analyzed the EEG for assessing emotions evoked during watching various pre-selected emotional music video clips. Specifically, they extracted the time domain and frequency domain features of the EEG signal, and then analyzed the subject-dependent and subject-independent correlations between extracted features and subjects self assessed emotions [42]. Trochidis and Bigand recorded the EEG activity during music listening in different regions without a-priori defining regions of interest and then analyzed the alpha and theta bands separately, which confirmed the hemispheric specialization hypothesis for emotional valence [72]. In [41], Koelstra et al. first presented a multimodal data set for the analysis of human affective states. The EEG and peripheral physiological signals during watching music videos were recorded. They then proposed a semi-automatic stimuli selection method using affective tags, and conducted single-trial classification to the features extracted from the EEG, peripheral and multimedia content analysis modalities. Duan et al. utilized pure music segments as stimuli to evoke the exciting or relaxing emotions of subjects and then extracted the EEG power spectrum as the features for the task of binary emotion classification [15]. Cabredo et al. collected EEG signals from subjects when they are listening to emotion-inducing music. Then the EEG signals were converted into emotion annotation by the emotion spectrum analysis method and C4.5 was used to build the emotion models [7]. Daly et al. employed a large set of musical stimuli drawn from different styles, and analyzed neural correlates of music-induced emotions based on the recorded EEG [12]. By combining the EEG dynamics and acoustic characteristics of musical contents, Lin et al. developed a multimodal approach for the classification of emotional valence and arousal [49].

2.2 Multi-label Dimensionality Reduction

In musical emotion analysis, a song sometimes conveys or evokes more than one emotions. Some researchers therefore formulate it as a multi-label learning problem [73], [86]. Unlike the single-label learning in which each data point belongs to only one category, the multi-label learning is more general than the single-label case that each data point might be associated with multiple labels [95]. More importantly, an implicit assumption in single-label learning is that the labels are mutually exclusive while in multi-label learning it is possible that the labels are correlated with each other [63]. Driven by various applications such as image classification [64] and text categorization [93], many multi-label learning algorithms have been proposed, such as multi-label k-nearest neighbor [94], multi-label support vector machines [24], multi-label neural networks [93], etc. A more comprehensive review on multi-label learning algorithms could be found in [95].

Furthermore, the music signals often have a huge number of features [8], [74], which may contain a large amount of redundant information and thus cause the high computational cost and poor performance of the analysis task. In order to discover the intrinsic features hidden in the original high-dimensional space, multi-label dimensionality reduction becomes our first choice for the task of music emotion analysis. To the best of our knowledge, there is no work on multi-label dimensionality reduction for music emotion analysis. However, multi-label dimensionality reduction itself is already an active research area in machine learning. Yu et al. proposed a method called multi-label informed latent semantic indexing to preserve the information of data and meanwhile capture the correlations between the multiple labels [91]. Arenas-Garca et al. presented the sparse kernel orthonormalized partial least squares to handle the multi-label data [2]. Sun et al. proposed the hypergraph spectral learning, which generalize the graph Laplacian to the hypergraph Laplacian for multi-label applications [69]. Park and Lee extended the traditional linear discriminant analysis to the multi-label version by applying the copy transformation [59]. Wang et al. proposed another multi-label linear discriminant analysis algorithm by taking advantage of label correlations [81]. Zhang and Zhou introduced a multi-label dimensionality reduction algorithm by maximizing the dependence between data and corresponding labels [96]. Ji et al. proposed a shared-subspace learning model for multi-label classification [33]. Other well-known dimensionality reduction schemes, such as nonnegative factorization, canonical correlation analysis, and sparse coding, have also been extended for multi-label classification [58], [70], [80].

3 Proposed Framework

3.1 EEG-based Emotion Smoothing

In music emotion analysis, the emotion labels are generally scored by users. In most of the situations, the score on each emotion label will be a binary choice, i.e., 0–1, where 1 indicates that the music is able to convey the corresponding emotion and 0 otherwise; or a multi-level choice, e.g., 0–1–2–3, where 0 denotes the weakest extent of the corresponding emotion while 3 denotes the strongest.

Although the user-provided scores can help to link the music and the corresponding emotions, sometimes they are too "hard" to reflect the extent of the evoked emotions accurately. In this subsection, we introduce a scheme called EEG-based emotion smoothing (E^{2}S) to refine the "hard" labels by using the recorded EEG signals, which have been proven to provide informative characteristics in responses to the emotional states [9], [60].

Let $\mathbb{X} = \mathbb{R}^{D_1 \times D_2}$ be the high-dimensional feature space of the music signal, and there is an emotion set $\mathbb{E}$ including m emotion labels. The original user-provided emotions associated with the i-th song can be represented as an m-dimensional vector $\mathbf{y}_i^o$, where $\mathbf{y}_i^o \in \{0,1\}^m$ or $\mathbf{y}_i^o \in \{0,1,...,c\}^m$. Here c denotes the number of levels of the emotion intension. The corresponding EEG signal is represented as a D_g-dimensional vector $\mathbf{g}_i$. Given the training dataset $\{(\mathbf{X}_1, \mathbf{y}_1^o, \mathbf{g}_1), \ ... \ , (\mathbf{X}_{n_g}, \mathbf{y}_{n_g}^o, \mathbf{g}_{n_g}), \ (\mathbf{X}_{n_g+1}, \mathbf{y}_{n_g+1}^o), \ ... \ , (\mathbf{X}_n, \mathbf{y}_n^o)\}$, where n_g denotes the number of songs having the EEG recording and n denotes the total number of songs. First, E^{2}S refines the emotion labels of the songs having the EEG recording by introducing the following objective function:

$$\begin{aligned}
\mathbf{Y}^g &= \arg\min_{\mathbf{Y}^g} Q_1(\mathbf{Y}^g) \\
&= \arg\min_{\mathbf{Y}^g} \frac{1}{2} \sum_{i,j=1}^{n_g} \left\| \frac{\mathbf{y}_i^g}{\sqrt{D_{ii}^g}} - \frac{\mathbf{y}_j^g}{\sqrt{D_{jj}^g}} \right\|^2 A_{ij}^g \\
&\quad + \alpha \sum_{i=1}^{n_g} \|\mathbf{y}_i^g - \mathbf{y}_i^o\|^2,
\end{aligned} \tag{1}$$

where $\mathbf{Y}^g = [\mathbf{y}_1^g, ..., \mathbf{y}_{n_g}^g]$ is composed of the n_g refined emotion label vectors of those songs who have the corresponding EEG recording, the coefficient $A_{ij}^g = exp(-||\mathbf{g}_i - \mathbf{g}_j||^2/2\sigma)$ measures the similarity between the i-th and the j-th EEG signal vectors, the coefficient $D_{ii}^g = \sum_{j=1}^{n_g} A_{ij}^g$ $(i = 1, ..., n_g)$ is used to remove the scaling factor, and $\alpha > 0$ is a regularization parameter. In (1), the first term of $Q_1(\mathbf{Y}^g)$ ensures that the songs who generate the similar EEG signals have similar emotion labels, while the second term of $Q_1(\mathbf{Y}^g)$ requires the consistency between the refined labels and initial labels [97].

Differentiating $Q_1(\mathbf{Y}^g)$ with respect to $\mathbf{Y}^g$, we have

$$\frac{\partial Q_1}{\partial \mathbf{Y}^g} = \mathbf{Y}^g - \mathbf{Y}^g(\mathbf{D}^g)^{-\frac{1}{2}}\mathbf{A}^g(\mathbf{D}^g)^{-\frac{1}{2}} + \alpha(\mathbf{Y}^g - \mathbf{Y}_{1:n_g}^o) = 0, \tag{2}$$

where $\mathbf{A}^g = [A_{ij}^g]$ is the $n_g \times n_g$ matrix, $\mathbf{D}^g = diag(D_{ii}^g)$ is the $n_g \times n_g$ diagonal matrix, and $\mathbf{Y}_{1:n_g}^o = [\mathbf{y}_1^o, ..., \mathbf{y}_{n_g}^o]$. Then we can obtain:

$$\mathbf{Y}^g = \alpha \mathbf{Y}_{1:n_g}^o \big((1+\alpha)\mathbf{I} - (\mathbf{D}^g)^{-\frac{1}{2}}\mathbf{A}^g(\mathbf{D}^g)^{-\frac{1}{2}}\big)^{\dagger}, \tag{3}$$

where $(\cdot)^{\dagger}$ denotes the Moore-Penrose pseudoinverse [25]. Now the refined emotion label matrix $\mathbf{Y}^{new} = [\mathbf{y}_1^g, ..., \mathbf{y}_{n_g}^g, \mathbf{y}_{n_g+1}^o, ..., \mathbf{y}_n^o]$.

Algorithm 1: EEG-based Emotion Smoothing (E^{2}S)

Input: Training dataset: $\{(\mathbf{X}_1, \mathbf{y}_1^o), ..., (\mathbf{X}_n, \mathbf{y}_n^o)\}$; the set of recorded EEG signals: $\{\mathbf{g}_1, ..., \mathbf{g}_{n_g}\}$; the regularization parameters: α, β

Output: The refined label vectors: $\{\mathbf{y}_1, ..., \mathbf{y}_n\}$

for $i = 1, ..., n_g$ **do**
  for $j = 1, ..., n_g$ **do**
    $A_{ij}^g \leftarrow exp(-||\mathbf{g}_i - \mathbf{g}_j||^2/2\sigma)$;
  $D_{ii}^g \leftarrow \sum_{j=1}^{n_g} A_{ij}^g$;
$\mathbf{A}^g \leftarrow [A_{ij}^g]_{n_g \times n_g}$;
$\mathbf{D}^g \leftarrow diag(D_{ii}^g)_{n_g \times n_g}$;
$\mathbf{Y}_{1:n_g}^o \leftarrow [\mathbf{y}_1^o, ..., \mathbf{y}_{n_g}^o]$;
$\mathbf{Y}^g \leftarrow \alpha \mathbf{Y}_{1:n_g}^o \big((1+\alpha)\mathbf{I} - (\mathbf{D}^g)^{-\frac{1}{2}}\mathbf{A}^g(\mathbf{D}^g)^{-\frac{1}{2}}\big)^{\dagger}$;
for $i = 1, ..., n$ **do**
  for $j = 1, ..., n$ **do**
    $A_{ij} \leftarrow exp(-||\mathbf{X}_i - \mathbf{X}_j||_F^2/2\sigma)$;
  $D_{ii} \leftarrow \sum_{j=1}^{n} A_{ij}$;
$\mathbf{A} \leftarrow [A_{ij}]_{n \times n}$;
$\mathbf{D} \leftarrow diag(D_{ii})_{n \times n}$;
$\mathbf{Y}^{new} \leftarrow [\mathbf{y}_1^g, ..., \mathbf{y}_{n_g}^g, \mathbf{y}_{n_g+1}^o, ..., \mathbf{y}_n^o]$;
$\mathbf{Y}^{\bar{g}} \leftarrow \beta \mathbf{Y}^{new} \big((1+\beta)\mathbf{I} - \mathbf{D}^{-\frac{1}{2}}\mathbf{A}\mathbf{D}^{-\frac{1}{2}}\big)^{\dagger}$;
$[\mathbf{y}_1, ..., \mathbf{y}_n] \leftarrow [\mathbf{y}_1^g, ..., \mathbf{y}_{n_g}^g, \mathbf{y}_{n_g+1}^{\bar{g}}, ..., \mathbf{y}_n^{\bar{g}}]$;

Then the proposed E^{2}S further refines the emotion labels of the remaining songs who do not have EEG recording by minimizing the following objective function:

$$\begin{aligned}
\mathbf{Y}^{\bar{g}} &= \arg\min_{\mathbf{Y}^{\bar{g}}} Q_2(\mathbf{Y}^{\bar{g}}) \\
&= \arg\min_{\mathbf{Y}^{\bar{g}}} \frac{1}{2} \sum_{i,j=1}^{n} \left\| \frac{\mathbf{y}_i^{\bar{g}}}{\sqrt{D_{ii}}} - \frac{\mathbf{y}_j^{\bar{g}}}{\sqrt{D_{jj}}} \right\|^2 A_{ij} \\
&\quad + \beta \Big(\sum_{i=1}^{n_g} \|\mathbf{y}_i^{\bar{g}} - \mathbf{y}_i^g\|^2 + \sum_{i=n_g+1}^{n} \|\mathbf{y}_i^{\bar{g}} - \mathbf{y}_i^o\|^2 \Big),
\end{aligned} \tag{4}$$

where $A_{ij} = exp(-||\mathbf{X}_i - \mathbf{X}_j||_F^2/2\sigma)$ measures the similarity between the i-th and the j-th data representations, $D_{ii} = \sum_{j=1}^{n} A_{ij}$ $(i = 1, ..., n)$ is used to remove the scaling factor, $\beta > 0$ is a regularization parameter, and $\|\cdot\|_F$ denotes the Frobenius norm. Similar to (3), the solution of (4) is given by:

$$\mathbf{Y}^{\bar{g}} = \beta \mathbf{Y}^{new} \big((1+\beta)\mathbf{I} - \mathbf{D}^{-\frac{1}{2}}\mathbf{A}\mathbf{D}^{-\frac{1}{2}}\big)^{\dagger}, \tag{5}$$

where $\mathbf{A} = [A_{ij}]$ is the $n \times n$ matrix and $\mathbf{D} = diag(D_{ii})$ is the $n \times n$ diagonal matrix.

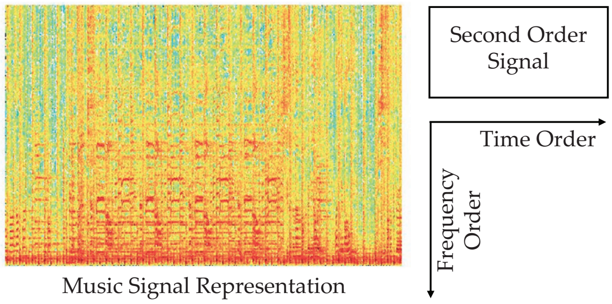

Fig. 1. Second-order representation of the music signal. The first order is the frequency order and the second order is the time order.

In order to emphasize the effect of EEG on label smoothing, we introduce the constraint that $\mathbf{y}_i^{\bar{g}} = \mathbf{y}_i^g$ for $i = 1, ..., n_g$ to keep the EEG-refined labels unchanged. The final label matrix is therefore given as follows:

$$\mathbf{Y} = [\mathbf{Y}^g, \mathbf{Y}^{\bar{g}}_{n_g+1:n}] = [\mathbf{y}_1^g, ..., \mathbf{y}_{n_g}^g, \mathbf{y}^{\bar{g}}_{n_g+1}, ..., \mathbf{y}_n^{\bar{g}}]. \tag{6}$$

The detailed procedure of E^{2}S is described in Algorithm 1.

3.2 Bilinear Multi-Emotion Similarity Preserving Embedding

In this subsection, we propose a multi-label dimensionality reduction algorithm to extract the intrinsic features embedded in the music signal that essentially evoke human emotions. In order to adapt to the music signals, which are naturally represented by the second-order tensors (i.e., matrices) in the time-frequency domain as shown in Fig. 1, the proposed algorithm employs the bilinear learning strategy and thus is able to take the second-order signals as the input directly[2].

The proposed Bilinear Multi-Emotion Similarity Preserving Embedding (BME-SPE) algorithm aims to map the original high-dimensional music signal representations into a low-dimensional feature subspace, in which we hope that a clearer linkage between the features and emotions could be discovered. The idea behind the proposed method is very simple: if two songs can convey similar emotions, they should possess some hidden features in common. Specifically, given the training set $\{(\mathbf{X}_1, \mathbf{y}_1), ..., (\mathbf{X}_n, \mathbf{y}_n)\}$, BME-SPE aims to learn two transformation matrices $\mathbf{U} = [\mathbf{u}_1, ..., \mathbf{u}_{d_1}] \in \mathbb{R}^{D_1 \times d_1}$ and $\mathbf{V} = [\mathbf{v}_1, ..., \mathbf{v}_{d_2}] \in \mathbb{R}^{D_2 \times d_2}$ to project data to a low-dimensional subspace $\mathbb{Z} = \mathbb{R}^{d_1 \times d_2}$, where the data with similar emotion labels are close to each other. The objective function of BME-SPE is formulated as follows:

$$\min \sum_{i=1}^{n} \sum_{j=1}^{n} \|\mathbf{Z}_i - \mathbf{Z}_j\|_F^2 \cdot S_{ij}, \tag{7}$$

2. An earlier version of the algorithm ME-SPE designed for the first-order input (i.e., vectors) has been appeared in [51].

where $\mathbf{Z}_i \in \mathbb{Z}$ denotes the low-dimensional representation of the i-th song, and S_{ij} denotes the emotional similarity between the i-th and the j-th songs $(i, j = 1, ..., n)$. Since the mapping from the original high-dimensional space to the low-dimensional subspace is given by $\mathbf{Z}_i = \mathbf{U}^T\mathbf{X}_i\mathbf{V}$, we rewrite the objective function as follows:

$$\begin{aligned}\mathbf{U}, \mathbf{V} &= \arg\min_{\mathbf{U},\mathbf{V}} J(\mathbf{U}, \mathbf{V}) \\ &= \arg\min_{\mathbf{U},\mathbf{V}} \sum_{i=1}^{n} \sum_{j=1}^{n} \|\mathbf{U}^T\mathbf{X}_i\mathbf{V} - \mathbf{U}^T\mathbf{X}_j\mathbf{V}\|_F^2 \cdot S_{ij}.\end{aligned} \tag{8}$$

There are many different ways to define the similarity function S. In our formulation, we choose the form of inner product as it is able to capture the information of correlations between different emotions.

Let $S_{ij} = \langle \mathbf{y}_i, \mathbf{y}_j \rangle$, where $\langle \cdot, \cdot \rangle$ denotes the inner product operation, then the similarity matrix $\mathbf{S} = [S_{ij}]_{n\times n} = \mathbf{Y}^T\mathbf{Y}$, where $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_n]$ is the refined label matrix generated by the E^{2}S scheme introduced in Section 3.1. Let $\mathbf{y}_{(i)}$ $(i = 1, ..., m)$ be the label indication vector for the i-th emotion. In fact, the transpose of $\mathbf{y}_{(i)}$ is the i-th row of $\mathbf{Y}$. Obviously, the matrix $\mathbf{S}^E = \mathbf{Y}\mathbf{Y}^T$ is an $m \times m$ matrix in which the (i, j)-th component indicates the similarity between the i-th emotion and the j-th emotion. Actually, the matrix $\mathbf{S}$ and $\mathbf{S}^E$ are closely related since they have the same non-zero eigenvalues. Suppose μ is a non-zero eigenvalue of $\mathbf{S}^E$, then we have

$$\mathbf{S}^E\mathbf{a} = \mathbf{Y}\mathbf{Y}^T\mathbf{a} = \mu\mathbf{a}, \tag{9}$$

where $\mathbf{a}$ is the eigenvector of $\mathbf{S}^E$ corresponding to μ. Left multiply above equation by $\mathbf{Y}^T$, we obtain

$$\mathbf{S}\mathbf{Y}^T\mathbf{a} = \mathbf{Y}^T\mathbf{Y}\mathbf{Y}^T\mathbf{a} = \mu\mathbf{Y}^T\mathbf{a}, \tag{10}$$

which means that μ is also the eigenvalue of $\mathbf{S}$, with the corresponding eigenvector $\mathbf{Y}^T\mathbf{a}$. Therefore, by formulating such a data similarity matrix $\mathbf{S}$, the correlation between different labels in $\mathbf{S}^E$ is well captured.

In order to normalize the similarity values into the interval $[0, 1]$, we define the normalized similarity matrix $\hat{\mathbf{S}}$ where

$$\hat{S}_{ij} = \langle \hat{\mathbf{y}}_i, \hat{\mathbf{y}}_j \rangle = \langle \mathbf{y}_i/||\mathbf{y}_i||, \mathbf{y}_j/||\mathbf{y}_j|| \rangle. \tag{11}$$

The objective function of BME-SPE now becomes

$$\begin{aligned}\mathbf{U}, \mathbf{V} &= \arg\min_{\mathbf{U},\mathbf{V}} J(\mathbf{U}, \mathbf{V}) \\ &= \arg\min_{\mathbf{U},\mathbf{V}} \sum_{i=1}^{n} \sum_{j=1}^{n} \|\mathbf{U}^T\mathbf{X}_i\mathbf{V} - \mathbf{U}^T\mathbf{X}_j\mathbf{V}\|_F^2 \cdot \hat{S}_{ij}.\end{aligned} \tag{12}$$

Since (12) is not a convex optimization problem, there is no closed-form solution for it. Instead, we utilize an alternating strategy [28], [44] to find a locally optimal solution.

Step 1: First, we fix $\mathbf{V}$ to obtain the optimal $\mathbf{U}$. Eq. (12) could be rewritten as follows:

$$\begin{aligned}\mathbf{U} &= \arg\min_{\mathbf{U}} J_{\mathbf{V}}(\mathbf{U}) \\ &= \arg\min_{\mathbf{U}} \sum_{i=1}^{n}\sum_{j=1}^{n} ||\mathbf{U}^T\mathbf{X}_i^{\mathbf{V}} - \mathbf{U}^T\mathbf{X}_j^{\mathbf{V}}||_F^2 \cdot \hat{S}_{ij},\end{aligned} \tag{13}$$

where $\mathbf{X}_i^{\mathbf{V}} = \mathbf{X}_i\mathbf{V}$. Then we have

$$\begin{aligned}&\sum_{i,j=1}^{n} ||\mathbf{U}^T\mathbf{X}_i^{\mathbf{V}} - \mathbf{U}^T\mathbf{X}_j^{\mathbf{V}}||_F^2 \cdot \hat{S}_{ij} \\ =&\sum_{i,j=1}^{n} tr\big((\mathbf{U}^T\mathbf{X}_i^{\mathbf{V}} - \mathbf{U}^T\mathbf{X}_j^{\mathbf{V}})(\mathbf{U}^T\mathbf{X}_i^{\mathbf{V}} - \mathbf{U}^T\mathbf{X}_j^{\mathbf{V}})^T\big)\hat{S}_{ij} \\ =&\sum_{i,j=1}^{n} tr\big(\mathbf{U}^T(\mathbf{X}_i^{\mathbf{V}} - \mathbf{X}_j^{\mathbf{V}})(\mathbf{X}_i^{\mathbf{V}} - \mathbf{X}_j^{\mathbf{V}})^T\mathbf{U}\big)\hat{S}_{ij} \\ =&2 \cdot tr\big(\mathbf{U}^T(\sum_{i=1}^{n} D_{ii}\mathbf{X}_i^{\mathbf{V}}(\mathbf{X}_i^{\mathbf{V}})^T - \sum_{i,j=1}^{n} \hat{S}_{ij}\mathbf{X}_i^{\mathbf{V}}(\mathbf{X}_j^{\mathbf{V}})^T)\mathbf{U}\big),\end{aligned} \tag{14}$$

where $D_{ii} = \sum_{j=1}^{n} \hat{S}_{ij}$ $(i = 1, ..., n)$, and $tr(\cdot)$ denotes the matrix trace operator. The problem in (13) now becomes

$$\mathbf{U} = \arg\min_{\mathbf{U}} tr\big(\mathbf{U}^T(\mathbf{D}^{\mathbf{V}} - \mathbf{S}^{\mathbf{V}})\mathbf{U}\big), \tag{15}$$

where $\mathbf{D}^{\mathbf{V}} = \sum_{i=1}^{n} D_{ii}\mathbf{X}_i^{\mathbf{V}}(\mathbf{X}_i^{\mathbf{V}})^T$ and $\mathbf{S}^{\mathbf{V}} = \sum_{i,j=1}^{n} \hat{S}_{ij}\mathbf{X}_i^{\mathbf{V}}(\mathbf{X}_j^{\mathbf{V}})^T$. Additionally, we introduce the constraint $\mathbf{U}^T\mathbf{D}^{\mathbf{V}}\mathbf{U} = \mathbf{I}_{d_1}$ to remove the scaling factor in the learning process, where $\mathbf{I}_{d_1}$ denotes the d_1-dimensional identity matrix. So for the first transformation vector $\mathbf{u}_1$, the problem becomes

$$\mathbf{u}_1 = \mathop{\arg\min}_{\substack{\mathbf{u}_1 \\ \mathbf{u}_1^T\mathbf{D}^{\mathbf{V}}\mathbf{u}_1=1}} \mathbf{u}_1^T(\mathbf{D}^{\mathbf{V}} - \mathbf{S}^{\mathbf{V}})\mathbf{u}_1. \tag{16}$$

Then we obtain the Lagrangian equation of (16):

$$L(\mathbf{u}_1, \lambda) = \mathbf{u}_1^T(\mathbf{D}^{\mathbf{V}} - \mathbf{S}^{\mathbf{V}})\mathbf{u}_1 - \lambda(\mathbf{u}_1^T\mathbf{D}^{\mathbf{V}}\mathbf{u}_1 - 1). \tag{17}$$

Letting $\partial L(\mathbf{u}_1, \lambda)/\partial \mathbf{u}_1 = 0$, the optimal $\mathbf{u}_1$ is therefore the eigenvector corresponding to the smallest non-zero eigenvalue of the generalized eigendecomposition problem

$$(\mathbf{D}^{\mathbf{V}} - \mathbf{S}^{\mathbf{V}})\mathbf{u} = \lambda\mathbf{D}^{\mathbf{V}}\mathbf{u}. \tag{18}$$

Similarly, $\mathbf{u}_2, ..., \mathbf{u}_d$ are the eigenvectors corresponding to the 2-nd, ..., d-th smallest non-zero eigenvalues of (18), respectively.

Step 2: After we have obtained the optimal $\mathbf{U}$, we fix it to find the optimal $\mathbf{V}$. Eq. (12) now could be rewritten as follows:

$$\begin{aligned}\mathbf{V} &= \arg\min_{\mathbf{V}} J_{\mathbf{U}}(\mathbf{V}) \\ &= \arg\min_{\mathbf{V}} \sum_{i=1}^{n}\sum_{j=1}^{n} ||\mathbf{X}_i^{\mathbf{U}}\mathbf{V} - \mathbf{X}_j^{\mathbf{U}}\mathbf{V}||_F^2 \cdot \hat{S}_{ij},\end{aligned} \tag{19}$$

where $\mathbf{X}_i^{\mathbf{U}} = \mathbf{U}^T\mathbf{X}_i$. Then we have

$$\begin{aligned}&\sum_{i,j=1}^{n} ||\mathbf{X}_i^{\mathbf{U}}\mathbf{V} - \mathbf{X}_j^{\mathbf{U}}\mathbf{V}||_F^2 \cdot \hat{S}_{ij} \\ =&\sum_{i,j=1}^{n} tr\big((\mathbf{X}_i^{\mathbf{U}}\mathbf{V} - \mathbf{X}_j^{\mathbf{U}}\mathbf{V})^T(\mathbf{X}_i^{\mathbf{U}}\mathbf{V} - \mathbf{X}_j^{\mathbf{U}}\mathbf{V})\big)\hat{S}_{ij} \\ =&\sum_{i,j=1}^{n} tr\big(\mathbf{V}^T(\mathbf{X}_i^{\mathbf{U}} - \mathbf{X}_j^{\mathbf{U}})^T(\mathbf{X}_i^{\mathbf{U}} - \mathbf{X}_j^{\mathbf{U}})\mathbf{V}\big)\hat{S}_{ij} \\ =&2 \cdot tr\big(\mathbf{V}^T(\sum_{i=1}^{n} D_{ii}(\mathbf{X}_i^{\mathbf{U}})^T\mathbf{X}_i^{\mathbf{U}} - \sum_{i,j=1}^{n} \hat{S}_{ij}(\mathbf{X}_i^{\mathbf{U}})^T\mathbf{X}_j^{\mathbf{U}})\mathbf{V}\big).\end{aligned} \tag{20}$$

The problem in (19) becomes

$$\mathbf{V} = \arg\min_{\mathbf{V}} tr\big(\mathbf{V}^T(\mathbf{D}^{\mathbf{U}} - \mathbf{S}^{\mathbf{U}})\mathbf{V}\big), \tag{21}$$

where $\mathbf{D}^{\mathbf{U}} = \sum_{i=1}^{n} D_{ii}(\mathbf{X}_i^{\mathbf{U}})^T\mathbf{X}_i^{\mathbf{U}}$ and $\mathbf{S}^{\mathbf{U}} = \sum_{i,j=1}^{n} \hat{S}_{ij}(\mathbf{X}_i^{\mathbf{U}})^T\mathbf{X}_j^{\mathbf{U}}$. Similarly, we introduce the constraint $\mathbf{V}^T\mathbf{D}^{\mathbf{U}}\mathbf{V} = \mathbf{I}_{d_2}$ to remove the scaling factor. Therefore, the optimal $\mathbf{V}$ that minimizes the objective function in (19) is composed of the eigenvectors corresponding to the d_2 smallest non-zero eigenvalues of the generalized eigendecomposition problem

$$(\mathbf{D}^{\mathbf{U}} - \mathbf{S}^{\mathbf{U}})\mathbf{v} = \lambda\mathbf{D}^{\mathbf{U}}\mathbf{v}. \tag{22}$$

Above two steps are alternately executed until the learning procedure converges. The proof of the convergency is provided in the Appendix. The detailed procedure of BME-SPE is described in Algorithm 2.

3.3 Computational Complexity Analysis

In this subsection, we analyze the computational complexity of the proposed framework.

E^{2}S: The time cost of E^{2}S mainly comes from four aspects: constructing $\mathbf{A}^g$ and $\mathbf{D}^g$, calculating $\mathbf{Y}^g$, constructing $\mathbf{A}$ and $\mathbf{D}$, and calculating $\mathbf{Y}^{\bar{g}}$. The cost of these four steps are $O(n_g^2D_g)$, $O(n_g^3 + n_g^2m)$, $O(n^2D_1D_2)$, and $O(n^3+n^2m)$, respectively. Since $n_g \leq n$, the total time cost of E^{2}S is $O(n_g^2D_g + n^2D_1D_2 + n^3 + n^2m)$.

BME-SPE: The training cost of BME-SPE mainly comes from three aspects: the calculation of $\hat{S}_{ij}$, constructing $\mathbf{D}^{\mathbf{V}}$, $\mathbf{S}^{\mathbf{V}}$, $\mathbf{D}^{\mathbf{U}}$, $\mathbf{S}^{\mathbf{U}}$, and solving (18) and (22). The cost of these three steps are $O(mn^2)$, $O(nT(D_1d_2 + D_2d_1)(D_1 + D_2))$, and $O(T(D_1^3 + D_2^3))$, respectively, where T is the number of iterations needed for algorithm convergence. Therefore, the total training cost of BME-SPE is $O(mn^2 + nT(D_1d_2 + D_2d_1)(D_1 + D_2) + T(D_1^3 + D_2^3))$. In the test phase of BME-SPE, the most demanding step is projecting the high-dimensional data to the learned subspace, whose cost is $O(D_1D_2\min(d_1, d_2))$.

Algorithm 2: Bilinear Multi-Emotion Similarity Preserving Embedding (BME-SPE)

Input: Training dataset: $\{(\mathbf{X}_1, \mathbf{y}_1), ..., (\mathbf{X}_n, \mathbf{y}_n)\}$; the dimensions of the subspace: d_1, d_2; the stop threshold: ε

Output: Transformation matrices: $\mathbf{U}$, $\mathbf{V}$

1 **for** $i = 1, ..., n$ **do**
2   **for** $j = 1, ..., n$ **do**
3     $\hat{S}_{ij} \leftarrow \langle \mathbf{y}_i/||\mathbf{y}_i||, \mathbf{y}_j/||\mathbf{y}_j|| \rangle$;
4 **for** $i = 1, ..., n$ **do**
5   $D_{ii} \leftarrow \sum_{j=1}^{n} \hat{S}_{ij}$;
6 $t \leftarrow 0$;
7 $\mathbf{U}_{(t)}, \mathbf{V}_{(t)} \leftarrow$ arbitrary column orthogonal matrices;
8 **for** $i = 1, ..., n$ **do**
9   $\mathbf{X}_i^{\mathbf{V}_{(t)}} \leftarrow \mathbf{X}_i \mathbf{V}_{(t)}$;
10 $\mathbf{D}^{\mathbf{V}} = \sum_{i=1}^{n} D_{ii} \mathbf{X}_i^{\mathbf{V}_{(t)}} (\mathbf{X}_i^{\mathbf{V}_{(t)}})^T$;
11 $\mathbf{S}^{\mathbf{V}} = \sum_{i,j=1}^{n} \hat{S}_{ij} \mathbf{X}_i^{\mathbf{V}_{(t)}} (\mathbf{X}_j^{\mathbf{V}_{(t)}})^T$;
12 **for** $i = 1, ..., d_1$ **do**
13   Solve $(\mathbf{D}^{\mathbf{V}} - \mathbf{S}^{\mathbf{V}})\mathbf{u}_i = \lambda \mathbf{D}^{\mathbf{V}} \mathbf{u}_i$;
14 $\mathbf{U}_{(\mathbf{t+1})} \leftarrow [\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_{d_1}]$;
15 **for** $i = 1, ..., n$ **do**
16   $\mathbf{X}_i^{\mathbf{U}_{(t+1)}} = \mathbf{U}_{(t+1)}^T \mathbf{X}_i$;
17 $\mathbf{D}^{\mathbf{U}} = \sum_{i=1}^{n} D_{ii} (\mathbf{X}_i^{\mathbf{U}_{(t+1)}})^T \mathbf{X}_i^{\mathbf{U}_{(t+1)}}$;
18 $\mathbf{S}^{\mathbf{U}} = \sum_{i,j=1}^{n} \hat{S}_{ij} (\mathbf{X}_i^{\mathbf{U}_{(t+1)}})^T \mathbf{X}_j^{\mathbf{U}_{(t+1)}}$;
19 **for** $i = 1, ..., d_2$ **do**
20   Solve $(\mathbf{D}^{\mathbf{U}} - \mathbf{S}^{\mathbf{U}})\mathbf{v}_i = \lambda \mathbf{D}^{\mathbf{U}} \mathbf{v}_i$;
21 $\mathbf{V}_{(\mathbf{t+1})} \leftarrow [\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_{d_2}]$;
22 **while** $|J_t(\mathbf{U}, \mathbf{V}) - J_{t+1}(\mathbf{U}, \mathbf{V})| > \varepsilon$ **do**
23   $t \leftarrow t + 1$;
24   Repeat lines 8 – 21;

4 Experiments

In this section, we evaluate the performance of the proposed framework on a standard dataset CAL-500 [75] and a self-collected Chinese music dataset. The CAL-500 dataset includes 502 popular western songs with 18 emotion labels. Table 1 lists the concrete emotion labels of this dataset. In order to demonstrate and analyze the performance of the proposed framework from the original feature space, we do not use the extracted MFCC-related features provided by this dataset. Instead, we generate the original STFT features for each song by ourselves. Specifically, we select 30-seconds duration from the center of each song. For each duration, we divide it into short frames of 300 ms (6,615 samples at 22,050 Hz sampling rate) with 50% length overlap. For each of these frames, we calculate the 512-point length STFT. We keep only the magnitude values of the STFTs, and considering the symmetry in the STFT, we end up with inputs of dimensions 257 for each short frame. Therefore, the dimension of the input data is 51,143 (257×199). All the values in the data matrix are normalized into the interval $[0, 1]$.

TABLE 1
Eighteen emotion labels of CAL-500 dataset.

angry/aggressive, arousing, bizarre/weird, calming, carefree/lighthearted, cheerful/Festive, emotional/passionate, exciting/thrilling, happy, laid-back/mellow, light/playful, loving/romantic, pleasant, positive/optimistic, powerful/strong, sad, tender/soft, touching/loving

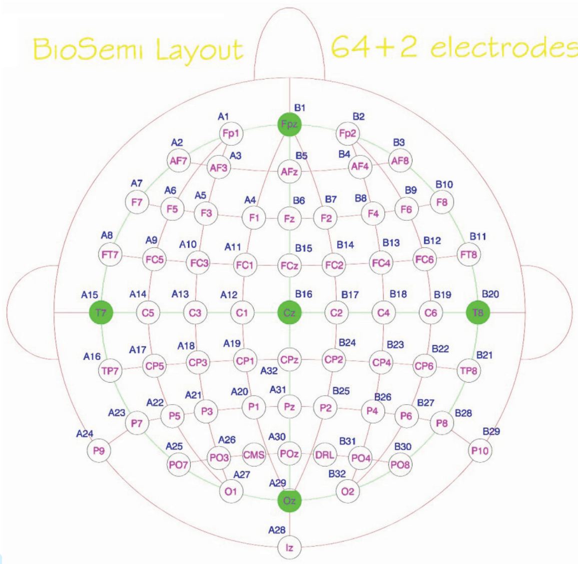

Fig. 2. Locations of 64 EEG electrodes of the BioSemi ActiveTwo system.

Besides the CAL-500 dataset, we collect a Chinese music dataset by ourselves. This dataset is composed of 100 classical and contemporary Chinese songs selected from 6 CDs of Enjoy Chinese Classical Music and 10 CDs of The Best Of Chinese Classical. To keep consistency between the CAL-500 dataset and the Chinese music dataset, we use the same way as that used in CAL-500 to generate the 257×199-dimensional feature matrix for each song in the Chinese music dataset. All the 18 emotion labels in CAL-500 dataset are covered by this Chinese music dataset. Specifically, the labels of the songs in this dataset are provided according to the descriptions and explanations of the Chinese music given by [16], [34], [47], [61], [87].

4.1 EEG Data Collection

We recorded the brain activity using 64 BioSemi pin-type active electrodes. Fig. 2 shows the locations of 64 EEG electrodes of the BioSemi ActiveTwo system[3]. We did not use a ground or reference electrode since the

3. http://www.biosemi.com/headcap.htm

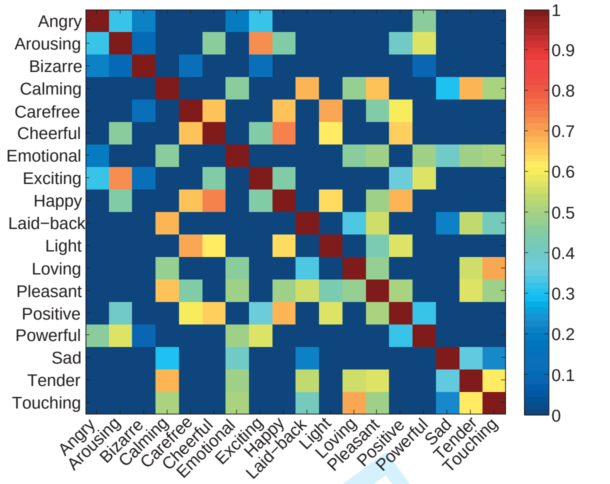

Fig. 3. Correlations of all pairs of 18 emotions from CAL-500 dataset.

BioSemi Common Mode Sense (CMS) active electrode and Driven Right Leg (DRL) passive electrode replace the ground electrodes used in conventional systems. The sampling rate and filter bandwidth were set to 4kHz and $0.16-100$Hz, respectively.

The EEG data were collected from 10 healthy subjects (7 males and 3 females with the age at 24.5 ± 3.5) during music listening. All the subjects are students or staffs at The Hong Kong Polytechnic University, with no or minimal formal musical education, and thus could be considered as nonmusicians. Subjects were instructed to keep their eyes closed and remain seated with minimal body movement during the process of music-listening. Twenty songs from the CAL-500 dataset were selected as stimuli, which cover all the 18 emotion labels. Therefore, we have $n_g = 20$. Each song was edited into a 30-second music segment. A 10-second silent rest was inserted between the music segments.

The recorded EEG signal were preprocessed via an automatic artification correction with $150\mu v$ for horizon electrooculography (HEOG) amplitude and $250\mu v$ for vertical electrooculography (VEOG) threshold to remove serious and obvious motion artifacts. Then for each of the 64 channels, we calculated the mean of the cleaned EEG signal over time as the feature of that channel. Therefore, the dimension of the EEG feature vector of each song is 64, i.e., $D_g = 64$.

4.2 Schematic Illustration of E^{2}S and BME-SPE on CAL-500 Dataset

In order to draw an intuitive picture on the relationship between different emotions, we schematically show the correlation matrix of all 18 emotions. From Fig. 3 we can observe that the most correlated emotions are "cheerful" and "happy" (the correlation coefficient is 0.7488) while the correlation of some other emotion pairs are very low, such as "happy" and "sad" (the correlation coefficient is 0).

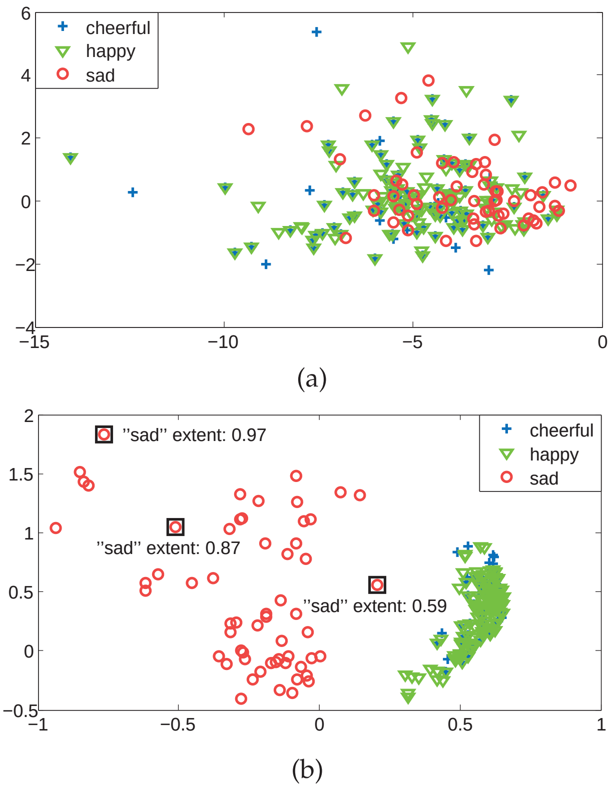

Fig. 4. 2-D representations for data points from three classes, "cheerful", "happy", and "sad", of CAL-500 dataset. (a) Visualization on 2-D principal component plane; (b) Visualization on 2-D plane learned by BME-SPE.

In the first experiment, we map the data points of above three classes, i.e., "cheerful", "happy", and "sad", onto the 2-D plane to illustrate the representation capability of BME-SPE. We also map these data onto the 2-D plane composed of the first two principal component axes (i.e., the PCA mapping) for comparison.

Fig. 4(a) shows the PCA mapping results. Most of the data points are mingled together and it is not easy to find clear boundaries to separate the "happy" class (represented by the green triangles) and the "sad" class (represented by the red circles), which should be clearly separated in the emotion space because of the low correlation coefficient.

Then we run the proposed BME-SPE on the data points of these three classes with the reduced dimension $d_1 \times d_2 = 2 \times 1$, and visualize the results in Fig. 4(b). Obviously, the data points from the "happy" class and the "sad" class are clearly separated. Moreover, the data points from the "cheerful" class and the "happy" class are largely overlapped, which is consistent with the correlation coefficient in the emotion space, and thus demonstrate that BME-SPE is able to catch the relationship between different emotions in the learning process.

TABLE 2
Performance evaluation of different dimensionality reduction methods using label-based metrics and k-nearest neighbor classifier on CAL-500 dataset. The number in the bracket denotes the reduced dimension corresponding to the best result of the algorithm under that criterion.

Methods \ Criteria	Macro average		Micro average	
	Precision	F1 Score	Precision	F1 Score
HSL	$0.285 \pm 0.057(8)$	$0.322 \pm 0.049(14)$	$0.274 \pm 0.058(13)$	$0.307 \pm 0.051(13)$
ML-LDA	$0.301 \pm 0.076(13)$	$0.266 \pm 0.041(17)$	$0.296 \pm 0.032(17)$	$0.263 \pm 0.059(14)$
ML-OPLS	$0.315 \pm 0.103(12)$	$0.270 \pm 0.030(5)$	$0.307 \pm 0.110(13)$	$0.275 \pm 0.062(4)$
MPCA	$0.429 \pm 0.103(10 \times 7)$	$0.388 \pm 0.021(10 \times 8)$	$0.398 \pm 0.099(10 \times 7)$	$0.376 \pm 0.007(16 \times 1)$
TLPP	$0.457 \pm 0.048(9 \times 8)$	$0.392 \pm 0.051(14 \times 2)$	$0.427 \pm 0.059(6 \times 13)$	$0.404 \pm 0.053(8 \times 2)$
BME-SPE	$0.462 \pm 0.039(8 \times 7)$	$0.411 \pm 0.041(9 \times 3)$	$0.432 \pm 0.040(9 \times 7)$	$0.426 \pm 0.042(7 \times 3)$
BME-SPE with E^{2}S	$\mathbf{0.479 \pm 0.054}(8 \times 6)$	$\mathbf{0.430 \pm 0.035}(12 \times 7)$	$\mathbf{0.455 \pm 0.052}(8 \times 7)$	$\mathbf{0.441 \pm 0.048}(10 \times 6)$

TABLE 3
Performance evaluation of different dimensionality reduction methods using example-based metrics and multi-label k-nearest neighbor classifier on CAL-500 dataset. The number in the bracket denotes the reduced dimension corresponding to the best result of the algorithm under that criterion.

Methods \ Criteria	Average precision	Hamming loss	One-error	Ranking loss
HSL	$0.496 \pm 0.031(2)$	$0.219 \pm 0.025(1)$	$0.585 \pm 0.050(1)$	$0.351 \pm 0.031(2)$
ML-LDA	$0.447 \pm 0.040(16)$	$0.234 \pm 0.014(14)$	$0.569 \pm 0.091(12)$	$0.388 \pm 0.039(17)$
ML-OPLS	$0.453 \pm 0.034(13)$	$0.234 \pm 0.019(17)$	$0.574 \pm 0.060(13)$	$0.380 \pm 0.043(17)$
MPCA	$0.539 \pm 0.009(16 \times 9)$	$0.241 \pm 0.034(2 \times 20)$	$0.440 \pm 0.073(12 \times 11)$	$0.342 \pm 0.026(19 \times 14)$
TLPP	$0.543 \pm 0.035(14 \times 8)$	$\mathbf{0.215 \pm 0.028}(5 \times 7)$	$0.477 \pm 0.053(6 \times 12)$	$0.304 \pm 0.029(5 \times 6)$
BME-SPE	$0.558 \pm 0.044(8 \times 10)$	$0.236 \pm 0.022(3 \times 3)$	$0.318 \pm 0.087(10 \times 10)$	$0.276 \pm 0.033(5 \times 1)$
BME-SPE with E^{2}S	$\mathbf{0.570 \pm 0.031}(8 \times 8)$	$0.218 \pm 0.026(6 \times 2)$	$\mathbf{0.313 \pm 0.069}(12 \times 7)$	$\mathbf{0.265 \pm 0.024}(5 \times 1)$

An interesting observation from Fig. 4(b) is that the variance of the data points representing "sad" songs is much larger than that of the data points representing "happy" songs, which indicates that in music emotions, the feelings of happy are all alike and every feeling of unhappy is unhappy in its own way[4].

In order to show the reasonableness of E^{2}S, we examine the refined label values of data points from the "sad" class. The original label value provided by human is a hard value, i.e., 1. We set $\alpha = \beta = 0.5$ in our experiments. Fig. 4(b) shows the refined "sad" extent of three data points (marked by the black squares). By taking the EEG consistency into account, the refined label values generated by E^{2}S seems more reasonable than the original ones: the closer the data point to the "happy" class, the lower extent of the "sad" emotion the song can evoke.

4.3 Statistical Evaluation of E^{2}S and BME-SPE on CAL-500 Dataset

In this subsection, we demonstrate the effectiveness of the proposed methods by comparing them with five dimensionality reduction algorithms, including hyper-graph spectral learning (HSL) [69], multi-label linear discriminant analysis (ML-LDA) [81], multi-label orthonormalized partial least squares (ML-OPLS) [70], multilinear PCA (MPCA) [53], and tensor LPP (TLPP) [11]. Here HSL, ML-LDA, and ML-OPLS are recently proposed multi-label dimensionality reduction methods with competent performance while MPCA and TLPP are typical tensor-based dimensionality reduction algorithms for second-order input. In order to demo the effectiveness of E^{2}S and BME-SPE clearly, we perform the dimensionality reduction on BME-SPE without E^{2}S (denoted as BME-SPE in Table 2 and Table 3) and BME-SPE with E^{2}S separately.

Two groups of criteria are used to evaluate the performance. In the first group, we use the standard label-based metrics, i.e., the precision and F1 score, as the evaluation criteria [81]. Since precision and F1 score are originally designed for binary classification, we use macro average and micro average to evaluate the overall performance across multiple labels [95]. The k-nearest neighbor classifier is used for final classification after dimensionality reduction. For all these four criteria, the larger the metric value the better the performance. In the second group, we use four standard example-based metrics, i.e., average precision, Hamming loss, one-error, and ranking loss, as the evaluation criteria [95]. The multi-label k-nearest neighbor classifier [94] is used for the final classification after dimensionality reduction. For average precision, the larger the metric value the better the performance. For Hamming loss, one-error, and ranking loss, the smaller the metric value the better the performance.

For both groups, we perform 10-fold cross valida-

4. The original sentence in the novel Anna Karenina by Russian writer Leo Tolstoy is "Happy families are all alike; every unhappy family is unhappy in its own way."

tion and set the number of nearest neighbor $k = 10$. For each algorithm, we test its performance on all the reduced dimensions and report the best result and the corresponding dimension. For HSL, ML-LDA, and ML-OPLS, the dimension of the input data is $51,143$. For MPCA, TLPP, BME-SPE, and BME-SPE with E^{2}S, the dimension of the input data is 257×199.

Table 2 reports the performance on label-based metrics with k-nearest neighbor classifier. Table 3 reports the performance on example-based metrics with multi-label k-nearest neighbor classifier. The proposed BME-SPE outperforms other algorithms on most of the evaluation criteria. Moreover, by considering the label smoothing, BME-SPE with E^{2}S further improves the performance.

4.4 EEG-Brain Mapping on Chinese Music Dataset

To further analyze the results the proposed algorithms, we conduct an experiment on the self-collected Chinese music dataset. Given the original representation of the Chinese song, i.e., $\mathbf{X}_i^C$ $(i = 1, ..., 100)$, instead of learning new transformation matrices via the proposed methods, we map it onto the low-dimensional space using the existing transformation matrices learned from the CAL-500 dataset, i.e., $\mathbf{Z}_i^C = \mathbf{U}^T\mathbf{X}_i^C\mathbf{V}$. Then for each $\mathbf{Z}_i^C$, we find its nearest neighbor from the low-dimensional representation of songs in CAL-500 dataset, i.e.,

$$\mathcal{N}(\mathbf{Z}_i^C) = \underset{\substack{\mathbf{Z}_j \\ j=1,\ldots,502}}{\arg\min} \|\mathbf{Z}_i^C - \mathbf{Z}_j\|_F^2, \tag{23}$$

where $\mathcal{N}(\mathbf{Z}_i^C)$ denotes the nearest neighbor of $\mathbf{Z}_i^C$ and $\mathbf{Z}_j$ denotes the low-dimensional representation of the j-th song in CAL-500.

We examine the distribution of EEG data of both datasets on each individual emotion. We select 20 songs from the Chinese music dataset, which cover all 18 emotions, together with the corresponding 20 Western songs, we have 40 songs in total. We record the EEG signals of subjects when listen to these songs. Then for each emotion, we use the traditional linear discriminant analysis (LDA) [21] to separate the songs who have the corresponding emotion from those who do not have in the 1-D space (the original dimension $D_g = 64$). We find that they can be clearly separated in such a low-dimensional space on all the emotions, which indicates that the Chinese and Western music share some common characteristics in evoking human emotions. Table 4 lists the most contributing electrode from all 64 as well as the corresponding Brodmann area [23] for classification on each emotion. There are 8 areas that contribute to the emotion classification, in which the areas 6, 37, 39, and 46 contribute to more than one emotions. The area 46, which is the dorsolateral prefrontal cortex area, contributes to the emotions of "happy" and "light/playful". This

TABLE 4
The most contributing electrode and the corresponding Brodmann area on each emotion.

Emotions	Most Contributing Electrode	Corresponding Brodmann Area
angry/aggressive	P6	39
arousing/awakening	P5	39
bizarre/weird	P6	39
calming/smoothing	P7	37
carefree/lighthearted	CP6	40
cheerful/festive	CZ	5
emotional/passionate	FCZ	6
exciting/thrilling	P7	37
happy	AF8	46
laid-back/mellow	P6	39
light/playful	AF8	46
loving/romantic	P7	37
pleasant/comfortable	FCZ	6
positive/optimistic	FCZ	6
powerful/strong	P7	37
sad	PO7	19
tender/soft	P7	37
touching/loving	AF3	9

observation is consistent with the findings in brain science, where the area 46 has been identified with the function of music enjoyment. More interestingly, for the areas 6, 37, and 39, it is not claimed that they have the functions closely related to music and emotion. However, all these areas have the functions on language comprehension, which might explain why they contribute to the music emotion understanding.

5 Conclusion and Future Work

This paper discovers the relationship between music and emotion via dimensionality reduction. A new learning scheme E^{2}S is proposed to refine the user-provided emotion labels by using the EEG consistency, followed by a novel multi-label dimensionality reduction technology named BME-SPE, which targets to find the genuine correlates of music emotions. The proposed methods find the influential correlates and show good performance in classification. We represent the Chinese music according to the identified correlates, and find that the music from different culture may share similar emotions.

In the future, we are specially interested in investigating the brain activities with other natural stimuli such as image browsing [76], [83] and video watching [26], [27], [31]. We will study how to combine the brain signals from different natural stimuli together, and how to apply them to various multimedia content analysis and affective computing tasks.

Appendix
Proof of Convergency of BME-SPE

Proof: We need to show that the objective function $J(\mathbf{U}, \mathbf{V})$ in (12) is nonincreasing in the learning procedure and has a lower bound.

On the one hand, the above alternating strategy indicates that in each iteration, $J(\mathbf{U}, \mathbf{V})$ is nonincreasing, i.e.,

$$\begin{aligned} J_t(\mathbf{U}, \mathbf{V}) =& J(\mathbf{U}_{(t)}, \mathbf{V}_{(t)}) = J_{\mathbf{V}_{(t)}}(\mathbf{U}_{(t)}) \\ \geq& \min J_{\mathbf{V}_{(t)}}(\mathbf{U}) = J_{\mathbf{V}_{(t)}}(\mathbf{U}_{(t+1)}) \\ =& J(\mathbf{U}_{(t+1)}, \mathbf{V}_{(t)}) = J_{\mathbf{U}_{(t+1)}}(\mathbf{V}_{(t)}) \\ \geq& \min J_{\mathbf{U}_{(t+1)}}(\mathbf{V}) = J_{\mathbf{U}_{(t+1)}}(\mathbf{V}_{(t+1)}) \\ =& J(\mathbf{U}_{(t+1)}, \mathbf{V}_{(t+1)}) = J_{t+1}(\mathbf{U}, \mathbf{V}), \end{aligned} \tag{24}$$

where $J_t(\mathbf{U}, \mathbf{V})$ denotes the value of $J(\mathbf{U}, \mathbf{V})$ after the t-th iteration, and $\mathbf{U}_{(t)}$ and $\mathbf{V}_{(t)}$ denote the matrices $\mathbf{U}$ and $\mathbf{V}$ after the t-th iteration, respectively.

On the other hand, for any i and j, we have $||\mathbf{U}^T\mathbf{X}_i\mathbf{V} - \mathbf{U}^T\mathbf{X}_j\mathbf{V}||_F^2 \geq 0$ and $\hat{S}_{ij} \geq 0$. Therefore,

$$J(\mathbf{U}, \mathbf{V}) = \sum_{i=1}^{n}\sum_{j=1}^{n} ||\mathbf{U}^T\mathbf{X}_i\mathbf{V} - \mathbf{U}^T\mathbf{X}_j\mathbf{V}||_F^2 \cdot \hat{S}_{ij} \geq 0, \tag{25}$$

which indicates that the objective function is lower bounded.

Since it has been proved that $J(\mathbf{U}, \mathbf{V})$ is nonincreasing and has a lower bound, we can conclude that the learning procedure will converge finally. □

REFERENCES

[1] E. Altenmüller, K. Schürmann, V. K. Lim, and D. Parlitz, "Hits to the left, flops to the right: different emotions during listening to music are reflected in cortical lateralisation patterns," *Neuropsychologia*, vol. 40, no. 13, pp. 2242–2256, 2002.

[2] J. Arenas-Garca, K. Petersen, and L. Hansen, "Sparse kernel orthonormalized pls for feature extraction in large data sets," in *NIPS 19*, 2007, pp. 33–40.

[3] F. G. Ashby, *Statistical analysis of fMRI data*. Cambridge, Mass.: MIT Press, 2011.

[4] T. Baumgartner, M. Esslen, and L. Jäncke, "From emotion perception to emotion experience: emotions evoked by pictures and classical music," *International Journal of Psychophysiology*, vol. 60, no. 1, pp. 34–43, 2006.

[5] A. J. Blood and R. J. Zatorre, "Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion," *Proc. Natl. Acad. Sci.*, vol. 98, no. 20, pp. 11 818–11 823, 2001.

[6] P. Brickman, J. Redfield, A. A. Harrison, and R. Crandall, "Drive and predisposition as factors in the attitudinal effects of mere exposure," *J. Exp. Soc. Psychol.*, vol. 8, no. 1, pp. 31–44, 1972.

[7] R. Cabredo, R. S. Legaspi, P. S. Inventado, and M. Numao, "Discovering emotion-inducing music features using EEG signals," *JACIII*, vol. 17, no. 3, pp. 362–370, 2013.

[8] M. Casey, C. Rhodes, and M. Slaney, "Analysis of minimum distances in high-dimensional musical spaces," *IEEE Trans. Audio, Speech, Language Process.*, vol. 16, no. 5, pp. 1015–1028, 2008.

[9] J. A. Coan and J. J. Allen, "Frontal EEG asymmetry as a moderator and mediator of emotion," *Biological Psychology*, vol. 67, no. 1C2, pp. 7–50, 2004.

[10] D. Cooke, *The language of music*. London: Oxford University Press, 1959.

[11] G. Dai and D.-Y. Yeung, "Tensor embedding methods." in *Proc. 21st AAAI*, 2006, pp. 330–335.

[12] I. Daly, A. Malik, F. Hwang, E. Roesch, J. Weaver, A. Kirke, D. Williams, E. Miranda, and S. J. Nasuto, "Neural correlates of emotional responses to music: An EEG study," *Neuroscience Letters*, vol. 573, pp. 52–57, 2014.

[13] D. Deutsch, *The Psychology of Music, 3rd Edition*. San Diego: Elsevier, 2013.

[14] J. Drever, *A Dictionary of Psychology*. Baltimore: Penguin Books Ltd, 1952.

[15] R.-N. Duan, X.-W. Wang, and B.-L. Lu, "EEG-based emotion recognition in listening music by using support vector machine and linear dynamic system." in *Proc. ICONIP*, 2012, pp. 468–475.

[16] Editorial Office of Dictionary of Chinese Music at Chinese National Academy of Arts, *Dictionary of Chinese Music (in Chinese)*. People's Music Publishing House, 2007.

[17] T. Eerola and J. K. Vuoskoski, "A comparison of the discrete and dimensional models of emotion in music," *Psychol. Music*, vol. 39, no. 1, pp. 18–49, 2011.

[18] P. Ekman, "Expression and the nature of emotion," in *Approaches to emotion*, K. Scherer and P. Ekman, Eds. Hillsdale, NJ: Erlbaum, 1984, pp. 319–344.

[19] J. Fang, X. Hu, J. Han, X. Jiang, D. Zhu, L. Guo, and T. Liu, "Data-driven analysis of functional brain interactions during free listening to music and speech," *Brain Imaging and Behavior*, pp. 1–16, 2014.

[20] Y. Feng, Y. Zhuang, and Y. Pan, "Popular music retrieval by detecting mood," in *Proc. 26th SIGIR*, 2003, pp. 375–376.

[21] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Ann. Eugen.*, vol. 7, pp. 179–188, 1936.

[22] W. Freeman and R. Quiroga, *Imaging Brain Function With EEG: Advanced Temporal and Spatial Analysis of Electroencephalographic Signals*. Springer, 2013.

[23] L. J. Garey, *Brodmann's Localisation in the Cerebral Cortex*. New York: Springer, 2006.

[24] S. Godbole and S. Sarawagi, "Discriminative methods for multi-labeled classification," in *Proc. 8th PAKDD*, 2004, pp. 22–30.

[25] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed. The Johns Hopkins University Press, 1996.

[26] J. Han, C. Chen, L. Shao, X. Hu, J. Han, and T. Liu, "Learning computational models of video memorability from fmri brain imaging," *IEEE Trans. Cybern.*, vol. PP, no. 99, pp. 1–1, 2014.

[27] J. Han, X. Ji, X. Hu, D. Zhu, K. Li, X. Jiang, G. Cui, L. Guo, and T. Liu, "Representing and retrieving video shots in human-centric brain imaging space," *IEEE Trans. Image Process.*, vol. 22, no. 7, pp. 2723–2736, 2013.

[28] X. He, D. Cai, and P. Niyogi, "Tensor subspace analysis," in *NIPS 18*, Y. Weiss, B. Schölkopf, and J. Platt, Eds. MIT Press, 2006, pp. 499–506.

[29] K. Hevner, "The affective character of the major and minor modes in music," *Am. J. Psychol.*, vol. 47, no. 1, pp. 103–118, 1935.

[30] ——, "Experimental studies of the elements of expression in music," *Am. J. Psychol.*, vol. 48, no. 2, pp. 246–268, 1936.

[31] X. Hu, K. Li, J. Han, X. Hua, L. Guo, and T. Liu, "Bridging the semantic gap via functional brain imaging," *IEEE Trans. Multimedia*, vol. 14, no. 2, pp. 314–325, 2012.

[32] C. E. Izard, *The face of emotion*. New York: Appleton-Century-Crofts, 1971.

[33] S. Ji, L. Tang, S. Yu, and J. Ye, "A shared-subspace learning framework for multi-label classification," *ACM Trans. Knowl. Discov. Data*, vol. 4, no. 2, pp. 8:1–8:29, 2010.

[34] J. Jin, *Chinese music*. Cambridge: Cambridge University Press, 2011.

[35] P. N. Juslin and J. A. Sloboda, *Music and Emotion: Theory and Research*. New York: Oxford University Press, 2001.

[36] P. N. Juslin and D. Västfjäll, "Emotional responses to music: The need to consider underlying mechanisms," *Behav. Brain Sci.*, vol. 31, pp. 559–575, October 2008.

[37] H. Katayose, M. Imai, and S. Inokuchi, "Sentiment extraction in music," in *Proc. 9th ICPR*, Nov 1988, pp. 1083–1087.

[38] Y. E. Kim, E. M. Schmidt, R. Migneco, B. G. Morton, P. Richardson, J. Scott, J. A. Speck, and D. Turnbull, "State of the art report: Music emotion recognition: A state of the art review," in *Proc. 11th ISMIR*, 2010, pp. 255–266.

[39] S. Koelsch, "Towards a neural basis of music-evoked emotions," *Trends in Cognitive Sciences*, vol. 14, no. 3, pp. 131–137, 2010.

[40] ——, "Brain correlates of music-evoked emotions," *Nature Reviews Neuroscience*, vol. 15, no. 3, pp. 170–180, 2014.

[41] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "Deap: A database for emotion analysis using physiological signals," *IEEE Trans. Affect. Comput.*, vol. 3, no. 1, pp. 18–31, 2012.

[42] E. Kroupi, A. Yazdani, and T. Ebrahimi, "EEG correlates of different emotional states elicited during watching music videos," in *Affective Computing and Intelligent Interaction*, ser. Lecture Notes in Computer Science, S. D'Mello, A. Graesser, B. Schuller, and J.-C. Martin, Eds., vol. 6975. Springer, 2011, pp. 457–466.

[43] C. L. Krumhansl, "An exploratory study of musical emotions and psychophysiology," *Can. J. Exp. Psychol.*, vol. 51, no. 4, pp. 336–353, 1997.

[44] L. D. Lathauwer, B. D. Moor, and J. Vandewalle, "A multilinear singular value decomposition," *SIAM J. Matrix Anal. Appl.*, vol. 21, no. 4, pp. 1253–1278, 2000.

[45] D. L. Levitin, "Psychology of Music," in *International Encyclopedia of Social Sciences*, W. A. Darity, Ed. Farmington Hills, MI: MacMillan Library Reference, 2007, pp. 345–346.

[46] T. Li and M. Ogihara, "Detecting emotion in music," in *Proc. 4th International Conference on Music Information Retrieval*, 2003.

[47] M. Liang, *Chinese Music of Today (in Chinese)*, 3rd ed. Shanghai Conservatory of Music Press, 2004.

[48] Y.-P. Lin, C.-H. Wang, T.-P. Jung, T.-L. Wu, S.-K. Jeng, J.-R. Duann, and J.-H. Chen, "EEG-based emotion recognition in music listening," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 7, pp. 1798–1806, 2010.

[49] Y.-P. Lin, Y.-H. Yang, and T.-P. Jung, "Fusion of electroencephalogram dynamics and musical contents for estimating emotional responses in music listening," *Frontiers in Neuroscience*, vol. 8, no. 94, 2014.

[50] D. Liu, L. Lu, and H.-J. Zhang, "Automatic mood detection from acoustic music data," in *Proc. 4th International Conference on Music Information Retrieval*, 2003, pp. 81–87.

[51] Y. Liu, Y. Liu, Y. Zhao, and K. A. Hua, "What strikes the strings of your heart?: Multi-label dimensionality reduction for music emotion analysis," in *Proc. 22nd ACM Multimedia*, 2014, pp. 1069–1072.

[52] C. Lord, *Aristotle The Politics. 1st edition*. Chicago: University of Chicago Press, 1984.

[53] H. Lu, K. Plataniotis, and A. Venetsanopoulos, "Mpca: Multilinear principal component analysis of tensor objects," *IEEE Trans. Neural Netw.*, vol. 19, no. 1, pp. 18–39, 2008.

[54] L. Lu, D. Liu, and H.-J. Zhang, "Automatic mood detection and tracking of music audio signals," *IEEE Trans. Audio, Speech, Language Process.*, vol. 14, no. 1, pp. 5–18, 2006.

[55] Q. Lu, X. Chen, D. Yang, and J. Wang, "Boosting for multimodal music emotion classification," in *Proc. 11th ISMIR*, 2010, pp. 105–110.

[56] G. Luck, P. Toiviainen, J. Erkkila, O. Lartillot, K. Riikkila, A. Makela, K. Pyhaluoto, H. Raine, L. Varkila, and J. Varri, "Modelling the relationships between emotional responses to, and musical content of, music therapy improvisations," *Psychol. Music*, vol. 36, no. 1, 2008.

[57] L. Meyer, *Emotion and Meaning in Music*. University of Chicago Press, 1956.

[58] Y. Panagakis, C. Kotropoulos, and G. R. Arce, "Sparse multi-label linear embedding within nonnegative tensor factorization applied to music tagging," in *Proc. 11th ISMIR*, 2010, pp. 393–398.

[59] C. H. Park and M. Lee, "On applying linear discriminant analysis for multi-labeled problems," *Pattern Recogn. Lett.*, vol. 29, no. 7, pp. 878–887, 2008.

[60] P. C. Petrantonakis and L. J. Hadjileontiadis, "A novel emotion elicitation index using frontal brain asymmetry for enhanced EEG-based emotion recognition," *IEEE Trans. Inform. Technol. Biomed.*, vol. 15, no. 5, pp. 737–746, 2011.

[61] R. C. Pian, *Sonq dynasty musical sources and their interpretation*. Harvard-Yenching Institute. Monograph Series, V. 16. Cambridge: Harvard University Press, 1967.

[62] R. Plutchik, "A general psychoevolutionary theory of emotion," in *Emotion: Theory, research, and experience: Vol. 1. Theories of emotion*, R. Plutchik and H. Kellerman, Eds. New York: Academic press, 1980, pp. 3–33.

[63] G.-J. Qi, X.-S. Hua, Y. Rui, J. Tang, T. Mei, and H.-J. Zhang, "Correlative multi-label video annotation," in *Proc. 15th ACM Multimedia*, 2007, pp. 17–26.

[64] G.-J. Qi, X.-S. Hua, Y. Rui, J. Tang, and H.-J. Zhang, "Two-dimensional multilabel active learning with an efficient online adaptation model for image classification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 10, pp. 1880–1897, 2009.

[65] P. J. Rentfrow and S. D. Gosling, "The do re mi's of everyday life: the structure and personality correlates of music preferences." *J. Pers. Soc. Psychol.*, vol. 84, no. 6, pp. 1236–1256, 2003.

[66] S. Rho, B.-j. Han, and E. Hwang, "Svr-based music mood classification and context-based music recommendation," in *Proc. 17th ACM Multimedia*, 2009, pp. 713–716.

[67] J. A. Russell, "A circumplex model of affect," *J. Pers. Soc. Psychol.*, vol. 39, pp. 1161–1178, 1980.

[68] L. A. Schmidt and L. J. Trainor, "Frontal brain electrical activity (EEG) distinguishes valence and intensity of musical emotions," *Cognition and Emotion*, vol. 15, no. 4, pp. 487–500, 2001.

[69] L. Sun, S. Ji, and J. Ye, "Hypergraph spectral learning for multi-label classification," in *Proc. 14th SIGKDD*, 2008, pp. 668–676.

[70] ——, *Multi-Label Dimensionality Reduction*, ser. Chapman & Hall/CRC Machine Learning & Pattern Recognition. Chapman and Hall/CRC, 2013.

[71] R. Thayer, *The biopsychology of mood and arousal*. Oxford University Press, USA, 1989.

[72] K. Trochidis and E. Bigand, "EEG-based emotion perception during music listening," in *Proc. ICMPC*, 2012, pp. 1018–1021.

[73] K. Trohidis, G. Tsoumakas, G. Kalliris, and I. Vlahavas, "Multi-label classification of music into emotions," in *Proc. 9th International Conference on Music Information Retrieval*, 2008, pp. 325–330.

[74] G. Tsoumakas, I. Katakis, and I. Vlahavas, "Mining multi-label data," in *Data Mining and Knowledge Discovery Handbook*, 2010, pp. 667–685.

[75] D. Turnbull, L. Barrington, D. Torres, and G. Lanckriet, "Semantic annotation and retrieval of music and sound effects," *IEEE Trans. Audio, Speech, Language Process.*, vol. 16, no. 2, pp. 467–476, 2008.

[76] M. Uščumlić, R. Chavarriaga, and J. d. R. Millán, "An iterative framework for EEG-based image search: Robust retrieval with weak classifiers," *PLoS ONE*, vol. 8, no. 8, p. e72018, 2013.

[77] B. L. van De, "Emotion Detection in Music, a Survey," in *6th Twente Student conference on IT*, 2006.

[78] E. Varèse, "The liberation of sound," *Perspect. New Music*, vol. 5, no. 1, pp. 11–19, 1966.

[79] T. I. Wallis, "Computer-Generating Emotional Music: The Design of an Affective Music Algorithm," 2008.

[80] C. Wang, S. Yan, L. Zhang, and H.-J. Zhang, "Multi-label sparse coding for automatic image annotation," *Proc. CVPR*, pp. 1643–1650, 2009.

[81] H. Wang, C. Ding, and H. Huang, "Multi-label linear discriminant analysis," in *Proc. 11th ECCV*, 2010, pp. 126–139.

[82] J.-C. Wang, Y.-H. Yang, H.-M. Wang, and S.-K. Jeng, "The acoustic emotion gaussians model for emotion-based music annotation and retrieval," in *Proc. 20th ACM Multimedia*, 2012, pp. 89–98.

[83] J. Wang, E. Pohlmeyer, B. Hanna, Y.-G. Jiang, P. Sajda, and S.-F. Chang, "Brain state decoding for rapid image retrieval," in *Proc. 17th ACM Multimedia*, 2009, pp. 945–954.

[84] X. Wang, D. Rosenblum, and Y. Wang, "Context-aware mobile music recommendation for daily activities," in *Proc. 20th ACM Multimedia*, 2012, pp. 99–108.

[85] N. M. Weinberger, "Neuroscience: Music and the brain," *Sci. Am.*, vol. 291, no. 5, pp. 88–95, 2004.

[86] A. Wieczorkowska, P. Synak, and Z. Raś, "Multi-Label Classification of Emotions in Music," in *Intelligent Information Processing and Web Mining*, ser. Advances in Intelligent and Soft Computing, M. Klopotek, S. Wierzchon, and K. Trojanowski, Eds. Springer Berlin/Heidelberg, 2006, vol. 35, pp. 307–315.

[87] H. Yang, *Contemporary Chinese Music*. Beijing: New Star Press, 2013.

[88] Y.-H. Yang and H. H. Chen, "Machine recognition of music emotion: A review," *ACM Trans. Intell. Syst. Technol.*, vol. 3, no. 3, pp. 40:1–40:30, May 2012.

[89] Y.-H. Yang, Y.-C. Lin, Y.-F. Su, and H. H. Chen, "A Regression Approach to Music Emotion Recognition," *IEEE Trans. Audio, Speech, Language Process.*, vol. 16, no. 2, pp. 448–457, 2008.

[90] Y.-H. Yang, C.-C. Liu, and H. H. Chen, "Music emotion classification: A fuzzy approach," in *Proc. 14th ACM Multimedia*, 2006, pp. 81–84.
[91] K. Yu, S. Yu, and V. Tresp, "Multi-label informed latent semantic indexing," in *Proc. 28th SIGIR*, 2005, pp. 258–265.
[92] M. Zentner, D. Grandjean, and K. R. Scherer, "Emotions Evoked by the Sound of Music: Characterization, Classification, and Measurement," *Emotion*, vol. 8, no. 4, pp. 494–521, 2008.
[93] M.-L. Zhang and Z.-H. Zhou, "Multilabel neural networks with applications to functional genomics and text categorization," *IEEE Trans. Knowl. Data Eng.*, vol. 18, no. 10, pp. 1338–1351, 2006.
[94] ——, "Ml-knn: A lazy learning approach to multi-label learning," *Pattern Recogn.*, vol. 40, no. 7, pp. 2038–2048, 2007.
[95] ——, "A review on multi-label learning algorithms," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 8, pp. 1819–1837, 2014.
[96] Y. Zhang and Z.-H. Zhou, "Multilabel dimensionality reduction via dependence maximization," *ACM Trans. Knowl. Discov. Data*, vol. 4, no. 3, pp. 14:1–14:21, 2010.
[97] D. Zhou, O. Bousquet, T. N. Lal, J. Weston, and B. Schölkopf, "Learning with local and global consistency," in *NIPS 16*, 2004.