

NOVEL FAVORITE MUSIC CLASSIFICATION USING EEG-BASED OPTIMAL AUDIO FEATURES SELECTED VIA KDLPPCA

Ryosuke Sawata, Takahiro Ogawa and Miki Haseyama

Graduate School of Information Science and Technology, Hokkaido University
N-14, W-9, Kita-ku, Sapporo, Hokkaido, 060-0814, Japan
Email: {sawata, ogawa}@imd.ist.hokudai.ac.jp, miki@ist.hokudai.ac.jp

ABSTRACT

This paper presents a novel method of favorite music classification using EEG-based optimal audio features. To select audio features related to user's music preference, our method utilizes a relationship between EEG features obtained from the user's EEG signals during listening to music and their corresponding audio features since EEG signals of human reflect his/her music preference. Specifically, cross-loadings, whose components denote the degree of the relationship, are calculated based on Kernel Discriminative Locality Preserving Canonical Correlation Analysis (KDLPPCA) which is newly derived in the proposed method. In contrast with standard CCA, KDLPPCA can consider (1) non-linear correlation, (2) class information and (3) local structures of input EEG and audio features, simultaneously. Therefore, KDLPPCA-based cross-loadings can reflect best correlation between the user's EEG and corresponding audio signals. Then an optimal set of audio features related to his/her music preference can be obtained by employing the cross-loadings as novel criteria for feature selection. Consequently, our method realizes favorite music classification successfully by using the EEG-based optimal audio features.

Index Terms— electroencephalogram (EEG), canonical correlation analysis (CCA), kernel method, music classification.

1. INTRODUCTION

In the field of music information retrieval, many recommendation methods have been researched in order to provide musical pieces desired by users [1–9]. Particularly, the recommendation methods based on affective phenomena of humans such as [7–9] have been researched in recent years since those of humans are closely related to individual music preference. Specifically, these methods model human feeling by using audio features to understand human emotions, and effective music recommendation is realized by favorite music classification based on the models. However, audio features, which are not related to user's music preference, used in such method may deteriorate the performance of favorite music classification. Furthermore, it is difficult to find a set of audio features related to each user's music preference adaptively since their music preference is diverse [10, 11]. Therefore, it is necessary to introduce a new idea such as solving the problem by using alternative features.

Recently, it has become obvious that electroencephalogram (EEG) features extracted from EEG signals of a human who listens to music have a close relationship to his/her music preference [12–18]. For example, Hadjidimitriou *et al.* reported discrimination between user's EEG signals depending on music preference, *i.e.*, whether the user listened to favorite or unfavorable musical

pieces [17, 18]. Furthermore, observation of EEG signals has become easier, and the quality of observed signals has become better due to the development and miniaturization of an EEG sensor. For instance, the *music inspiration from your subconsciousness (mico)*¹, which is a headphone with an embedded EEG sensor, produced by *neurowear*², can easily provide EEG signals from the user when the headphone is used like standard ones. Therefore, we consider that EEG features obtained from the user's EEG signals during listening to music is utilizable and important to find a set of audio features related to user's music preference. However, so far, such research has not been studied adequately as far as we know.

Motivated by the aforementioned discussion, we have proposed favorite music classification using audio features selected by monitoring the relationship between EEG and audio features [19]. Specifically, we employed Canonical Correlation Analysis (CCA) [20] to obtain the relationship between EEG features of a user during listening to music and corresponding audio features. However, CCA may be insufficient for extraction of the relationship since CCA extracts only linear correlation, *i.e.*, the relationship may contain not only linear but also non-linear correlation. In fact, Li *et al.* reported that extracting non-linear correlation between EEG and other features is important to analyze a relationship between brain states and assigned tasks accurately [21]. Furthermore, although training features generally have class labels, CCA cannot consider those since CCA is an unsupervised method.

In this paper, we propose a novel method of favorite music classification using EEG-based optimal audio features selected via novel CCA. Our method firstly obtains EEG and audio features, and next selects a set of audio features related to user's music preference. To select the audio features, we utilize a relationship between EEG features obtained from the user's EEG signals during listening to favorite/unfavorite music and their corresponding audio features. Specifically, cross-loadings [22], whose components denote the degree of the relationship, are calculated based on Kernel Discriminative Locality Preserving CCA (KDLPPCA) which is newly derived in the proposed method. In contrast with standard CCA, KDLPPCA can consider non-linear correlation by using the kernel method, class information denoting whether input training samples are obtained from favorite or unfavorable music and local structures of input data. Thus, an optimal selection of audio features which are suitable for individual music preference can be expected by employing the KDLPPCA-based cross-loadings as criteria. Consequently, we train a Support Vector Machine (SVM) [23] classifier using the selected audio features, and successful classification of the user's favorite musical pieces becomes feasible.

This work was partly supported by Grant-in-Aid for Scientific Research (B) 25280036 and Grant-in-Aid for Challenging Exploratory Research 15K12023, Japan Society for the Promotion of Science (JSPS).

¹http://www.neurowear.com/projects_detail/mico.html

²<http://www.neurowear.com>

Table 1. EEG features calculated by the proposed method: C denotes the number of channels of EEG signals and C_p represents the number of symmetric electrode pairs placed on the scalp.

DESCRIPTION		DIMENSION	
Zero Crossing Rate		C	
Content Percentage of The Power Spectrum	θ wave (4-7 Hz)	C	
	slow- α wave (7-10 Hz)	C	
	fast- α wave (10-13 Hz)	C	
	α wave (7-13 Hz)	C	
	slow- β wave (13-19 Hz)	C	
	fast- β wave (19-30 Hz)	C	
	β wave (13-30 Hz)	C	
	γ wave (30-49 Hz)	C	
	Power Spectrum of The Hemispheric Asymmetry [15]	θ wave (4-7 Hz)	$2C_p$
		slow- α wave (7-10 Hz)	$2C_p$
fast- α wave (10-13 Hz)		$2C_p$	
α wave (7-13 Hz)		$2C_p$	
slow- β wave (13-19 Hz)		$2C_p$	
fast- β wave (19-30 Hz)		$2C_p$	
β wave (13-30 Hz)		$2C_p$	
γ wave (30-49 Hz)		$2C_p$	
Mean Frequency		C	
Mean Frequency of γ wave		C	
Spectral Entropy		C	
Spectral Entropy of γ wave		C	
TOTAL		$13C + 16C_p$	

2. FAVORITE MUSIC CLASSIFICATION USING EEG-BASED OPTIMAL AUDIO FEATURES

In this section, we explain favorite music classification using EEG-based optimal audio features. In our method, we first extract user's EEG features from EEG signals recorded while the user listens to musical pieces and audio features from corresponding musical pieces. Then we apply our novel CCA, *i.e.*, KDLPPCA to obtained EEG and audio features in order to extract the relationship between them. Next, we select a set of audio features related to user's music preference by monitoring KDLPPCA-based novel criteria, called cross-loadings [22], whose components denotes the degree of relationship between canonical variates of EEG and each original audio feature. KDLPPCA has three superiority to CCA: 1) the kernel method is employed to solve the linearly inseparable problem; 2) more discriminative variates can be extracted by considering the class information, *i.e.*, whether each feature vector is obtained from favorite or unfavorable music; 3) consideration of local structures is employed so that KDLPPCA can capture the intrinsic data structures of EEG and audio, globally. Thus, KDLPPCA-based cross-loadings are expected to be effective criteria for selection of audio features that are suitable for the user's music preference. Then favorite music classification based on SVM [23] using EEG-based optimal audio features is conducted. Note that our method does not need to use SVM invariably. In this paper, we experimentally employ SVM as a classifier since it is broadly used in some studies.

2.1. EEG-based Audio Feature Selection via KDLPPCA

In this subsection, we describe an EEG-based selection of audio features related to user's music preference using KDLPPCA-based cross-loadings.

First, we obtain two sets of N feature vectors $\mathbf{X}^E = [\mathbf{x}_1^E, \mathbf{x}_2^E, \dots, \mathbf{x}_N^E] \in \mathbb{R}^{d_E \times N}$ and $\mathbf{X}^A = [\mathbf{x}_1^A, \mathbf{x}_2^A, \dots, \mathbf{x}_N^A] \in \mathbb{R}^{d_A \times N}$ from EEG and audio signals, where d_E and d_A are the dimensions of the EEG and the audio feature vector, respectively. Note that the sets of EEG and audio features, used in this paper, were acquired in an authors' previous work [24] and summarized in Tables 1 and 2. Since the space of the paper is limited, the details of calculation of EEG and audio feature are given there. Each obtained feature vector has a label representing whether it is obtained from "Favorite" or "Unfavorite" musical pieces.

Table 2. Audio features used in the proposed method.

CATEGORY	DESCRIPTION	STATISTICS	DIMENSION
dynamics	Root Means Square	Mean, Std	2
spectral	Centroid	Mean, Std	2
	Brightness	Mean, Std	2
	Spread	Mean, Std	2
	Skewness	Mean, Std	2
	Kurtosis	Mean, Std	2
	Rolloff	Mean, Std	4
	Entropy	Mean, Std	2
	Flatness	Mean, Std	2
	Roughness	Mean, Std	2
	Irregularity	Mean, Std	2
timbre	Zero Crossing Rate	Mean, Std	2
	MFCC	Mean, Std	26
tonal	Low Energy	Mean, Std	2
	Key Strength	Mean, Std	48
	Chromagram	Mean, Std	24
	Key	Mean, Std	2
	Tonal Centroid	Mean, Std	12
rhythm	Mode	Mean, Std	2
	Tempo	Mean	1
	Pulse Clarity	Mean, Std	2
	Event Density	Mean, Std	2
	Attack Time	Mean, Std	2
	Attack Slope	Mean, Std	2
TOTAL			151

2.1.1. Applying KDLPPCA

We firstly apply non-linear mapping to obtained feature vectors. Note that the dimension of audio feature vector needs to be preserved in order to apply the following feature selection. Hence, feature vectors of only EEG set, *i.e.*, \mathbf{x}_j^E ($j = 1, 2, \dots, N$), are transformed into Hilbert space via non-linear maps $\phi_E : \mathbf{x}^E \mapsto \phi_E(\mathbf{x}^E) \in \mathbb{R}^{d_{\phi_E}}$. From the aforementioned mapped results, we obtain $\Phi^E = [\phi_E(\mathbf{x}_1^E), \phi_E(\mathbf{x}_2^E), \dots, \phi_E(\mathbf{x}_N^E)] \in \mathbb{R}^{d_{\phi_E} \times N}$. Then we apply Discriminative Locality Preserving CCA (DLPPCA) [25] to \mathbf{X}^A and Φ^E . Specifically, we seek to maximize the following objective function by using the projections $\mathbf{u}_\phi^E \in \mathbb{R}^{d_{\phi_E}}$ and $\mathbf{u}^A \in \mathbb{R}^{d_A}$:

$$(\hat{\mathbf{u}}_\phi^E, \hat{\mathbf{u}}^A) = \arg \max_{\mathbf{u}_\phi^E, \mathbf{u}^A} \frac{\mathbf{u}_\phi^{E\top} (\mathbf{C}_w - \zeta \mathbf{C}_b) \mathbf{u}^A}{\sqrt{\mathbf{u}_\phi^{E\top} \mathbf{C}_\phi^E \mathbf{u}_\phi^E} \sqrt{\mathbf{u}^{A\top} \mathbf{C}^A \mathbf{u}^A}}, \quad (1)$$

where ζ is a tunable parameter and

$$\mathbf{C}_w = \Phi^E \mathbf{H} (\mathbf{S}_\phi^E \circ \mathbf{S}^A) \mathbf{H} \mathbf{X}^A \mathbf{X}^A \mathbf{T}, \quad \mathbf{C}_b = \Phi^E \mathbf{H} (\bar{\mathbf{S}}_\phi^E \circ \bar{\mathbf{S}}^A) \mathbf{H} \mathbf{X}^A \mathbf{X}^A \mathbf{T}, \quad (2)$$

$$\mathbf{C}_\phi^E = \Phi^E \mathbf{H} (\mathbf{L}_\phi^E + \bar{\mathbf{L}}_\phi^E) \mathbf{H} \Phi^{E\top}, \quad \mathbf{C}^A = \mathbf{X}^A \mathbf{H} (\mathbf{L}^A + \bar{\mathbf{L}}^A) \mathbf{H} \mathbf{X}^A \mathbf{T}, \quad (3)$$

where the symbol \circ denotes the Hadamard product. Furthermore, $\mathbf{H} = \mathbf{I} - \frac{1}{N} \mathbf{1}\mathbf{1}^\top$ is a centering matrix, $\mathbf{I} \in \mathbb{R}^{N \times N}$ is an $N \times N$ identity matrix, and $\mathbf{1} = [1, \dots, 1]^\top \in \mathbb{R}^N$ is an N -dimensional vector. In Eq. (1), $\mathbf{C}_w \in \mathbb{R}^{d_{\phi_E} \times d_A}$ and $\mathbf{C}_b \in \mathbb{R}^{d_{\phi_E} \times d_A}$, defined as shown in Eq. (2), are respectively the within-class covariance matrix and the between-class covariance matrix, which they are derived by considering the class similarity matrices $\mathbf{S}_\phi^E = \{S_{ij}^{\phi_E}\}_{i,j=1}^N$, $\bar{\mathbf{S}}_\phi^E = \{\bar{S}_{ij}^{\phi_E}\}_{i,j=1}^N$, $\mathbf{S}^A = \{S_{ij}^A\}_{i,j=1}^N$ and $\bar{\mathbf{S}}^A = \{\bar{S}_{ij}^A\}_{i,j=1}^N$ defined as follows:

$$S_{ij}^{\phi_E} = \begin{cases} \exp(-\delta_{ij}^{\phi_E} / t_{\phi_E}), & \text{label}(\phi_E(\mathbf{x}_i^E)) = \text{label}(\phi_E(\mathbf{x}_j^E)) \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

$$\bar{S}_{ij}^{\phi_E} = \begin{cases} \exp(-\delta_{ij}^{\phi_E} / t_{\phi_E}), & \text{label}(\phi_E(\mathbf{x}_i^E)) \neq \text{label}(\phi_E(\mathbf{x}_j^E)) \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

$$S_{ij}^A = \begin{cases} \exp(-\delta_{ij}^A / t_A), & \text{label}(\mathbf{x}_i^A) = \text{label}(\mathbf{x}_j^A) \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

$$\bar{S}_{ij}^A = \begin{cases} \exp(-\delta_{ij}^A / t_A), & \text{label}(\mathbf{x}_i^A) \neq \text{label}(\mathbf{x}_j^A) \\ 0, & \text{otherwise} \end{cases} \quad (7)$$

where $\delta_{ij}^{\phi_E} = \|\phi_E(\mathbf{x}_i^E) - \phi_E(\mathbf{x}_j^E)\|^2$, $t_{\phi_E} = \frac{1}{N(N-1)} \sum_{i=1}^N \sum_{j=1}^N \delta_{ij}^{\phi_E}$, $\delta_{ij}^A = \|\mathbf{x}_i^A - \mathbf{x}_j^A\|^2$ and $t_A = \frac{1}{N(N-1)} \sum_{i=1}^N \sum_{j=1}^N \delta_{ij}^A$. Furthermore, "label(\bullet)" is

the \bullet 's class label. By using these class similarity matrices, we can also derive Laplacian matrices $\mathbf{L}_\phi^E \in \mathbb{R}^{N \times N}$, $\bar{\mathbf{L}}_\phi^E \in \mathbb{R}^{N \times N}$, $\mathbf{L}^A \in \mathbb{R}^{N \times N}$ and $\bar{\mathbf{L}}^A \in \mathbb{R}^{N \times N}$ defined in Eq. (3) as follows:

$$\mathbf{L}_\phi^E = \mathbf{D}_\phi^E - \mathbf{S}_\phi^E \circ \mathbf{S}_\phi^E, \quad \bar{\mathbf{L}}_\phi^E = \bar{\mathbf{D}}_\phi^E - \bar{\mathbf{S}}_\phi^E \circ \bar{\mathbf{S}}_\phi^E, \quad (8)$$

$$\mathbf{L}^A = \mathbf{D}^A - \mathbf{S}^A \circ \mathbf{S}^A, \quad \bar{\mathbf{L}}^A = \bar{\mathbf{D}}^A - \bar{\mathbf{S}}^A \circ \bar{\mathbf{S}}^A, \quad (9)$$

where $\mathbf{D}_\phi^E = \text{diag}[\sum_i (S_{1i}^{\phi_E})^2, \sum_i (S_{2i}^{\phi_E})^2, \dots, \sum_i (S_{Ni}^{\phi_E})^2]$, $\bar{\mathbf{D}}_\phi^E = \text{diag}[\sum_i (\bar{S}_{1i}^{\phi_E})^2, \sum_i (\bar{S}_{2i}^{\phi_E})^2, \dots, \sum_i (\bar{S}_{Ni}^{\phi_E})^2]$, $\mathbf{D}^A = \text{diag}[\sum_i (S_{1i}^A)^2, \sum_i (S_{2i}^A)^2, \dots, \sum_i (S_{Ni}^A)^2]$ and $\bar{\mathbf{D}}^A = \text{diag}[\sum_i (\bar{S}_{1i}^A)^2, \sum_i (\bar{S}_{2i}^A)^2, \dots, \sum_i (\bar{S}_{Ni}^A)^2]$. Solving Eq. (1) corresponds to the maximization of the correlation considering the class information and the local structures of input training sets of EEG and audio features.

However, finding \mathbf{u}_ϕ^E in Eq. (1) is generally difficult since the dimension of Hilbert space's vector, *i.e.*, d_{ϕ_E} is very high. Therefore, we employ the kernel trick which is denoted as the kernel function $k^E(\mathbf{x}_i^E, \mathbf{x}_j^E) = \phi_E(\mathbf{x}_i^E)^T \phi_E(\mathbf{x}_j^E)$, and we can thus rewrite the objective function of Eq. (1) by assuming $\mathbf{u}_\phi^E = \Phi^E \mathbf{H} \alpha$ based on dual representation [26, 27] as follows:

$$\begin{aligned} & \frac{\mathbf{u}_\phi^{E^T} (\mathbf{C}_w - \zeta \mathbf{C}_b) \mathbf{u}^A}{\sqrt{\mathbf{u}_\phi^{E^T} \mathbf{C}_\phi^E \mathbf{u}_\phi^E} \sqrt{\mathbf{u}^A \mathbf{C}^A \mathbf{u}^A}} \\ &= \frac{\mathbf{u}_\phi^{E^T} \{ \Phi^E (\mathbf{S}_\phi^E \circ \mathbf{S}^A) \mathbf{X}^A \mathbf{T} - \zeta \{ \Phi^E (\bar{\mathbf{S}}_\phi^E \circ \bar{\mathbf{S}}^A) \mathbf{X}^A \mathbf{T} \} \} \mathbf{u}^A}{\sqrt{\alpha^T \tilde{\mathbf{C}}_\phi^E \alpha} \sqrt{\mathbf{u}^A \mathbf{C}^A \mathbf{u}^A}} \\ &= \frac{\alpha^T \mathbf{K}^E \{ (\mathbf{S}_\phi^E \circ \mathbf{S}^A) - \zeta (\bar{\mathbf{S}}_\phi^E \circ \bar{\mathbf{S}}^A) \} \mathbf{X}^A \mathbf{T} \mathbf{u}^A}{\sqrt{\alpha^T \tilde{\mathbf{C}}_\phi^E \alpha} \sqrt{\mathbf{u}^A \mathbf{C}^A \mathbf{u}^A}}, \end{aligned} \quad (10)$$

where

$$\tilde{\mathbf{C}}_\phi^E = \mathbf{H} \mathbf{K}^E \mathbf{H} (\mathbf{L}_\phi^E + \bar{\mathbf{L}}_\phi^E) \mathbf{H} \mathbf{K}^E \mathbf{H} + \xi \mathbf{H} \mathbf{K}^E \mathbf{H}, \quad (11)$$

and $\alpha \in \mathbb{R}^N$ is a coefficient vector. Furthermore, $\mathbf{K}^E = \Phi^{E^T} \Phi^E \in \mathbb{R}^{N \times N}$ is gram matrix whose (i, j) th component is $k^E(\mathbf{x}_i^E, \mathbf{x}_j^E)$, and ξ is a regularization parameter. Then $\delta_{ij}^{\phi_E}$ in Eqs. (4) and (5) becomes calculable as $\delta_{ij}^{\phi_E} = \|\phi_E(\mathbf{x}_i^E) - \phi_E(\mathbf{x}_j^E)\|^2 = (\mathbf{K}^E)_{ii} - 2(\mathbf{K}^E)_{ij} + (\mathbf{K}^E)_{jj}$, where $(\mathbf{K}^E)_{ij}$ denotes the (i, j) th component of \mathbf{K}^E . Consequently, we can obtain α and \mathbf{u}^A as the optimal solutions by solving the following Lagrange function:

$$\begin{aligned} \mathcal{L}(\alpha, \mathbf{u}^A) &= \alpha^T \mathbf{K}^E \{ (\mathbf{S}_\phi^E \circ \mathbf{S}^A) - \zeta (\bar{\mathbf{S}}_\phi^E \circ \bar{\mathbf{S}}^A) \} \mathbf{X}^A \mathbf{T} \mathbf{u}^A \\ &\quad - \frac{\lambda_E}{2} (\alpha^T \tilde{\mathbf{C}}_\phi^E \alpha - 1) - \frac{\lambda_A}{2} (\mathbf{u}^A \mathbf{C}^A \mathbf{u}^A - 1), \end{aligned} \quad (12)$$

where $\lambda_E = \lambda_A$, and they are equivalent to the optimal solution of Eq. (1).

In this way, we can obtain α and \mathbf{u}^A which can optimally reflect the user's music preference with preserving the local structures and considering the class information in a subspace by solving Eq. (12).

2.1.2. Feature Selection Using KDLPCA-based cross-loadings

In order to select audio features considering the relationship between EEG and audio features, class information and local structures of those features, we newly derive KDLPCA-based cross-loadings $\mathbf{r} \in \mathbb{R}^{d_A}$ based on Eq. (10) defined as follows:

$$\mathbf{r} = \mathbf{X}^A \{ (\mathbf{S}_\phi^E \circ \mathbf{S}^A) - \zeta (\bar{\mathbf{S}}_\phi^E \circ \bar{\mathbf{S}}^A) \} \mathbf{K}^E \alpha. \quad (13)$$

Cross-loadings can measure how well a feature loads on the canonical variate from the opposing feature set instead of its own. Specifically, an i th component r_i of \mathbf{r} denotes the degree of relevance between KDLPCA-based canonical variates of EEG and corresponding i th original audio feature. Thus, we employ these cross-loadings

Table 3. The details of music dataset used in the experiment.

	Number of FAVORITE music	Number of UNFAVORITE music	TOTAL
Subject A	23	23	46
Subject B	23	23	46
Subject C	17	17	34
Subject D	20	20	40
Subject E	16	16	32

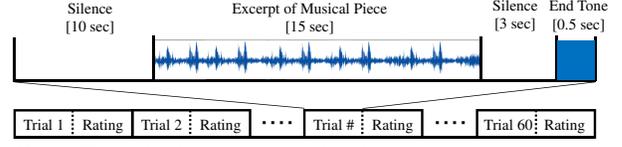


Fig. 1. The task which we required each subject in the experiment.

as novel criteria to select audio features related to the user's music preference. Specifically, we select audio features according to the value of the cross-loadings in descending order.

In this way, selection of audio features which are optimally suitable for individual music preference is realized since the KDLPCA-based canonical variates used for calculation of cross-loadings reflect the best correlation between the user's EEG and audio features with the class information and the local structures as explained in the previous subsection.

2.2. Favorite Music Classification

In this subsection, we explain favorite music classification using the EEG-based optimal audio features explained in the previous subsection. In our method, we classify an unknown label of an EEG-based optimal audio feature vector $\mathbf{x}^{A_{test}}$ by using the following hyperplane:

$$f(\mathbf{x}^{A_{test}}) = \text{sgn}(\langle \mathbf{w}, \mathbf{x}^{A_{test}} \rangle + b). \quad (14)$$

Given training music database $\mathcal{D} = \{(\mathbf{x}_l^{A_{train}}, y_l)\}_{l=1}^{N_{train}}$, \mathbf{w} and b can be obtained by solving the SVM formulation [23] as follows:

$$\min_{\mathbf{w}} \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{l=1}^{N_{train}} \eta_l, \quad (15)$$

$$\text{s.t. } y_l (\langle \mathbf{w}, \mathbf{x}_l^{A_{train}} \rangle + b) + \eta_l \geq 1, \eta_l \geq 0, \forall (\mathbf{x}_l^{A_{train}}, y_l) \in \mathcal{D},$$

where $\mathbf{x}_l^{A_{train}}$ is obtained from the l th piece of music database, η_l is a slack variable, and $y_l \in \{+1, -1\}$ denotes a correct class label, *i.e.*, "Favorite" or "Unfavorite".

In this way, we realize effective classification of favorite musical pieces for each user adaptively since audio features used for SVM are selected based on our novel criteria which are optimally suitable for his/her music preference.

3. EXPERIMENTAL RESULTS

In this section, we show experimental results to verify the effectiveness of the proposed method. First, we explain music dataset for observation of EEG signals. In our experiment, we used 60 musical pieces³, each of which has 15 sec length. Furthermore, all subjects rated each musical piece by a value of 5 levels, *i.e.*, 5 (very favorite), 4 (favorite), 3 (undecided), 2 (unfavorite) and 1 (unfavorite at all). Therefore, audio feature vectors could be grouped with respect to two classes, *i.e.*, "Favorite" and "Unfavorite". Class "Favorite" consisted of the audio feature vectors corresponding to the musical pieces rated 5 or 4 by a subject. On the other hand, class "Unfavorite" consisted of the audio feature vectors corresponding to the

³The database has 12 genres: Pops, Rock, Jazz, Latin, Classic, March, etc. We use 5 musical pieces for each genre, *i.e.*, the total number of musical pieces used in our experiment is 60.

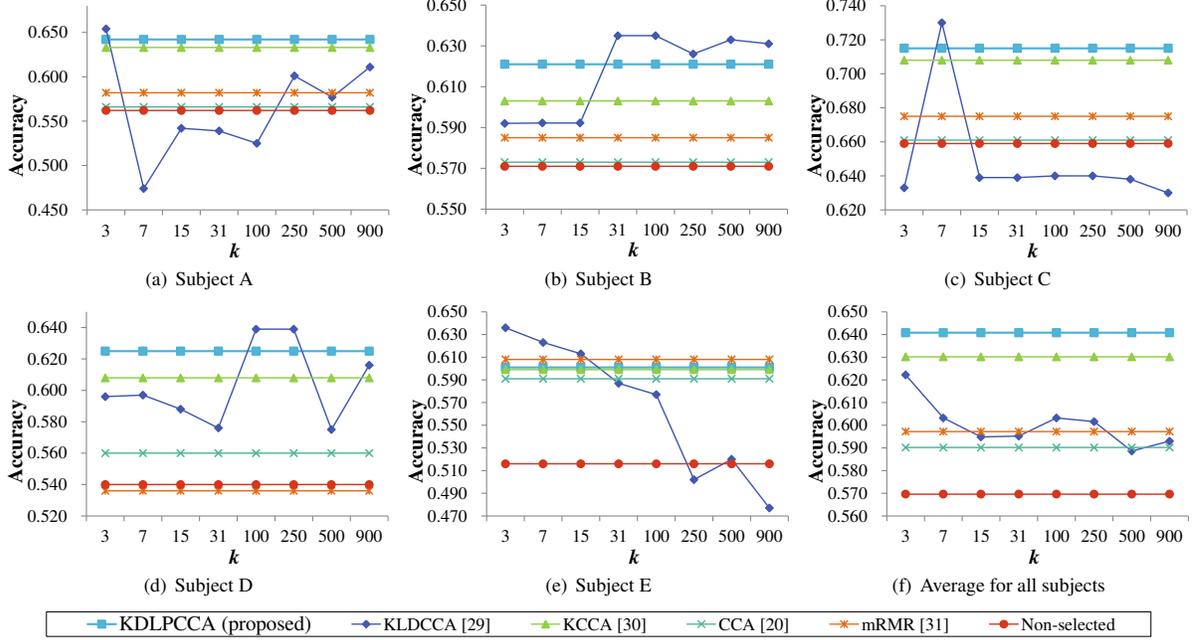


Fig. 2. Results of favorite music classification: The horizontal axis denotes the number of nearest neighbors ($= k$) for considering class labels. Meanwhile, the vertical axis denotes Accuracy. Note that the methods except for KLDCCA do not need to set the parameter k , *i.e.*, not depend on k . The methods except Non-selected varied the dimension of audio feature selection and presented the best results regarding each subject.

musical pieces rated 2 or 1 by a subject. Note that we did not use the musical pieces rated 3. Furthermore, we equalized the number of musical pieces of each class per subject by excluding some musical pieces randomly to prevent imbalance problem [28–30]. The details of the aforementioned music dataset are summarized in Table 3, and the task implemented for each subject in our experiment is shown in Fig. 1. These conditions were adopted in related studies [17, 18].

Next, we describe how to collect EEG signals in our experiment. EEG signals were collected from five healthy subjects, and their average age was about 23 years. We recorded EEG signals from 12 channels (Fp1, Fp2, F7, F8, C3, C4, P3, P4, O1, O2, T3 and T4) decided based on [17, 31] according to the international 10-20 system. Since EEG signals are weak, we amplified the signals by utilizing an amplifier (MEG-6116M, NIHON KOHDEN). All leads were referenced to linked earlobes, and a ground electrode was located on the forehead. We also applied a band-pass filter to recorded EEG signals in order to avoid artifacts, and we set the filter bandwidth to 0.04-100 Hz. The subjects were instructed to keep their eyes closed and to relax and remain seated during listening to music.

Furthermore, we explain the experimental conditions. In our experiment, we adopted the Gaussian kernel $k^E(\mathbf{x}_i^E, \mathbf{x}_j^E) = \exp^{-\|\mathbf{x}_i^E - \mathbf{x}_j^E\|^2 / 2\sigma^2}$ for KDLPPCA, where the kernel width σ^2 was chosen by searching the following parameter space: $\sigma^2 \in [2^{-15}, 2^{-12}, 2^{-9}, 2^{-6}, 2^{-3}, 2^0, 2^3]$. In addition, we respectively searched the following spaces to obtain the optimal parameters ζ in Eq. (10) and ξ in Eq. (11): $\zeta \in [0.01, 0.2, 0.5, 1.0]$ and $\xi \in [10^{-3}, 10^{-2}, 10^{-1}, 10^0]$. Moreover, we divided all of the training sets of audio feature vectors per musical piece in order to prevent overfitting caused by learning similar vectors extracted from the same musical piece. Thus, we classified the audio feature vectors per each musical piece, and calculated classification accuracy.

Results of our experiment are shown in Fig. 2. In this figure, we

also show the results of five comparative methods: methods using audio features selected via (1) Kernel Local Discrimination CCA (KLDCCA) [32], (2) Kernel CCA (KCCA) [33], (3) CCA, (4) the Max-Relevance and Min-Redundancy (mRMR) algorithm [34], and (5) method using non-selected audio features. This is because CCA and KCCA are benchmarks, and KLDCCA which can consider non-linear correlation and k -nearest neighbors' class information of input data is state-of-the-art CCA. Furthermore, (4) the mRMR algorithm is well-known and widely used non-CCA-based feature selection algorithm. Note that we adopted linear kernels for all SVMs experimentally, and parameters used in SVMs were determined via Grid Search [35]. As shown in Figs. 2(a)-(e), it is confirmed that KLDCCA-based comparative method outperforms the other methods, but it severely depends on the parameter k . Generally, it is difficult for KLDCCA to decide the optimal value of k according to each subject adaptively since the optimal k is different per each subject as shown in Figs. 2(a)-(e). Meanwhile, our method, *i.e.*, KDLPPCA-based method does not depend on k . Nonetheless, the performance of our method is relatively high as shown in Figs. 2(a)-(e). In addition, our method's average accuracy for all subjects outperforms all other methods as shown in Fig. 2(f). Therefore, our method realizes robust favorite music classification successfully.

4. CONCLUSIONS

In this paper, we proposed novel favorite music classification using EEG-based optimal audio features selected via KDLPPCA. KDLPPCA realizes the extraction of non-linear correlation with preserving the local structures of input EEG and audio features and considering the class labels. Hence, we can select an optimal set of audio features which is suitable for individual music preference by employing the KDLPPCA-based cross-loadings. Consequently, successful favorite music classification is realized by using the selected audio features.

5. REFERENCES

- [1] P. Åman and L. A. Liikkanen, "A survey of music recommendation aids," in *Proc. of Workshop on Music Recommendation And Discovery (WOMRAD)*, Sept. 2010.
- [2] Y. Song, S. Dixon, and M. Pearce, "A survey of music recommendation systems and future perspectives," in *Proc. of the 9th International Symposium on Computer Music Modelling and Retrieval (CMMR)*, 2012.
- [3] Z. Fu, G. Lu, K. M. Ting, and D. Zhang, "A survey of audio-based music classification and annotation," *IEEE Trans. on Multimedia*, vol. 13, no. 2, pp. 303–319, April 2011.
- [4] J. Shen, J. Shepherd, and A. H. H. Ngu, "Towards effective content-based music retrieval with multiple acoustic feature combination," *IEEE Trans. on Multimedia*, vol. 8, no. 6, pp. 1179–1189, Dec. 2006.
- [5] Z. Fu, G. Lu, K. M. Ting, and D. Zhang, "A survey of audio-based music classification and annotation," *IEEE Trans. on Multimedia*, vol. 13, no. 2, pp. 303–319, April 2011.
- [6] M. Korhonen, D. Clausi, and M. Jernigan, "Modeling emotional content of music using system identification," *IEEE Trans. on Systems, Man, and Cybernetics*, vol. 36, no. 3, pp. 588–599, June 2005.
- [7] K. Yoon, J. Lee, and M.-U. Kim, "Music recommendation system using emotion triggering low-level features," *IEEE Trans. on Consumer Electronics*, vol. 58, no. 2, pp. 612–618, May 2012.
- [8] J. Deng and C. Leung, "Emotion-based music recommendation using audio features and user playlist," in *Proc. of the 6th International Conference on New Trends in Information Science and Service Science and Data Mining (ISSDM)*, 2012, pp. 796–801.
- [9] Y. H. Chin, S. H. Lin, C. H. Lin, E. Siahaan, A. Frisky, and J. C. Wang, "Emotion profile-based music recommendation," in *Proc. of the 7th International Conference on Ubi-Media Computing and Workshops (UMEDIA)*, 2014, pp. 111–114.
- [10] X. Hu and J. S. Downie, "Exploring mood metadata: Relationships with genre, artist and usage metadata," in *Proc. of the 8th International Conference on Music Information Retrieval*, 2007, pp. 67–72.
- [11] D. Levitin, *This is your brain on music: the science of a human obsession*, Dutton Penguin, 2006.
- [12] R. Kadir, M. Ghazali, Z. Murat, M. Taib, H. Rahman, and S. Aris, "The preliminary study on the effect of nasyid music and rock music on brainwave signal using EEG," in *Proc. of the 2nd International Congress on Engineering Education (ICEED)*, 2010, pp. 58–63.
- [13] M. Sreedevi, A. Ajesh, R. Ajithnath, and L. Binu, "A study of effect of music pitch variation in EEG using factor analysis and neural networks," in *Proc. of the 2nd International Conference on Biomedical Engineering and Informatics (BMEI)*, 2009, pp. 1–3.
- [14] M. Naji, M. Firoozabadi, and P. Azadfallah, "Emotion classification based on forehead biosignals using support vector machines in music listening," in *Proc. of the IEEE 12th International Conference on Bioinformatics Bioengineering (BIBE)*, Nov. 2012, pp. 396–400.
- [15] 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. on Biomedical Engineering*, vol. 57, no. 7, pp. 1798–1806, 2010.
- [16] J.-L. Hsu, Y.-L. Zhen, T.-C. Lin, and Y.-S. Chiu, "Personalized music emotion recognition using electroencephalography (EEG)," in *Proc. of IEEE International Symposium on Multimedia (ISM)*, Dec. 2014, pp. 277–278.
- [17] S. Hadjidimitriou and L. Hadjileontiadis, "Toward an EEG-based recognition of music liking using time-frequency analysis," *IEEE Trans. on Biomedical Engineering*, vol. 59, no. 12, pp. 3498–3510, Dec. 2012.
- [18] S. Hadjidimitriou and L. Hadjileontiadis, "EEG-based classification of music appraisal responses using time-frequency analysis and familiarity ratings," *IEEE Trans. on Affective Computing*, vol. 4, no. 2, pp. 161–172, April 2013.
- [19] R. Sawata, T. Ogawa, and M. Haseyama, "Music recommendation via eeg-based individual preference extraction," in *Proc. of International Workshop on Advanced Image Technology (WAIT)*, 2015.
- [20] H. Hotelling, "Relations between two sets of variates," *Biometrika*, vol. 28, no. 3, pp. 321–377, 1936.
- [21] Y. Li, S. Liu, N. Xi, and X. Zhang, "A study of the relationship between brain states and skill level of teleoperator," in *Proc. of IEEE International Conference on Robotics and Biomimetics (ROBIO)*, 2013, pp. 390–395.
- [22] J. F. Hair, R. L. Tatham, R. E. Anderson, and W. Black, *Multivariate Data Analysis*, 5th edition, Prentice Hall, March 1998.
- [23] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, Sept. 1995.
- [24] R. Sawata, T. Ogawa, and M. Haseyama, "Human-centered favorite music estimation: EEG-based extraction of audio features reflecting individual preference," in *Proc. of IEEE International Conference on Digital Signal Processing (DSP)*, 2015, pp. 818–822.
- [25] X. Zhang, N. Guan, Z. Luo, and L. Lan, "Discriminative locality preserving canonical correlation analysis," in *Pattern Recognition*, vol. 321, pp. 341–349, Communications in Computer and Information Science, Springer Berlin Heidelberg, 2012.
- [26] T. Melzer, M. Reiter, and H. Bischof, "Appearance models based on kernel canonical correlation analysis," *Pattern Recognition*, vol. 36, no. 9, pp. 1961–1971, 2003.
- [27] B. Schölkopf, A. Smola, E. Smola, and K.-R. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Computation*, vol. 10, pp. 1299–1319, 1998.
- [28] K. Veropoulos, C. Campbell, and N. Cristianini, "Controlling the sensitivity of support vector machines," in *Proc. of the International Joint Conference on AI*, 1999, pp. 55–60.
- [29] G. Wu and E. Y. Chang, "Adaptive Feature-Space Conformal Transformation for Imbalanced Data Learning," in *Proc. of the 12th International Conference on Machine Learning*, 2003, pp. 816–823.
- [30] R. Akbani, S. Kwek, and N. Japkowicz, "Applying support vector machines to imbalanced datasets," in *Proc. of the 15th European Conference on Machine Learning (ECML)*, 2004, pp. 39–50.
- [31] T. Kawakami, T. Ogawa, and M. Haseyama, "Vocal segment estimation in music pieces based on collaborative use of EEG and audio features," in *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2013, pp. 1197–1201.
- [32] Y. Peng, D. Zhang, and J. Zhang, "A new canonical correlation analysis algorithm with local discrimination," *Neural Processing Letters*, vol. 31, no. 1, pp. 1–15, 2010.
- [33] S. Akaho, "A kernel method for canonical correlation analysis," in *Proc. of the International Meeting of the Psychometric Society (IMPS)*, 2001.
- [34] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1226–1238, 2005.
- [35] C.-W. Hsu, C.-C. Chang, and C.-J. Lin, "A practical guide to support vector classification," *Technical Report*, Department of Computer Science, 2003.