Novel Audio Feature Projection Using KDLPCCA-based Correlation with EEG Features for Favorite Music Classification

Ryosuke Sawata, *Member, IEEE,* Takahiro Ogawa, *Member, IEEE,* and Miki Haseyama, *Senior Member, IEEE*

Abstract—A novel audio feature projection using Kernel Discriminative Locality Preserving Canonical Correlation Analysis (KDLPCCA)-based correlation with electroencephalogram (EEG) features for favorite music classification is presented in this paper. The projected audio features reflect individual music preference adaptively since they are calculated by considering correlations with the user's EEG signals during listening to musical pieces that the user likes/dislikes via a novel CCA proposed in this paper. The novel CCA, called KDLPCCA, can consider not only a non-linear correlation but also local properties and discriminative information of each class sample, namely, music likes/dislikes. Specifically, local properties reflect intrinsic data structures of the original audio features, and discriminative information enhances the power of the final classification. Hence, the projected audio features have an optimal correlation with individual music preference reflected in the user's EEG signals, adaptively. If the KDLPCCA-based projection that can transform original audio features into novel audio features is calculated once, our method can extract projected audio features from a new musical piece without newly observing individual EEG signals. Our method therefore has a high level of practicability. Consequently, effective classification of user's favorite musical pieces via a Support Vector Machine (SVM) classifier using the new projected audio features becomes feasible. Experimental results show that our method for favorite music classification using projected audio features via the novel CCA outperforms methods using original audio features, EEG features and even audio features projected by other state-of-the-art CCAs.

Index Terms—Electroencephalogram (EEG), music liking/disliking, canonical correlation analysis (CCA), kernel method, locality preservation, support vector machine (SVM)

1 Introduction

Recently, a large amount of music has become available to users, and efficient music information retrieval technology is therefore necessary to retrieve musical pieces desired by the users. Many methods related to genre classification [1], [2], [3], [4], [5], [6], [7], artist identification [8], [9], music mood classification [10], [11], [12], [13], instrument recognition [14], [15], [16], [17] and music annotation [18], [19] have been proposed to help users retrieve desired musical pieces. Although these methods can help users to retrieve desired musical pieces from music databases, retrieving desired musical pieces from enormous music databases is still a laborious task.

More recently, many music recommendation methods that aim to provide desired musical pieces automatically have been proposed [20], [21] in order to solve the aforementioned problem. Generally, there are two typical algorithms for music recommendation: (1) collaborative filtering (CF) [22] and (2) content-based filtering (CBF) [23]. These recommendation methods are simple and have a relatively high level of practicability, and thus some music discovery websites such as *Amazon*[1] and *Last.fm*[2] have employed these methods. In addition, novel approaches using audio features have been proposed in order to recommend musical pieces effectively [24], [25]. Chiliguano *et al.* proposed a music recommender system that applies deep neural networks (DNN) to audio features in order to obtain new audio features considering the relationship between a user's preference and music genre [24]. Specifically, a genre vector that is not limited to a specific genre can be obtained from a musical piece by the DNN output layer, each component of which represents the probability corresponding to a certain music genre. In this way, musical pieces can be flexibly dealt with without limiting them to specific genres, and effective music recommendation can be realize by using the DNN-based genre vectors. Benzi *et al.* proposed a music recommender system that uses the graph theory and Non-negative Matrix Factorization (NMF) [25]. Their goal is to find an approximate low-rank and non-negative representation of a user-item matrix by considering the users' playlists and the contents of musical pieces listened to. In order to achieve their goal, they introduced audio feature vectors obtained from music listened to and playlist graphs between users to NMF optimization. As a result of this, a new effective playlist is recommended to a user by using the obtained low-rank and non-negative matrices.

Since the services and the experimental results of studies showed high performances, effective music recommendations were actually realized. However, there is a critical problem of recommendation performance depending on each user, *i.e.*, there is low adaptivity for individuals. For instance, Chiliguano *et al.* and Benzi *et al.* experimentally used the traditional audio features, and their audio features are not always suitable for individual music preference. These problems have been discussed in many reports including [26], [27], [28], [29], [30]. Therefore, it is necessary to exploit an effective method that can classify favorite musical pieces for each user adaptively in order to solve

• *The authors are with the Graduate School of Information Science and Technology, Hokkaido University, Sapporo 060-0814, Japan.*
E-mail: sawata@lmd.ist.hokudai.ac.jp, ogawa@lmd.ist.hokudai.ac.jp, miki@ist.hokudai.ac.jp

*Manuscript received June 15, 2016; revised *** **, 2016; accepted *** **, 2016. Date of publication *** **, 2016.*

1. http://www.amazon.com/
2. http://www.last.fm/

the aforementioned conventional problem. In order to satisfy this necessity, calculation of novel audio features suitable for representing each user's music preference is inevitable. We argue that utilization of heterogeneous features, which are features other than audio features and can represent each user's music preference, is important to provide a solution to the above problem.

From this background, context-aware recommendation approaches that aim at personalization and enhancement of the recommendation performance by utilizing contextual information, *e.g.*, related tags, history of rating and so on, have been proposed [31], [32], [33], [34], [35]. For instance, Zhang *et al.* proposed a virtual rating method in order to enhance the performance of CF-based recommendation. Guo *et al.* utilized social information as contexts to improve data selection for recommendation. Novel context-aware approaches that leverage biosignals as context information have also recently been proposed [36], [37], [38]. This is because recording some biosignals within naturalistic environments has become feasible [36]. In particular, observation of electroencephalogram (EEG) signals has become easier, and the quality of observed signals has become better in recent years. This is because there have recently been some studies aimed at the development of wearable devices, which enable a user's EEG signals to be easily observed and do not hinder everyday life [39], [40], [41]. For instance, Lin *et al.* proposed a wearable and wireless EEG device named *Mindo* [39]. Small wearable EEG acquisition devices, which are capable of recording EEG signals without hindering the user who listens to music, have gradually become available. In addition, it is considered that observing EEG signals is particularly an effective one of the current techniques aiming to observe biosignals since they can provide high temporal resolution to directly reflect the dynamics of brain activity as described in [42], [43]. Therefore, there are currently many novel approaches to estimate internal information of humans, such as information on emotion, comfort and preference, based on features extracted from affective phenomena of humans via EEG signals [44], [45], [46], [47], [48], [49]. For instance, Hadjidimitriou *et al.* [46], [47] studied discrimination between a user's EEG signals depending on music preference, *i.e.*, whether the user listened to musical pieces the user liked or disliked. The results of those studies showed that EEG features can reflect individual music preference, which is important for our purpose. Thus, we are convinced that EEG signals will become effective context information for recommendation systems, and we consider that calculating a set of novel audio features will be feasible by monitoring the relationship between the EEG and audio features. Actually, Koelstra *et al.* classified emotions of subjects when they were watching videos by mutually using multimodal features containing EEG and audio features [50]. However, those approaches including the context-aware recommendations using biosignals are different from our idea since Koelstra *et al.*, for example, firstly used feature fusion, which is realized by concatenating heterogeneous feature vectors simply, and finally applied decision fusion. Namely, such methods effectively utilizing the correlations between affective phenomena of a human listening to musical pieces and those audio signals have not been researched adequately as far as we know. Hence, we argue that utilization of the correlations with EEG features extracted from the user's EEG signals during listening to music will be necessary and become a powerful solution for calculation of novel audio features suitable for representing the user's music preference. Based on the above discussion, we regard "affective phenomena" described in the above as the responses of the user's EEG signals during listening to musical pieces that the user likes or dislikes, and we utilize the responses for calculation of novel audio features based on the correlation.

In order to extract the correlation quantitatively, we pay attention to Canonical Correlation Analysis (CCA) [51]. CCA can extract a correlation via canonical variates obtained from a pair of multivariate datasets by maximizing a linear correlation. In [52], we previously proposed CCA-based audio feature selection, by which audio features suitable for individual music preference are selected. However, Yeh *et al.* [53] reported that canonical variates projected by applying CCA to heterogeneous sets of features show better discriminative performance than that of original features if the heterogeneous sets have semantic relevancy. Inspired by CCA and that report, we consider that novel audio features that are more suitable for the user's music preference should be realized by not feature selection but projection by applying CCA to audio and EEG features. However, if there is a non-linear relationship between the EEG and audio features, standard CCA will not always extract useful features. Moreover, standard CCA may miss intrinsic data structures since EEG signals are generally noisy, and CCA cannot consider class information of music preference, *i.e.*, whether the user likes or dislikes a musical piece, although training EEG and audio features have class information. Thus, exploiting a novel CCA that can avoid the above concern is inevitable for our purpose.

Motivated by the aforementioned background, in this paper, we firstly propose a novel CCA that can satisfy our purpose and next propose novel audio features based on our CCA for favorite music classification. The novel CCA, called Kernel Discriminative Locality Preserving CCA (KDLPCCA), can consider not only a non-linear correlation but also local properties and discriminative information of each class sample, namely music likes/dislikes. Our purpose is effective music recommendation realized by classifying user's music preference, *i.e.*, favorite music, and thus consideration of the relationship between EEG signals obtained from a user during listening to music and audio signals listened to is important. However, the relationship based on standard linear correlation may be insufficient since (a) the relationship generally contains a non-linear correlation and (b) the standard formulation of correlation does not consider class labels, *i.e.*, "favorite" or "unfavorite" for the user. Furthermore, even if a supervised scheme that can consider class information is introduced, (c) classes such as "favorite" and "unfavorite" have multimodality, *e.g.*, different and similar genres are in a class. Therefore, we solve these problems by applying our novel KDLPCCA. Specifically, non-linear correlation based on a kernel method solves problem (a), and consideration of class labels and local structure of input data simultaneously solves problems (b) and (c) since Sugiyama reported that locality preserving approaches can work well with data having a multimodal class [54]. Therefore, the projected audio features by KDLPCCA are expected to reflect individual music preference effectively.

In the proposed method, we obtain a projection that can transform original audio features into novel audio features by applying KDLPCCA to EEG and original audio features. Hence, if the KDLPCCA-based projection is calculated once, our method can extract novel audio features from a new musical piece by using the already calculated projection without newly observing individual EEG signals. Our method therefore has a high level of practicability. Consequently, we train a Support Vector Machine (SVM) [55] classifier using these projected audio features, and

then effective classification of the user's favorite musical pieces becomes feasible.

The rest of this paper is organized as follows. In Section **2**, a brief review of some CCAs and their expansions is given. In Section **3**, KDLPCCA is explained. In Section **4**, classification of favorite musical pieces using projected audio features by KDLPCCA is explained. In Section **5**, we show the effectiveness of novel projected audio features for favorite music classification. Finally, we summarize this paper in Section **6**.

2 Review of CCA, KCCA and Their Expansions

In this section, we give a brief review of standard CCA [51], Kernel CCA (KCCA) [56] and their expansions [57], [58], [59], [60] in order to explain our newly derived KDLPCCA. In this section, we assume that there are two heterogeneous sets of N feature vectors having class information: $\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_N] \in \mathbb{R}^{d_x \times N}$ and $\boldsymbol{Y} = [\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_N] \in \mathbb{R}^{d_y \times N}$, where d_x and d_y are respectively the dimensions of corresponding feature vectors.

2.1 CCA and Its Expansions

CCA and its expansions aim at finding pair-wise projection vectors $\boldsymbol{w}_x \in \mathbb{R}^{d_x}$ and $\boldsymbol{w}_y \in \mathbb{R}^{d_y}$ for $\boldsymbol{X}$ and $\boldsymbol{Y}$ to maximize the following objective function:

$$(\hat{\boldsymbol{w}}_x, \hat{\boldsymbol{w}}_y) = \arg \max_{\boldsymbol{w}_x, \boldsymbol{w}_y} \frac{C(\boldsymbol{w}_x, \boldsymbol{w}_y)}{C(\boldsymbol{w}_x)\, C(\boldsymbol{w}_y)}. \tag{1}$$

The terms of this objective function, *i.e.*, $C(\boldsymbol{w}_x, \boldsymbol{w}_y)$, $C(\boldsymbol{w}_x)$ and $C(\boldsymbol{w}_y)$, are defined in each method as follows.

(i) CCA [51]:

$$C(\boldsymbol{w}_x, \boldsymbol{w}_y) = \boldsymbol{w}_x{}^{\mathrm{T}} \boldsymbol{X}\boldsymbol{H}\boldsymbol{H}\boldsymbol{Y}^{\mathrm{T}} \boldsymbol{w}_y, \tag{2}$$

$$C(\boldsymbol{w}_x) = \sqrt{\boldsymbol{w}_x{}^{\mathrm{T}} \boldsymbol{X}\boldsymbol{H}\boldsymbol{H}\boldsymbol{X}^{\mathrm{T}} \boldsymbol{w}_x}, \tag{3}$$

$$C(\boldsymbol{w}_y) = \sqrt{\boldsymbol{w}_y{}^{\mathrm{T}} \boldsymbol{Y}\boldsymbol{H}\boldsymbol{H}\boldsymbol{Y}^{\mathrm{T}} \boldsymbol{w}_y}, \tag{4}$$

where $\boldsymbol{H} = \boldsymbol{I} - \frac{1}{N}\mathbf{1}\mathbf{1}^{\mathrm{T}}$ is a centering matrix, $\boldsymbol{I}$ is the $N \times N$ identity matrix, and $\mathbf{1} = [1, \cdots, 1]^{\mathrm{T}} \in \mathbb{R}^N$ is an N-dimensional vector.

(ii) Locality Preserving CCA (LPCCA) [57]:

$$C(\boldsymbol{w}_x, \boldsymbol{w}_y) = \boldsymbol{w}_x{}^{\mathrm{T}} \boldsymbol{X}\boldsymbol{H}\boldsymbol{L}_{xy}\boldsymbol{H}\boldsymbol{Y}^{\mathrm{T}} \boldsymbol{w}_y, \tag{5}$$

$$C(\boldsymbol{w}_x) = \sqrt{\boldsymbol{w}_x{}^{\mathrm{T}} \boldsymbol{X}\boldsymbol{H}\boldsymbol{L}_{xx}\boldsymbol{H}\boldsymbol{X}^{\mathrm{T}} \boldsymbol{w}_x}, \tag{6}$$

$$C(\boldsymbol{w}_y) = \sqrt{\boldsymbol{w}_y{}^{\mathrm{T}} \boldsymbol{Y}\boldsymbol{H}\boldsymbol{L}_{yy}\boldsymbol{H}\boldsymbol{Y}^{\mathrm{T}} \boldsymbol{w}_y}, \tag{7}$$

where $\boldsymbol{L}_\bullet$ is the Laplacian matrix that considers local structures of the input data, and its details are described in [57].

(iii) Discriminative LPCCA (DLPCCA) [60]:

$$C(\boldsymbol{w}_x, \boldsymbol{w}_y) = \boldsymbol{w}_x{}^{\mathrm{T}} (\boldsymbol{C}_w^d - \eta \boldsymbol{C}_b^d) \boldsymbol{w}_y, \tag{8}$$

$$C(\boldsymbol{w}_x) = \sqrt{\boldsymbol{w}_x{}^{\mathrm{T}} \boldsymbol{X}\boldsymbol{H}(\boldsymbol{L}_{xx} + \bar{\boldsymbol{L}}_{xx})\boldsymbol{H}\boldsymbol{X}^{\mathrm{T}} \boldsymbol{w}_x}, \tag{9}$$

$$C(\boldsymbol{w}_y) = \sqrt{\boldsymbol{w}_y{}^{\mathrm{T}} \boldsymbol{Y}\boldsymbol{H}(\boldsymbol{L}_{yy} + \bar{\boldsymbol{L}}_{yy})\boldsymbol{H}\boldsymbol{Y}^{\mathrm{T}} \boldsymbol{w}_y}, \tag{10}$$

where η is a tunable parameter, and

$$\boldsymbol{C}_w^d = \boldsymbol{X}(\boldsymbol{S}_x \circ \boldsymbol{S}_y)\boldsymbol{Y}^{\mathrm{T}}, \tag{11}$$

$$\boldsymbol{C}_b^d = \boldsymbol{X}(\bar{\boldsymbol{S}}_x \circ \bar{\boldsymbol{S}}_y)\boldsymbol{Y}^{\mathrm{T}}, \tag{12}$$

then $\boldsymbol{S}_\bullet$ and $\bar{\boldsymbol{S}}_\bullet$ are similarity matrices representing the relationships between within-class and between-class samples, respectively, and the symbol $\circ$ denotes the Hadamard product. Furthermore, $\boldsymbol{L}_\bullet$ and $\bar{\boldsymbol{L}}_\bullet$ are Laplacian matrices based on $\boldsymbol{S}_\bullet$ and $\bar{\boldsymbol{S}}_\bullet$, respectively. The details of these computations are explained in [60].

Since the objective function of standard CCA denotes the correlation between $\boldsymbol{X}$ and $\boldsymbol{Y}$, CCA can find pair-wise projection vectors $\boldsymbol{w}_x$ and $\boldsymbol{w}_y$ that maximize their correlations. Thus, CCA realizes effective feature extraction considering the semantic relevancy between heterogeneous sets of features and has been used in several research fields such as information retrieval, transfer learning and pattern recognition [61], [62], [63], [64]. However, standard CCA cannot find a non-linear correlation and deal with class information if original sets have such characteristics.

As a solution for the above non-linear problem, Sun *et al.* proposed **(ii)** LPCCA [57] based on Locality Preserving Projection (LPP) [65]. Generally, global non-linear structures are locally linear, and local structures can be aligned. Therefore, LPCCA can find a non-linear correlation by introducing $\boldsymbol{L}_\bullet$, which can preserve the local structures of original data, to the standard CCA's objective function like Eqs. (5)-(7). As a solution for the above class information problem, Sun *et al.* and Peng *et al.* proposed DCCA [58] and LDCCA [59], respectively. Sun *et al.* firstly defined the within-class covariance matrix and between-class covariance matrix. Then DCCA can find pairs of projection vectors considering the correlation and class information mutually by using the defined within-class and between-class covariance matrices. However, DCCA does not deal with these matrices fairly since the between-class covariance matrix becomes the same matrix as the within-class matrix by expanding these original formulas, *i.e.*, DCCA deals with only the within/between-class matrix. On the other hand, Peng *et al.* defined the local within-class covariance matrix and local between-class covariance matrix, aiming at considering the local structures and class information of the original data simultaneously. These matrices not only avoid the problem of DCCA but also can consider local structures of original data. Therefore, LDCCA has better performance than that of DCCA in general but does not consider the local structures as sufficiently as LPCCA does due to just using k nearest neighborhoods in the defined local within-class and between-class matrices. In other words, using k nearest neighborhoods as the local structures may be insufficient. In order to consider the correlation, class information and local structures as considered in LPCCA, Zhang *et al.* proposed **(iii)** DLPCCA [60]. Since DLPCCA calculates the within-class covariance matrix $\boldsymbol{C}_w^d$ and between-class covariance matrix $\boldsymbol{C}_b^d$ using $\boldsymbol{S}_\bullet$ and $\bar{\boldsymbol{S}}_\bullet$ as LPCCA does, DLPCCA can find a correlation by considering class information and local structures more effectively than can LDCCA.

2.2 KCCA and Its Expansions

In terms of a non-linear problem, there is a kernel trick that is a more major solution than LPP. CCA using a kernel trick, *i.e.*, Kernel CCA (KCCA) [56], firstly maps data points into high dimensional Hilbert space by $\phi_x : \boldsymbol{x} \mapsto \phi_x(\boldsymbol{x}) \in \mathbb{R}^{d_{\phi_x}}$ and $\phi_y : \boldsymbol{y} \mapsto \phi_y(\boldsymbol{y}) \in \mathbb{R}^{d_{\phi_y}}$ in order to find a non-linear correlation. Thus, we obtain $\boldsymbol{\Phi}_x = [\phi_x(\boldsymbol{x}_1), \phi_x(\boldsymbol{x}_2), \cdots, \phi_x(\boldsymbol{x}_N)] \in \mathbb{R}^{d_{\phi_x} \times N}$ and $\boldsymbol{\Phi}_y = [\phi_y(\boldsymbol{y}_1), \phi_y(\boldsymbol{y}_2), \cdots, \phi_y(\boldsymbol{y}_N)] \in \mathbb{R}^{d_{\phi_y} \times N}$. Then we can

extract the non-linear correlation by maximizing the correlation

$$\rho = \frac{{\boldsymbol{w}_{\phi_x}}^{\mathrm{T}} \boldsymbol{\Phi}_x \boldsymbol{H}\boldsymbol{H} \boldsymbol{\Phi}_y^{\mathrm{T}} \boldsymbol{w}_{\phi_y}}{\sqrt{{\boldsymbol{w}_{\phi_x}}^{\mathrm{T}} \boldsymbol{\Phi}_x \boldsymbol{H}\boldsymbol{H} \boldsymbol{\Phi}_x^{\mathrm{T}} \boldsymbol{w}_{\phi_x}}} \sqrt{{\boldsymbol{w}_{\phi_y}}^{\mathrm{T}} \boldsymbol{\Phi}_y \boldsymbol{H}\boldsymbol{H} \boldsymbol{\Phi}_y^{\mathrm{T}} \boldsymbol{w}_{\phi_y}}} \tag{13}$$

over the projection directions $\boldsymbol{w}_{\phi_x} \in \mathbb{R}^{d_{\phi_x}}$ and $\boldsymbol{w}_{\phi_y} \in \mathbb{R}^{d_{\phi_y}}$. However, obtaining $\boldsymbol{w}_{\phi_x}$ and $\boldsymbol{w}_{\phi_y}$ is generally difficult since the dimensions of each Hilbert space's vector, *i.e.*, d_{ϕ_x} and d_{ϕ_y}, are very high. By using the kernel trick that is represented as the kernel function $k_x(\bullet,\bullet) = \phi_x(\bullet)^{\mathrm{T}}\phi_x(\bullet)$ and $k_y(\bullet,\bullet) = \phi_y(\bullet)^{\mathrm{T}}\phi_y(\bullet)$, KCCA and its expansions can extract the non-linear correlation hidden between the original $\boldsymbol{X}$ and $\boldsymbol{Y}$. For instance, KCCA extracts the non-linear correlation by solving the following objective function:

$$(\hat{\boldsymbol{\alpha}}_x, \hat{\boldsymbol{\alpha}}_y) = \arg \max_{\boldsymbol{\alpha}_x, \boldsymbol{\alpha}_y} \frac{{\boldsymbol{\alpha}_x}^{\mathrm{T}} \boldsymbol{C}_{xy}^{\phi} \boldsymbol{\alpha}_y}{\sqrt{{\boldsymbol{\alpha}_x}^{\mathrm{T}} \boldsymbol{C}_{xx}^{\phi} \boldsymbol{\alpha}_x}\ \sqrt{{\boldsymbol{\alpha}_y}^{\mathrm{T}} \boldsymbol{C}_{yy}^{\phi} \boldsymbol{\alpha}_y}}, \tag{14}$$

where Eq. (14) is obtained by rewriting $\boldsymbol{w}_{\phi_x}$ and $\boldsymbol{w}_{\phi_y}$ as $\boldsymbol{w}_{\phi_x} = \boldsymbol{\Phi}_x \boldsymbol{H} \boldsymbol{\alpha}_x$ and $\boldsymbol{w}_{\phi_y} = \boldsymbol{\Phi}_y \boldsymbol{H} \boldsymbol{\alpha}_y$ respectively, based on dual representation [66], [67]. Note that $\boldsymbol{\alpha}_x \in \mathbb{R}^N$ and $\boldsymbol{\alpha}_y \in \mathbb{R}^N$ are coefficient vectors. Moreover, $\boldsymbol{C}_{xy}^{\phi} \in \mathbb{R}^{N\times N}$, $\boldsymbol{C}_{xx}^{\phi} \in \mathbb{R}^{N\times N}$ and $\boldsymbol{C}_{yy}^{\phi} \in \mathbb{R}^{N\times N}$ are defined as follows:

$$\boldsymbol{C}_{xy}^{\phi} = \boldsymbol{H}\boldsymbol{K}_x\boldsymbol{H}\boldsymbol{H}\boldsymbol{K}_y\boldsymbol{H}, \tag{15}$$

$$\boldsymbol{C}_{xx}^{\phi} = \boldsymbol{H}\boldsymbol{K}_x\boldsymbol{H}\boldsymbol{H}\boldsymbol{K}_x\boldsymbol{H} + \xi_x \boldsymbol{H}\boldsymbol{K}_x\boldsymbol{H}, \tag{16}$$

$$\boldsymbol{C}_{yy}^{\phi} = \boldsymbol{H}\boldsymbol{K}_y\boldsymbol{H}\boldsymbol{H}\boldsymbol{K}_y\boldsymbol{H} + \xi_y \boldsymbol{H}\boldsymbol{K}_y\boldsymbol{H}, \tag{17}$$

where $\boldsymbol{K}_x \in \mathbb{R}^{N\times N}$ and $\boldsymbol{K}_y \in \mathbb{R}^{N\times N}$ are gram matrices whose (i,j)th elements are $k_x(\boldsymbol{x}_i,\boldsymbol{x}_j)$ and $k_y(\boldsymbol{y}_i,\boldsymbol{y}_j)$, respectively. Furthermore, ξ_x and ξ_y are regularization parameters.

KCCA can solve a non-linear problem in the original space by applying the "kernel trick" that can compute the corresponding CCA on the mapped data without actually knowing the mapping of ϕ_x and ϕ_y themselves. A sufficient condition for a kernel function $k_\bullet(\bullet,\bullet)$ corresponding to an inner product in the Hilbert space is given by Mercer's theorem [68]. Therefore, KCCA can derive not only a linear correlation but also a non-linear correlation between $\boldsymbol{X}$ and $\boldsymbol{Y}$ by using $\boldsymbol{\alpha}_x$ and $\boldsymbol{\alpha}_y$. Furthermore, Kernel DCCA (KDCCA) [58] and Kernel LDCCA (KLDCCA) [59], which are kernelized versions of DCCA and LDCCA, respectively, have been proposed as expansions of KCCA. KDCCA and KLDCCA can consider not only a non-linear correlation but also class information since these methods inherit the merits from DCCA and LDCCA. These kernelized versions of CCA generally show more efficient performance than that of original CCAs, *i.e.*, CCA, DCCA and LDCCA. Therefore, introducing the kernel trick into the non-kernel CCAs is expected to enhance the performance of final classification, estimation, recognition, etc. However, so far, a kernelized CCA using LPP such as LPCCA and DLPCCA has not been researched adequately. DLPCCA, which generally has the best performance in recent non-kernel CCAs as far as we know, is expected to become a powerful method for feature extraction if a kernel trick is introduced into it. Motivated by these factors, we propose Kernel DLPCCA (KDLPCCA) in the following section.

3 KDLPCCA

In this section, we show the derivation of KDLPCCA and briefly explain its merits by comparing it with other CCAs explained in Section **2**. In this section, we also assume that there are two heterogeneous sets of N feature vectors that are the same as those shown in Section **2**, *i.e.*, $\boldsymbol{X} \in \mathbb{R}^{d_x\times N}$ and $\boldsymbol{Y} \in \mathbb{R}^{d_y\times N}$.

KDLPCCA firstly maps these data points into the Hilbert space in the same manner as that shown in Section **2**, namely, we obtain $\boldsymbol{\Phi}_x \in \mathbb{R}^{d_{\phi_x}\times N}$ and $\boldsymbol{\Phi}_y \in \mathbb{R}^{d_{\phi_y}\times N}$. Then we apply DLPCCA to $\boldsymbol{\Phi}_x$ and $\boldsymbol{\Phi}_y$ using a kernel trick as the inner product in the Hilbert space. First, we compute similarities between the ith and jth class samples in the mapped Hilbert space as follows:

$$S_{ij}^{\phi_x} = \begin{cases} \exp(-\delta_{ij}^x / t_x^{\phi}), & \mathrm{label}(\phi_x(\boldsymbol{x}_i)) = \mathrm{label}(\phi_x(\boldsymbol{x}_j)) \\ 0, & \text{otherwise} \end{cases} \tag{18}$$

$$\bar{S}_{ij}^{\phi_x} = \begin{cases} \exp(-\delta_{ij}^x / t_x^{\phi}), & \mathrm{label}(\phi_x(\boldsymbol{x}_i)) \neq \mathrm{label}(\phi_x(\boldsymbol{x}_j)) \\ 0, & \text{otherwise} \end{cases} \tag{19}$$

where $\delta_{ij}^x = \|\phi_x(\boldsymbol{x}_i) - \phi_x(\boldsymbol{x}_j)\|^2 = (\boldsymbol{K}_x)_{ii} - 2(\boldsymbol{K}_x)_{ij} + (\boldsymbol{K}_x)_{jj}$, and $(\boldsymbol{K}_x)_{ij}$ denotes the (i,j)th element of $\boldsymbol{K}_x$. Moreover, "label$(\phi_x(\boldsymbol{x}_\bullet))$" is the $\boldsymbol{x}_\bullet$'s class label, and $t_x^{\phi} = \frac{1}{N(N-1)} \sum_{i=1}^{N} \sum_{j=1}^{N} \delta_{ij}^x$. As a result of this, we can obtain similarity matrices $\boldsymbol{S}_x^{\phi} = \{S_{ij}^{\phi_x}\}_{i,j=1}^N$ and $\bar{\boldsymbol{S}}_x^{\phi} = \{\bar{S}_{ij}^{\phi_x}\}_{i,j=1}^N$ representing the similarities between samples having the same and different labels, respectively. Note that similarity matrices with respect to $\boldsymbol{\Phi}_y$, *i.e.*, $\boldsymbol{S}_y^{\phi} = \{S_{ij}^{\phi_y}\}_{i,j=1}^N$ and $\bar{\boldsymbol{S}}_y^{\phi} = \{\bar{S}_{ij}^{\phi_y}\}_{i,j=1}^N$, are calculated in the same manner. Hereafter, we explain the calculation related to $\boldsymbol{\Phi}_x$ only since the calculation related to $\boldsymbol{\Phi}_y$ can be performed in the same manner. Based on these similarity matrices, we next calculate Laplacian matrices in order to consider the LPP-based local structures and the class information of input data shown as follows:

$$\boldsymbol{L}_{xx}^{\phi} = \boldsymbol{D}_{xx}^{\phi} - \boldsymbol{S}_x^{\phi} \circ \boldsymbol{S}_x^{\phi}, \tag{20}$$

$$\bar{\boldsymbol{L}}_{xx}^{\phi} = \bar{\boldsymbol{D}}_{xx}^{\phi} - \bar{\boldsymbol{S}}_x^{\phi} \circ \bar{\boldsymbol{S}}_x^{\phi}, \tag{21}$$

where $\boldsymbol{D}_{xx}^{\phi} = \mathrm{diag}[\sum_i (S_{1i}^{\phi_x})^2, \sum_i (S_{2i}^{\phi_x})^2, \cdots, \sum_i (S_{Ni}^{\phi_x})^2]$ and $\bar{\boldsymbol{D}}_{xx}^{\phi} = \mathrm{diag}[\sum_i (\bar{S}_{1i}^{\phi_x})^2, \sum_i (\bar{S}_{2i}^{\phi_x})^2, \cdots, \sum_i (\bar{S}_{Ni}^{\phi_x})^2]$. The definition of $\boldsymbol{L}_{xx}^{\phi}$ $(\bar{\boldsymbol{L}}_{xx}^{\phi}, \boldsymbol{L}_{yy}^{\phi}, \bar{\boldsymbol{L}}_{yy}^{\phi})$ is similar to that of the Laplacian matrix in LPP [65] except that the Laplacian matrix in KDLPCCA can consider not only local structures but also class information. Hence, KDLPCCA can extract a non-linear correlation considering local structures and class information simultaneously by solving the following function:

$$(\hat{\boldsymbol{\alpha}}_x, \hat{\boldsymbol{\alpha}}_y) = \arg \max_{\boldsymbol{\alpha}_x, \boldsymbol{\alpha}_y} \frac{{\boldsymbol{\alpha}_x}^{\mathrm{T}} (\boldsymbol{C}_w^{\phi_d} - \eta \boldsymbol{C}_b^{\phi_d}) \boldsymbol{\alpha}_y}{\sqrt{{\boldsymbol{\alpha}_x}^{\mathrm{T}} \boldsymbol{C}_{xx}^{\phi_d} \boldsymbol{\alpha}_x}\ \sqrt{{\boldsymbol{\alpha}_y}^{\mathrm{T}} \boldsymbol{C}_{yy}^{\phi_d} \boldsymbol{\alpha}_y}}, \tag{22}$$

where

$$\boldsymbol{C}_w^{\phi_d} = \boldsymbol{H}\boldsymbol{K}_x\boldsymbol{H}(\boldsymbol{S}_x^{\phi} \circ \boldsymbol{S}_y^{\phi})\boldsymbol{H}\boldsymbol{K}_y\boldsymbol{H}, \tag{23}$$

$$\boldsymbol{C}_b^{\phi_d} = \boldsymbol{H}\boldsymbol{K}_x\boldsymbol{H}(\bar{\boldsymbol{S}}_x^{\phi} \circ \bar{\boldsymbol{S}}_y^{\phi})\boldsymbol{H}\boldsymbol{K}_y\boldsymbol{H}, \tag{24}$$

$$\boldsymbol{C}_{xx}^{\phi_d} = \boldsymbol{H}\boldsymbol{K}_x\boldsymbol{H}(\boldsymbol{L}_{xx}^{\phi} + \bar{\boldsymbol{L}}_{xx}^{\phi})\boldsymbol{H}\boldsymbol{K}_x\boldsymbol{H} + \xi_x \boldsymbol{H}\boldsymbol{K}_x\boldsymbol{H}, \tag{25}$$

$$\boldsymbol{C}_{yy}^{\phi_d} = \boldsymbol{H}\boldsymbol{K}_y\boldsymbol{H}(\boldsymbol{L}_{yy}^{\phi} + \bar{\boldsymbol{L}}_{yy}^{\phi})\boldsymbol{H}\boldsymbol{K}_y\boldsymbol{H} + \xi_y \boldsymbol{H}\boldsymbol{K}_y\boldsymbol{H}. \tag{26}$$

In this way, we can calculate $\boldsymbol{\alpha}_x$ and $\boldsymbol{\alpha}_y$ as the optimal solutions by solving the following Lagrange function:

$$\begin{aligned} \mathcal{L}(\boldsymbol{\alpha}_x, \boldsymbol{\alpha}_y) = {} & \boldsymbol{\alpha}_x^{\mathrm{T}} (\boldsymbol{C}_w^{\phi_d} - \eta \boldsymbol{C}_b^{\phi_d}) \boldsymbol{\alpha}_y \\ & - \frac{\lambda_x}{2} (\boldsymbol{\alpha}_x^{\mathrm{T}} \boldsymbol{C}_{xx}^{\phi_d} \boldsymbol{\alpha}_x - 1) - \frac{\lambda_y}{2} (\boldsymbol{\alpha}_y^{\mathrm{T}} \boldsymbol{C}_{yy}^{\phi_d} \boldsymbol{\alpha}_y - 1), \end{aligned} \tag{27}$$

where $\lambda = \lambda_x = \lambda_y$, and they become equivalent to the optimal solution of Eq. (22). Therefore, we can finally obtain the following

TABLE 1: Characteristics of the proposed KDLPCCA and other CCAs.

	CCA	LPCCA	DCCA	LDCCA	DLPCCA	KCCA	KDCCA	KLDCCA	**KDLPCCA**
Kernel-based non-linear correlation	×	×	×	×	×	✓	✓	✓	✓
Local stuctures of original data	×	✓	×	×	✓	×	×	×	✓
Class information	×	×	✓	✓	✓	×	✓	✓	✓

eigenvalue problems based on Eq. (27):

$$\begin{bmatrix} & (\boldsymbol{C}_w^{\phi_d} - \eta \boldsymbol{C}_b^{\phi_d}) \\ (\boldsymbol{C}_w^{\phi_d} - \eta \boldsymbol{C}_b^{\phi_d})^{\mathrm{T}} & \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}_x \\ \boldsymbol{\alpha}_y \end{bmatrix} = \lambda \begin{bmatrix} \boldsymbol{C}_{xx}^{\phi_d} & \\ & \boldsymbol{C}_{yy}^{\phi_d} \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}_x \\ \boldsymbol{\alpha}_y \end{bmatrix}. \tag{28}$$

As solutions of Eq. (28), we can obtain some eigenvectors as $\boldsymbol{\alpha}_x$ and $\boldsymbol{\alpha}_y$, respectively. By using these vectors, we can obtain projection vectors $\boldsymbol{w}_x^{\phi_d} \in \mathbb{R}^{d_{\phi_x}}$ and $\boldsymbol{w}_y^{\phi_d} \in \mathbb{R}^{d_{\phi_y}}$ that can project the original vectors $\boldsymbol{x}_i$ and $\boldsymbol{y}_i$ into the subspace considering the non-linear correlation, class information and local structures as follows:

$$\boldsymbol{w}_x^{\phi_d} = \boldsymbol{\Phi}_x \boldsymbol{H} \boldsymbol{\alpha}_x, \tag{29}$$

$$\boldsymbol{w}_y^{\phi_d} = \boldsymbol{\Phi}_y \boldsymbol{H} \boldsymbol{\alpha}_y. \tag{30}$$

In summary, the characteristics of KDLPCCA and other CCAs explained in Section **2** are shown in TABLE 1. As shown in this table, KDLPCCA is expected to have more efficient feature extraction ability for classification than that of the other CCAs since only KDLPCCA can deal with all of the characteristics, *i.e.*, kernel-based non-linear correlation, local structures of original data and class information.

TABLE 2: EEG features used in the proposed method. U denotes the number of channels of EEG signals and U_P represents the number of symmetric electrode pairs placed on the scalp.

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

4 Favorite Music Classification Using Novel Projected Audio Features

In our method, the user's EEG features are extracted from EEG signals recorded while the user listens to favorite musical pieces, and audio features are extracted from the corresponding musical pieces. Next, a projection that can transform audio features into features reflecting the user's music preference is obtained by applying KDLPCCA to the EEG and audio features. Then an SVM classifier is trained by using the projected audio features to realize effective classification of the user's favorite musical pieces.

This section is organized as follows. In **4.1**, we explain the EEG and audio features used in our method. In **4.2**, calculation of the novel audio features projected by KDLPCCA is explained. In **4.3**, we describe a method to classify favorite musical pieces using the SVM trained by the KDLPCCA-based projected audio features.

4.1 Feature Extraction

EEG Feature Extraction

EEG signals are electrical signals recorded as multiple channel signals from multiple electrodes placed on the scalp. We calculate EEG features based on [44], [45], [48]. Furthermore, we utilize the baseline based on [46] for EEG features and apply a feature selection method to enhance the final performance of the classification.

First, segmentation of each channel's EEG signal is performed at a fixed interval as preprocessing. Next, short-time Fourier transform (STFT) is applied to each channel's EEG signal, and some kinds of EEG features are computed as shown in TABLE 2. "Zero Crossing Rate" denotes the rate of sign changes in the duration of each task. Next, some frequency domain-based EEG features, which are shown in TABLE 2, are calculated. Specifically, "Content Percentage of The Power Spectrum" and "Power Spectrum of The Hemispheric Asymmetry" according to each frequency band are respectively calculated since Lin *et al.* reported that these kinds of EEG features have close relations with the affective phenomena of humans [44]. The characteristics of each band have been reported [69], [70], [71]. Most researchers studied frequency bands including θ wave (4-7 Hz), α wave (7-13 Hz) and β wave (13-30 Hz). Sammler *et al.* reported that the power of θ wave is related to listening to pleasant or unpleasant music [69]. It has also been reported that the power of α wave is closely related to valence and intensity of music emotion [70]. Furthermore, Nakamura *et al.* reported that the power of β wave is related to beginning to listen to music from a rest condition [71]. Although these EEG bands have been studied as described above, there have been few studies on γ wave (30-49 Hz). This is because it was reported that γ wave tends to contain artifacts evoked by muscle potential and eye movement [72], [73], [74]. However, Hadjidimitriou *et al.* and Pan *et al.* reported that γ wave is essential for representing music preference [46], [75]. Furthermore, observation of EEG signals has become easier, and the quality of observed signals has become better in recent years due to the development of new equipment (See **1**). We therefore carefully observed EEG signals containing the band of γ wave from a subject listening to music and used γ wave for EEG feature extraction. In particular, "Mean Frequency of γ Wave" and "Spectral Entropy of γ Wave" are newly calculated in our method since γ wave is essential for representing music preference.

TABLE 3: 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
	Low Energy	Mean, Std	2
tonal	Key Strength	Mean, Std	48
	Chromagram	Mean, Std	24
	Key	Mean, Std	2
	Tonal Centroid	Mean, Std	12
	Mode	Mean, Std	2
rhythm	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

Next, we introduce the idea of baseline. As Hadjidimitriou *et al.* pointed out in [46], [47], brain activity just before listening to music is regarded as noise in order to obtain music preference. Therefore, we regard a resting period of 3 sec just before beginning to listen to music as the baseline, and we calculate the baseline's EEG feature vector for reduction of the effect of brain activity just before starting to listen to music. Specifically, we normalize EEG features obtained in the user's listening duration by using those extracted from the user's baseline (3 sec) based on the results of a study on event-related synchronization/desynchronization (ERS/ERD) [76]. Note that this length (3 sec) of the resting period is sufficient since the selection of this resting interval complies with the precedents adopted in studies in which ERS/ERD patterns were investigated [77], [78]. First, we obtain the average EEG feature vector $\bar{\boldsymbol{v}}^R \in \mathbb{R}^d$ ($d = 13U + 16U_p$; U and U_p being defined in the caption of TABLE 2) from a user's baseline:

$$\bar{\boldsymbol{v}}^R = \frac{1}{N^R} \sum_{n^R=1}^{N^R} \boldsymbol{v}^R_{n^R}, \tag{31}$$

where $\boldsymbol{v}^R_{n^R} \in \mathbb{R}^d (n^R = 1,2,\cdots,N^R; N^R$ being the number of EEG segments in the baseline) denotes the n^Rth EEG feature vector. Then, the normalized EEG feature vector $\boldsymbol{v} \in \mathbb{R}^d$ is calculated as follows:

$$\boldsymbol{v}(i) = \frac{\boldsymbol{v}^L(i) - \bar{\boldsymbol{v}}^R(i)}{\bar{\boldsymbol{v}}^R(i)}, \tag{32}$$

where $\boldsymbol{v}^L \in \mathbb{R}^d$ denotes an EEG feature vector extracted from the user's EEG signals during listening to musical pieces. Moreover, $\boldsymbol{v}(i)$, $\boldsymbol{v}^L(i)$ and $\bar{\boldsymbol{v}}^R(i)$ ($i = 1,2,\cdots,d$) represent each vector's ith element, respectively. Consequently, we can reflect a user's brain activity (ERS/ERD) in $\boldsymbol{v}$ effectively since this normalization handles the user's baseline based on the results of the study on ERS/ERD [76].

Finally, we apply the feature selection method based on the Max-Relevance and Min-Redundancy (mRMR) algorithm [79] in order to reduce noise, *i.e.*, obtain an efficient EEG feature set for reflecting a user's preference. Specifically, a user gives a label $+1$ or -1 indicating that he/she either likes or dislikes each musical piece. Then we obtain a set of final EEG features by inputting the normalized EEG feature vectors and their corresponding labels to the mRMR algorithm. In this way, effective classification of a user's favorite musical pieces is expected since only effective EEG features for our purpose are selected via the mRMR algorithm.

Audio Feature Extraction

First, audio signal segmentation is performed at a fixed interval as preprocessing. Next, audio features used in [80], [81] are extracted from each audio segment by applying STFT to each audio segment. In the proposed method, the "Music Information Retrieval (MIR) toolbox" for MATLAB is used to extract audio features[3]. The MIR toolbox has a function named *mirframe(•)* that divides the input audio signal into audio segments, each of which is 0.05 sec in length. Generally, that length of an audio segment (0.05 sec), which is set by the function *mirframe(•)* of the MIR toolbox, is widely used in the field of MIR. Thus, we adopted the default length of the audio segment set by the MIR toolbox, *i.e.*, 0.05 sec, in order to extract audio features. On the other hand, a set of EEG features is derived per an EEG segment (T sec, *e.g.*, 1-2 sec). The audio segment is much shorter than the EEG segment. Therefore, we compute the mean and standard deviation from the audio features per T sec (the same length as the EEG segment) since the durations of EEG and audio, which are used to extract their features by applying the following KDLPCCA, must have equal lengths. In this way, we can obtain audio features shown in TABLE 3.

Note that we do not apply feature selection to audio features unlike EEG features. In the proposed method, we used the feature selection based on the mRMR algorithm with the expectation of two roles: a) denoising and b) extraction of effective features for classification.

In terms of a), application of feature selection is not necessary for audio features since the audio signals obtained from music are not measurement signals, *i.e.*, having no noise originally. On the other hand, EEG signals are measurement signals. Namely, denoising needs to be applied to them as preprocessing since they have some noise and measurement errors due to imperfections of the equipment for observation, the subject's eye or body movements and so on. The importance of denoising as preprocessing for EEG signals was pointed out in [82]

In terms of b), we conducted a preliminary experiment to confirm the effect of applying feature selection, *i.e.*, the mRMR algorithm, to EEG and audio features. In our preliminary experiment, we changed the dimensions of EEG and audio features in the order of the output of mRMR and classified each user's favorite musical pieces by separately using selected EEG and audio features. To confirm the effect of applying feature selection to EEG and audio features only, we set the same experimental conditions except for using EEG and audio features separately. The data of EEG and audio signals are exactly the same as those obtained in our experiments (See **5**). The classifier used in this preliminary experiment was also the same as that used in our experiments, *i.e.*, the linear SVM classifier. First, we applied the mRMR algorithm to EEG and audio features and then obtained

3. Available at https://www.jyu.fi/hum/laitokset/musiikki/en/research/coe/materials/mirtoolbox

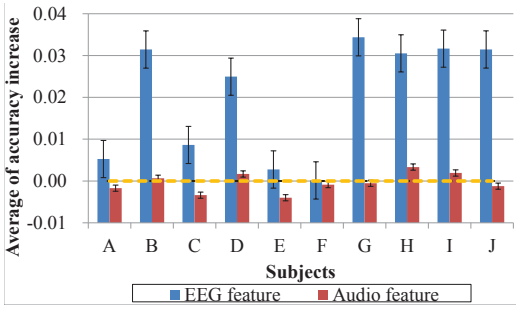


Fig. 1: Average of "accuracy increase" with changes in the dimensions of EEG and audio features. The horizontal dotted orange line denotes zero.

sets of selected EEG and audio features that have max-relevance and min-redundancy with respect to the input labels indicating the user's music preference. Next, we trained the linear SVM classifiers, which aim to classify the vectors obtained from the musical piece for testing into LIKE or DISLIKE, by separately using the sets of selected EEG and audio features. Finally, we classified the user's favorite and unfavorite musical pieces by using these trained SVM classifiers and then calculated the accuracy. To confirm the effect of selecting features, we repeated this experiment with changes in the selected dimensions of EEG and audio features. Specifically, we changed the dimensions of EEG and audio features from 1 to 20 respectively. Then we calculated "accuracy increase" as $\mathrm{ACC}(i+1) - \mathrm{ACC}(i)$, where "$\mathrm{ACC}(i)$" denotes the classification accuracy when i ($i = 1,2,\cdots,19$) dimensions were selected from the original set of features.

The results are shown in Fig. 1. As shown in the figure, most of the accuracy increases for audio features are lower than the horizontal dotted orange line, *i.e.*, zero. On the other hand, all of the accuracy increases for EEG features are higher than the horizontal dotted orange line. Furthermore, to disclose the statistical verification in the effect of feature selection for the set of EEG features, we applied Jonkheere-Terpstra test to the results of feature selection. Jonkheere-Terpstra test is a rank-based nonparametric test and can be used to determine whether there is a statistically significant trend between an ordinal independent variable and a continuous or ordinal dependent variable. Hence, we considered that Jonkheere-Terpstra test is appropriate for our purpose, *i.e.*, testing null hypothesis that there is no performance trend corresponding to changing feature dimensions used for classification. Specifically, we applied Jonkheere-Terpstra test to the classification results respectively using EEG and audio features in ascending order of selecting feature dimensions. Namely, we tested whether there is a statistically significant increasing trend in the classification accuracy using EEG and audio features corresponding to adding a feature one by one based on the results of feature selection. As a result of Jonkheere-Terpstra test, the value p with respect to EEG features was under 0.01. On the other hand, the value p with respect to audio features was 0.124, *i.e.*, over 0.1. Therefore, we can consider that the effect of applying feature selection, *i.e.*, the mRMR algorithm, to EEG features is greater than the effect of applying feature selection to audio features.

From the above preliminary experimental results and discussion, if the effect of selection of audio features is small, it is considered that remaining as many original dimensions of audio features as possible is eligible for the following calculation of projection via KDLPCCA. Therefore, we applied feature selection to EEG features but not to audio features.

4.2 KDLPCCA-based Projection of Audio Features

In this subsection, we describe a method for novel audio feature projection using KDLPCCA-based correlation with EEG features for representing a user's music preference. As described in the previous subsection, two sets of N feature vectors $\boldsymbol{X}^E = [\boldsymbol{x}_1^E, \boldsymbol{x}_2^E, \cdots, \boldsymbol{x}_N^E] \in \mathbb{R}^{d_E \times N}$ and $\boldsymbol{X}^A = [\boldsymbol{x}_1^A, \boldsymbol{x}_2^A, \cdots, \boldsymbol{x}_N^A] \in \mathbb{R}^{d_A \times N}$ are obtained from EEG and audio signals, where d_E and d_A are the dimensions of the EEG feature vector and the audio feature vector, respectively. Note that each feature vector has a label representing whether the user likes the musical pieces, *i.e.*, "like" or "dislike". First, feature vectors of each set, *i.e.*, $\boldsymbol{x}_j^E$ and $\boldsymbol{x}_j^A$ ($j = 1,2,\cdots,N$), are transformed into Hilbert space via nonlinear maps $\phi_E : \boldsymbol{x}^E \mapsto \phi_E(\boldsymbol{x}^E) \in \mathbb{R}^{d_{\phi_E}}$ and $\phi_A : \boldsymbol{x}^A \mapsto \phi_A(\boldsymbol{x}^A) \in \mathbb{R}^{d_{\phi_A}}$, respectively. From the aforementioned mapped results, we obtain $\boldsymbol{\Phi}_E = [\phi_E(\boldsymbol{x}_1^E), \phi_E(\boldsymbol{x}_2^E), \cdots, \phi_E(\boldsymbol{x}_N^E)] \in \mathbb{R}^{d_{\phi_E} \times N}$ and $\boldsymbol{\Phi}_A = [\phi_A(\boldsymbol{x}_1^A), \phi_A(\boldsymbol{x}_2^A), \cdots, \phi_A(\boldsymbol{x}_N^A)] \in \mathbb{R}^{d_{\phi_A} \times N}$. Next, we apply KDLPCCA explained in Section **3** to these EEG and audio features. Based on Eqs. (22)-(28), we can finally obtain the following generalized eigenvalue problem.

$$\begin{bmatrix} & (\boldsymbol{C}_w^{EA} - \eta \boldsymbol{C}_b^{EA}) \\ (\boldsymbol{C}_w^{EA} - \eta \boldsymbol{C}_b^{EA})^{\mathrm{T}} & \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}_E \\ \boldsymbol{\alpha}_A \end{bmatrix} = \lambda \begin{bmatrix} \boldsymbol{C}_{\phi_E} & \\ & \boldsymbol{C}_{\phi_A} \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}_E \\ \boldsymbol{\alpha}_A \end{bmatrix}, \tag{33}$$

where

$$\boldsymbol{C}_w^{EA} = \boldsymbol{H}\boldsymbol{K}_E\boldsymbol{H}(\boldsymbol{S}_E^{\phi} \circ \boldsymbol{S}_A^{\phi})\boldsymbol{H}\boldsymbol{K}_A\boldsymbol{H}, \tag{34}$$

$$\boldsymbol{C}_b^{EA} = \boldsymbol{H}\boldsymbol{K}_E\boldsymbol{H}(\bar{\boldsymbol{S}}_E^{\phi} \circ \bar{\boldsymbol{S}}_A^{\phi})\boldsymbol{H}\boldsymbol{K}_A\boldsymbol{H}, \tag{35}$$

$$\boldsymbol{C}_{\phi_E} = \boldsymbol{H}\boldsymbol{K}_E\boldsymbol{H}(\boldsymbol{L}_E^{\phi} + \bar{\boldsymbol{L}}_E^{\phi})\boldsymbol{H}\boldsymbol{K}_E\boldsymbol{H} + \xi_E\boldsymbol{H}\boldsymbol{K}_E\boldsymbol{H}, \tag{36}$$

$$\boldsymbol{C}_{\phi_A} = \boldsymbol{H}\boldsymbol{K}_A\boldsymbol{H}(\boldsymbol{L}_A^{\phi} + \bar{\boldsymbol{L}}_A^{\phi})\boldsymbol{H}\boldsymbol{K}_A\boldsymbol{H} + \xi_A\boldsymbol{H}\boldsymbol{K}_A\boldsymbol{H}. \tag{37}$$

Note that $\boldsymbol{K}_E$ and $\boldsymbol{K}_A$ denote the gram matrices of EEG and audio features, respectively. Furthermore, $\boldsymbol{S}_E^{\phi}$, $\bar{\boldsymbol{S}}_E^{\phi}$, $\boldsymbol{L}_E^{\phi}$, $\bar{\boldsymbol{L}}_E^{\phi}$ ($\boldsymbol{S}_A^{\phi}$, $\bar{\boldsymbol{S}}_A^{\phi}$, $\boldsymbol{L}_A^{\phi}$, $\bar{\boldsymbol{L}}_A^{\phi}$) correspond to $\boldsymbol{S}_x^{\phi}$, $\bar{\boldsymbol{S}}_x^{\phi}$, $\boldsymbol{L}_x^{\phi}$, $\bar{\boldsymbol{L}}_x^{\phi}$ ($\boldsymbol{S}_y^{\phi}$, $\bar{\boldsymbol{S}}_y^{\phi}$, $\boldsymbol{L}_y^{\phi}$, $\bar{\boldsymbol{L}}_y^{\phi}$), respectively, and are computed on the basis of equations explained in Section **3**. As solutions of Eq. (33), we can derive some eigenvectors as $\boldsymbol{\alpha}_{E_i}$ and $\boldsymbol{\alpha}_{A_j}$, respectively. Then we obtain $\boldsymbol{A}_E = [\boldsymbol{\alpha}_{E_1}, \boldsymbol{\alpha}_{E_2}, \cdots, \boldsymbol{\alpha}_{E_D}] \in \mathbb{R}^{N \times D}$ and $\boldsymbol{A}_A = [\boldsymbol{\alpha}_{A_1}, \boldsymbol{\alpha}_{A_2}, \cdots, \boldsymbol{\alpha}_{A_D}] \in \mathbb{R}^{N \times D}$ by extracting the D coefficient vectors for which the corresponding values of λ_i ($i = 1,2,\cdots,D$) are larger than the others. In this way, we can derive projections $\boldsymbol{P}_E$ and $\boldsymbol{P}_A$ that can transform EEG and audio features into subspaces having the maximum correlation and consider the local structures and class information by using $\boldsymbol{A}_E$ and $\boldsymbol{A}_A$, respectively, as:

$$\boldsymbol{P}_E = \boldsymbol{\Phi}_E\boldsymbol{H}\boldsymbol{A}_E, \tag{38}$$

$$\boldsymbol{P}_A = \boldsymbol{\Phi}_A\boldsymbol{H}\boldsymbol{A}_A. \tag{39}$$

We use this projection $\boldsymbol{P}_A$ to calculate novel audio features reflecting a user's music preference in our method. Specifically, given a new audio feature vector $\boldsymbol{x}^{A_{test}}$, its projected audio feature vector is obtained as:

$$\begin{aligned} \boldsymbol{z}^{A_{test}} &= \boldsymbol{P}_A^{\mathrm{T}}\{\phi_A(\boldsymbol{x}^{A_{test}}) - \bar{\phi}_A\} \\ &= \boldsymbol{A}_A{}^{\mathrm{T}}\boldsymbol{H}\boldsymbol{\Phi}_A^{\mathrm{T}}\{\phi_A(\boldsymbol{x}^{A_{test}}) - \frac{1}{N}\boldsymbol{\Phi}_A\boldsymbol{1}\} \\ &= \boldsymbol{A}_A{}^{\mathrm{T}}\boldsymbol{H}\{\boldsymbol{\Phi}_A^{\mathrm{T}}\phi_A(\boldsymbol{x}^{A_{test}}) - \frac{1}{N}\boldsymbol{K}_A\boldsymbol{1}\}, \end{aligned} \tag{40}$$

where $\bar{\boldsymbol{\phi}}_A = \frac{1}{N}\boldsymbol{\Phi}_A\boldsymbol{1}$ is the mean vector of $\boldsymbol{\phi}_A(\boldsymbol{x}_i^{A_{train}})$ ($i = 1,2,\cdots,N_{train}$).

Since KDLPCCA can deal with not only a non-linear correlation but also local structures and class information of original

feature sets, *i.e.*, audio and EEG features extracted from audio signals and the user's EEG signals during listening to corresponding musical pieces, calculation of an effective projection for each user's music preference is expected. Furthermore, our method has a high level of practicability since the projected audio features $z^{A_{test}}$ can be obtained from new musical pieces without the user's EEG signals as shown in Eq. (40) if $\boldsymbol{P}_A$ has already been calculated.

4.3 Favorite Music Classification

In this subsection, we explain favorite music classification using the KDLPCCA-based projected audio features. In our method, an unknown label of a projected audio feature vector $z^{A_{test}}$ is estimated by using the following hyperplane:

$$f(z^{A_{test}}) = \boldsymbol{w}^{\mathrm{T}} z^{A_{test}} + b. \tag{41}$$

Given a training music database $\mathcal{D} = \{(z_l^{A_{train}}, y_l)\}_{l=1}^{N_{train}}$, $\boldsymbol{w}$ and b can be obtained by solving the SVM formulation [55]. Note that $z_l^{A_{train}}$ is calculated from the original training audio feature vector $\boldsymbol{x}_l^{A_{train}}$ in the same way as Eq. (40), and $y_l \in \{+1, -1\}$ denotes a user's music preference, *i.e.*, "Like" or "Dislike". Furthermore, a linear kernel is used for SVM since a non-linear kernel has already been used in KDLPCCA as shown in the previous section.

In this way, effective classification of favorite musical pieces is realized for each user adaptively since the projection $\boldsymbol{P}_A$ can derive the projected audio features having the best KDLPCCA-based correlation with the user's EEG signals.

5 Experiments

Experimental results of our favorite music classification using KDLPCCA are shown in this section. The performance of KDLPCCA using a well-known benchmark dataset, Multiple Features Datase (MFD), was also investigated, and the results are shown in Supplemental Material.

Important factors in kernel-based algorithms are determination of the type of kernel and its related parameters. In our experiments, we adopted the Gaussian kernel, *i.e.*, $k_x(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp^{-\|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2/2\sigma_x^2}$ and $k_y(\boldsymbol{y}_i, \boldsymbol{y}_j) = \exp^{-\|\boldsymbol{y}_i - \boldsymbol{y}_j\|^2/2\sigma_y^2}$. Here, the kernel widths σ_x^2 and σ_y^2 were chosen by searching the following parameter space: $\sigma_x^2 \in [2^{-19.5}, 2^{-15}, 2^{-10.5}, 2^{-6}, 2^{-1.5}, 2^3]$ and $\sigma_y^2 \in [2^{-19.5}, 2^{-15}, 2^{-10.5}, 2^{-6}, 2^{-1.5}, 2^3]$. We also searched the following space to obtain the optimal regularization parameters of KDLPCCA, *i.e.*, ξ_x and ξ_y in Eqs. (25) and (26): $\xi_x \in [10^{-3}, 10^{-2}, 10^{-1}, 10^0]$ and $\xi_y \in [10^{-3}, 10^{-2}, 10^{-1}, 10^0]$. In order to control the relative contributions of $\boldsymbol{C}_w^{\phi_d}$ and $\boldsymbol{C}_b^{\phi_d}$ in KDLPCCA, a balancing parameter η is introduced in Eq. (22). In our experiments, we experimentally used $\eta \in [0.5, 1.0]$. Furthermore, we searched the parameter space of SVM's penalty $C \in [2^{-5}, 2^{-2.5}, 2^0, 2^{2.5}, 2^5, 2^{7.5}, 2^{10}, 2^{12.5}, 2^{15}]$.

We used 10 healthy subjects (See **5.1**), a number that we considered to be sufficient since less than 10 subjects were used in recent studies using EEG: 4 subjects in [83] and [84], 5 subjects in [85] and [86], and 6 subjects in [87]. Furthermore, we used 60 musical pieces as a music dataset, each of which was 15 sec in length[4]. In some related works using biosignals obtained from a user during listening to music, the same setting, *i.e.*, listening to music for 15 sec [46], [47], [88], [89], [90], and it has been reported that the average time for a listener to start making emotional judgements for musical pieces is 8.31 seconds [91]. Therefore, we also set a period of 15 seconds for observing EEG signals from a subject listening to a musical piece. All of the subjects evaluated each musical piece by 5 levels, *i.e.*, 5 (like very much), 4 (like), 3 (undecided), 2 (do not like) and 1 (do not like at all). Therefore, audio feature vectors could be grouped with respect to two classes, *i.e.*, "Like" and "Dislike". The class "Like" consisted of audio feature vectors corresponding to the musical pieces rated 5 or 4 by a subject, and the class "Dislike" consisted of audio feature vectors corresponding to the musical pieces rated 2 or 1 by a subject. Therefore, we regarded the class "Like" as favorite musical pieces in our experiment. Note that we did not use musical pieces rated 3 by subjects. These experiment conditions were used in related studies [46], [47], and we thus regarded the number of musical pieces as being sufficient for our purpose. We equalized the number of musical pieces of each class per subject by randomly excluding some musical pieces from a class with more musical pieces in order to prevent an imbalance problem [92], [93], [94]. Thus, N_{music}, which is the total number of musical pieces used in our experiment, is slightly different according to each user, and the details about this are shown in TABLE 4. The task implemented by each subject in our experiment is shown in Fig. 2. As shown in Fig. 2, a subject rated each musical piece after listening to each one.

4. The details of the music dataset used in our experiments are presented in Supplemental Material.

TABLE 4: Details of the dataset used in the experiment per subject.

Subject	Number of LIKE music	Number of DISLIKE music	TOTAL ($=N_{music}$)
A	23	23	46
B	23	23	46
C	17	17	34
D	20	20	40
E	16	16	32
F	23	23	46
G	21	21	42
H	18	18	36
I	25	25	50
J	25	25	50

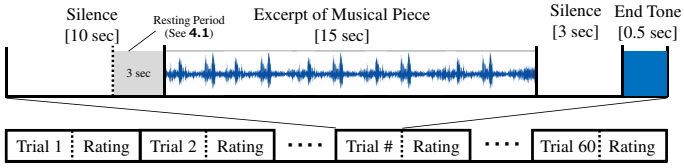


Fig. 2: The task that each subject was required to perform in the experiment.

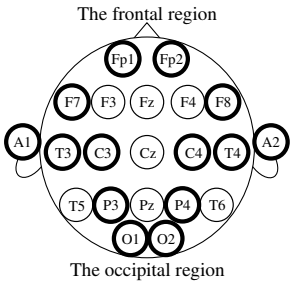
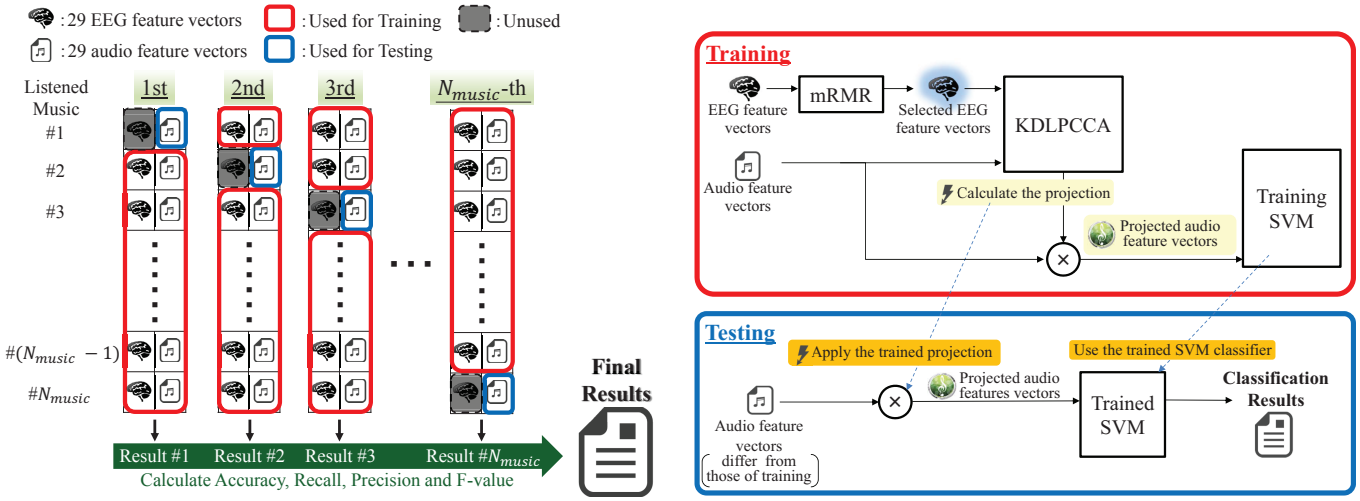


Fig. 3: Electrode layout of the international 10-20 system. Channels shown by thick lines were used in our experiments. They were decided on the basis of [45], [46].



(a) Details of the leave-trial out validation conducted for each subject in our experiments. The 29 feature vectors of EEG and audio are respectively extracted from a musical piece since it is divided into 29 segments by the preprocessing (See **4.1**).

(b) Details of the training and testing in our experiments. We repeated the training and testing using each combination of parameters to find the best set of parameters.

Fig. 4: An overview of our validation and parameter determination.

5.1 EEG Signal Collection and Experimental Procedures

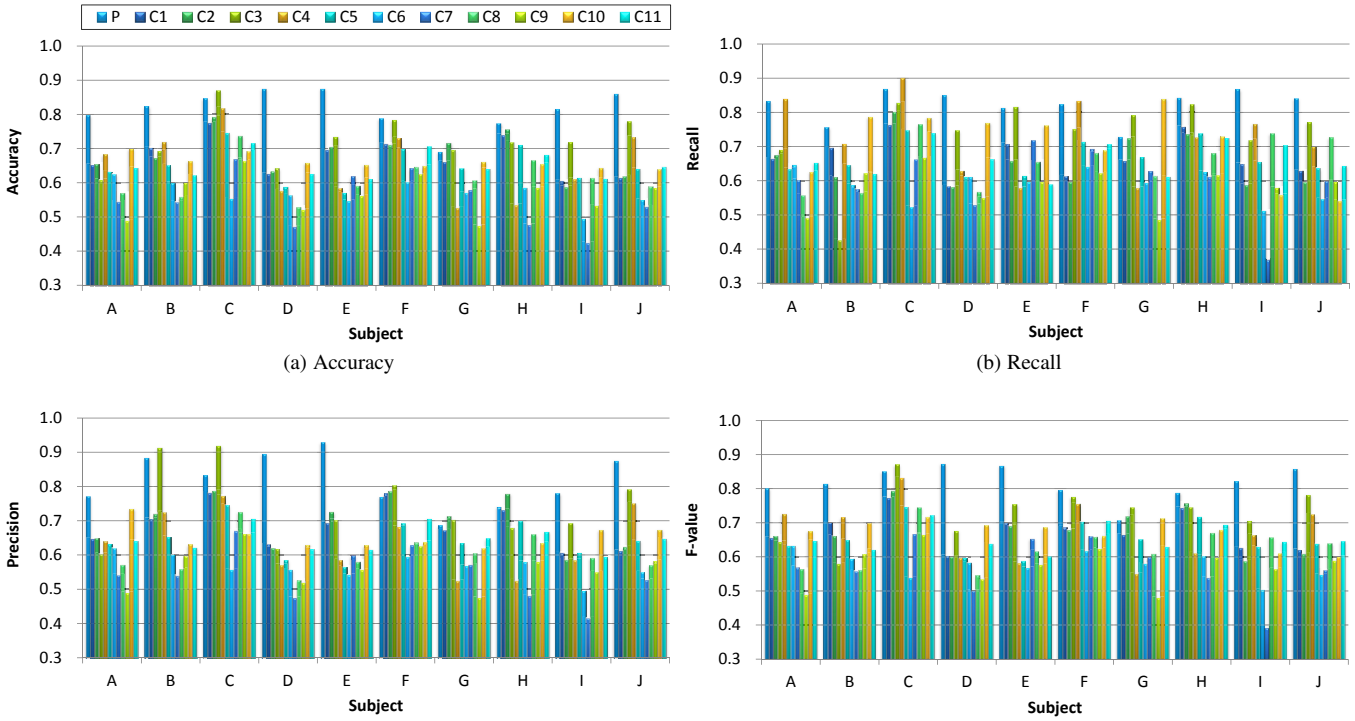
EEG signals in our experiment were collected from 10 healthy subjects while they were listening to musical pieces. The average age of the subjects was about 23 years. All of the subjects in our experiments were males and had no experience of musical training. Therefore, it is considered that our method is effective for favorite music classification even if a target user has no specialized music training. We recorded EEG signals from 12 channels (Fp1, Fp2, F7, F8, C3, C4, P3, P4, O1, O2, T3 and T4) according to the international 10-20 system shown in Fig. 3. We chose these 12 channels based on the study by Lin *et al.* [44]. Lin *et al.* classified emotional states of humans into four states (Joy, Anger, Sadness and Pleasure) by using many types of EEG features, *i.e.*, changing the channel and the frequency band to perform the classification. Then they investigated the degree of use of each channel in the top 30 results across the 26 subjects. The important regions for observing affective phenomena corresponded to the channels of Fp1, Fp2, F7, F8, C3, C4, P3, P4, O1, O2, T3 and T4, which are used in our method (See Fig. 3). Thus, we adopted these 12 EEG channels to observe affective phenomena of humans, *i.e.*, individual music preference in the case of our research. Since EEG signals are weak, we amplified the signals by using 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 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 while they were listening to the musical pieces.

5.2 Results of Classification of Favorite Musical Pieces

In our experiment, the lengths of an EEG segment and an overlapping segment were 1.0 and 0.5 sec, respectively. Since CCA needs the same number of samples from two heterogeneous sets, the lengths of an audio segment and an overlapping segment to calculate the mean and standard deviation of audio features were also in the same as those of EEG. Furthermore, we divided all of the audio feature vectors per musical piece in order to prevent overfitting caused by learning similar vectors extracted from the same musical piece, *i.e.*, we employed a leave-trial-out validation. A trial is corresponding to listening to one musical piece. In our experiments, the testing data consisted of 29 vectors obtained from the same musical piece. Therefore, the training data did not contain samples that were similar to those of the testing data. The details of the validation explained above are shown in Fig. 4(a). We employed linear kernels for all SVMs since we have already applied non-linear kernels to the EEG and audio features in the step of KDLPCCA. The parameters used in SVM were determined via Grid Search [95] based on final accuracy of the classification. Namely, we determined the kernel parameters of SVM and KDLPCCA upon the best classification accuracy for training data only. Specifically, we firstly divided the dataset as

TABLE 5: Details of each method in our experiment: abbreviated title of each method, kinds of features used and dimensions of the features.

METHOD	FEATURE	DIMENSION
P (proposed)	audio features projected by KDLPCCA	2
C1	audio features projected by KLDCCA	2
C2	audio features projected by KDCCA	2
C3	audio features projected by KCCA	2
C4	audio features projected by DLPCCA	2
C5	audio features projected by LDCCA	2
C6	audio features projected by DCCA	2
C7	audio features projected by LPCCA	2
C8	audio features projected by CCA	2
C9	Original audio features	151
C10	Selected original EEG features	depending on each subject
C11	Selected audio features by our previous method [52]	depending on each subject



(a) Accuracy (b) Recall

(c) Precision (d) F-value

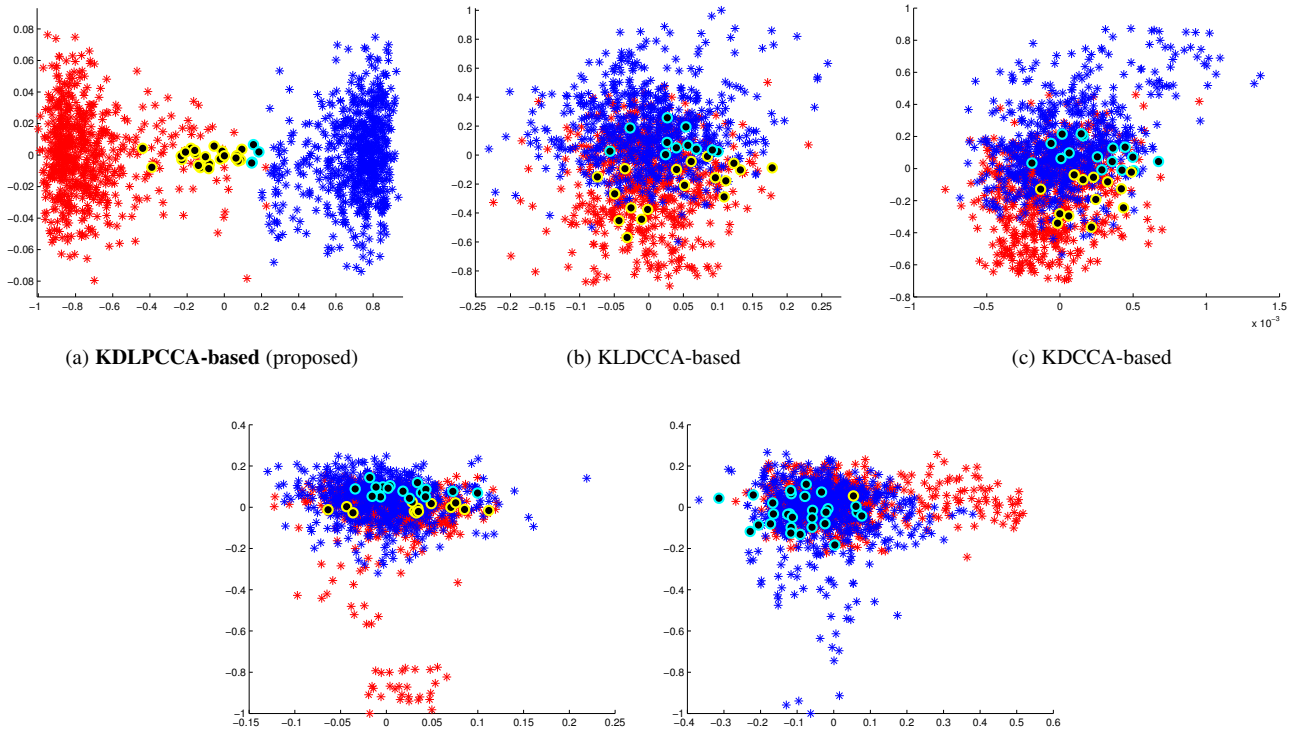
Fig. 5: Results of favorite music classification. We selected the EEG features' dimensions recording the best accuracy as Comparative method 10 (C10; using only EEG features) and also used the same dimensions of EEG features before applying the CCAs for each subject. On the other hand, after the projecting, DCCA and KDCCA are limited to extract the number of dimensions equaling the number of classes (= 2) due to their formulas [58], and we thus used two dimensions of projected audio features in all CCAs (C1-C8).

TABLE 6: Averaged results for all subjects regarding favorite music classification.

	P (proposed)	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11
Accuracy	**0.814**±0.06	0.678±0.06	0.684±0.06	0.724±0.07	0.651±0.10	0.649±0.05	0.568±0.04	0.549±0.08	0.610±0.06	0.563±0.06	0.661±0.02	0.650±0.04
Recall	**0.822**±0.05	0.671±0.06	0.656±0.08	0.736±0.12	0.726±0.11	0.666±0.05	0.587±0.05	0.598±0.10	0.655±0.08	0.581±0.06	0.708±0.10	0.665±0.05
Precision	**0.816**±0.08	0.685±0.06	0.699±0.07	0.742±0.11	0.635±0.09	0.645±0.06	0.565±0.04	0.544±0.08	0.602±0.06	0.563±0.06	0.652±0.03	0.646±0.04
F-value	**0.817**±0.05	0.677±0.05	0.675±0.07	0.727±0.08	0.675±0.09	0.655±0.05	0.576±0.04	0.569±0.08	0.626±0.06	0.572±0.06	0.673±0.04	0.655±0.04

shown in Fig. 4(a) and next searched the parameter spaces that have been shown in the second paragraph of this section. Note that we also applied the mRMR feature selection to the set of EEG features obtained from only training music trials per each leave-trial-out validation as shown in Fig. 4(b). Therefore, the test data did not contain any EEG feature vectors and were entirely disjointed from the parameter optimization. Then we calculated classification accuracy regarding favorite music classification for each combination of parameters. Eventually, we determined the best combination of parameters based on all of the calculated results. To evaluate the performance of our method, we used Accuracy, Recall, Precision and F-value. We used the following ten comparative methods (C1-C11): methods using KLDCCA (C1), KDCCA (C2), KCCA (C3), DLPCCA (C4), LDCCA (C5), DCCA (C6), LPCCA (C7) and standard CCA (C8) instead of KDLPCCA, a method that uses original audio features, *i.e.*, without the any CCA-based projection, (C9) and a method that uses only subjects' EEG features selected by the mRMR algorithm (C10). We also used our previous method [52] that is realized by KDLPCCA-based audio feature selection as C11. Since DCCA and KDCCA are limited to extract the number of dimensions equaling the number of classes (= 2) due to their formulas [58], we used 2 dimensions of projected audio features in all CCAs (C1-C8). The details of the proposed method and all comparative methods are summarized in TABLE 5.

The results are shown in Fig. 5 and TABLE 6. Since we confirmed that all of the results for C10 were better than those of C9, one of our most important ideas, *i.e.*, utilizing individual EEG features for favorite music classification, is valid. Actually, all of the methods utilizing projected audio features (P and C1-C8) outperformed the method using only original audio features (C9) as shown in TABLE 6. Among the comparative methods, the results obtained by using kernelized CCAs (C1-C3) were much better than the results obtained by using corresponding non-kernelized CCAs (C4-C8). Therefore, we can argue that another idea, *i.e.*, considering not only linear but also non-linear correlations between EEG and audio features for favorite music classification, is important. However, from Fig. 5 and TABLE 6, we can confirm that C1-C8 do not significantly outperform the method using only original EEG features (C10). In other words, C1-C8 are equal to C10 at the most. Meanwhile, we confirmed that our proposed method using KDLPCCA outperforms even C10 in addition to C1-C9. This is because KDLPCCA can deal with not only a non-linear correlation but also local structures and class information of original data as shown in TABLE 1. In fact, in non-kernelized methods (C4-C8), DLPCCA (C4) also outperforms C5-C8 as shown in TABLE 6 since DLPCCA can deal with local structures and class information. In addition, we confirmed that the hypothesis stated in Section **1**, *i.e.*, that KDLPCCA-based projected audio features are more suitable for the user's



(a) **KDLPCCA-based** (proposed) (b) KLDCCA-based (c) KDCCA-based

(d) KCCA-based (e) CCA-based

Fig. 6: Visualization results using each CCA's 1st projected audio features (horizontal axis) and 2nd projected audio features (vertical axis). The results ware obtained from the same training and testing datasets, *i.e.*, subject A's EEG signals and the musical pieces liked/disliked by subject A. RED $*$ denotes training samples obtained from the musical pieces that were liked, while BLUE $*$ denotes training samples obtained from the musical pieces that were disliked. Circles with yellow edges denote testing samples classified as a correct class, and circles with cyan edges denote testing samples classified as an incorrect class. Note that true classes of all test samples are RED classes.

music preference than KDLPCCA-based selected audio features, is correct since the result of P is superior to that for C11.

Next, we visualize subject A's projected samples (See Fig. 6) using some CCAs: our KDLPCCA, KLDCCA, KDCCA, KCCA and standard CCA since KCCA and CCA are benchmarks, and KLDCCA and KDCCA showed high performances in previous experiments. We used the same training and testing samples for the visualization of all CCAs. In Fig. 6, RED $*$ denotes training samples obtained from the musical pieces liked by subject A, and BLUE $*$ denotes training samples obtained from the musical pieces disliked by subject A. Circles with yellow edges denote testing samples classified as a correct class, and circles with cyan edges denote testing samples classified as an incorrect class. Note that all testing samples' true class is "Like", and they thus should be projected into the RED region and become circles with yellow edges, ideally. A comparison of Figs. 6(d) and 6(e) shows that KCCA has better discriminative power than that of standard CCA since KCCA has more testing samples classified as a correct class, *i.e.*, circles with yellow edges, than does CCA. Hence, using the kernel trick for CCA is effective for enhancing discriminative power by considering the correlation between EEG signals and audio signals. However, the KCCA's result cannot separate the RED and BLUE training samples well since KCCA does not consider class information. This deteriorates the discriminative power of KCCA-based projected audio features. On the other hand, the results obtained by using KLDCCA and KDCCA can classify test samples better than the results obtained by using KCCA as shown in Figs. 6(b) and 6(c). In fact, there are more circles with yellow edges in Figs. 6(b) and 6(c) than in Fig. 6(d). However, from Figs. 6(b) and 6(c), we can confirm that the regions of almost all RED and BLUE training samples still overlap. Meanwhile, our novel CCA, *i.e.*, KDLPCCA, can separate the training samples much better than can KLDCCA or KDCCA and can classify the test samples more correctly than can the other CCAs as shown in Fig. 6(a) since KDLPCCA can consider not only a non-linear correlation and class information but also local structures of original data as shown in TABLE 1.

For CCA-based audio features (C1-C8), the results of our KDLPCCA, KLDCCA and KDCCA are better than those of other CCAs since these CCAs can deal with class information and a non-linear correlation simultaneously. Among these three CCAs, the results using our KDLPCCA are notably better. The major difference between our KDLPCCA and KLDCCA and KDCCA is the consideration of local structures of original data based on LPP. Generally, it is considered that there are both similar and non-similar musical pieces in the same class, *i.e.*, a multimodal class. As Sugiyama pointed out in [54], locality preserving approaches can work well with data having a multimodal class. Therefore, consideration of the similarity among samples is important for favorite music classification even if the samples have the same labels. KDLPCCA can consider the similarity among samples by using Laplacian matrices based on LPP, while KLDCCA and

KDCCA cannot consider the similarity. Therefore, we consider that KDLPCCA is a powerful tool for extraction of features for favorite music classification.

The results presented in this subsection show that our KDLPCCA-based approach is powerful and effective for favorite music classification since KDLPCCA can extract the non-linear correlation between user's EEG and audio features considering class information and local structures of original data.

6 Future Work

In the field of neuroscience, considering the profiles of subjects, especially gender, is very important and desirable. In fact, there are some reports that there is a difference of EEG signals during listening to music between a male and a female. However, it has been reported that the difference is minor since it can be observed during listening to limited musical pieces, *i.e.*, only having song lyrics [96]. In our music dataset, most of musical pieces had no song lyrics. Furthermore, most of song lyrics were not subjects' mother tongue even if there are few musical pieces having duration with song lyrics in our music dataset. In addition, Miles *et al.* have recently reported that a female has the advantage at recognizing familiar melodies, *i.e.*, storing and retrieving knowledge about specific melodies, since the hippocampus seems to develop at a faster rate in girls than in boys [97]. Therefore, we consider that dealing with a comprehensiveness of a dataset is an enormous and important future work which should devote the time sufficiently. Namely, in the future work, it is necessary that re-collecting the dataset by considering the gender balance, age, their mother tongue, familiar or non-familiar melodies, experience of musical training and so on.

7 Conclusions

In this paper, we have proposed a novel audio feature projection using KDLPCCA-based correlation with EEG features for favorite music classification. The proposed method calculates new projected audio features that are suitable for representing a user's music preference by applying our novel CCA, *i.e.*, KDLPCCA, to audio features and EEG features that are extracted from the user's EEG signals during listening to musical pieces. Since EEG features reflect individual music preference, favorite music classification that adaptively considers a user's preference becomes feasible via an SVM classifier using the projected audio features. Since our method does not need acquisition of EEG signals for obtaining new audio features from new musical pieces after calculating the projection, our method has a high level of practicability. The experimental results show that (1) KDLPCCA is a globally powerful tool for extracting effective features for classification (See Supplemental Material) and (2) our projection method via KDLPCCA can reflect individual music preference and thus realize favorite music classification effectively. In addition, we consider that existing methods using audio features such as the methods in [24], [25] may be enhanced by using new projected audio features obtained by our method.

Acknowledgements

This work was partly supported by JSPS KAKENHI Grant Numbers JP17H01744, JP15K12023.

References

[1] G. Tzanetakis and P. Cook, "Musical genre classification of audio signals," *IEEE Trans. on Speech and Audio Processing*, vol. 10, no. 5, pp. 293–302, July 2002.

[2] T. Li, M. Ogihara, and Q. Li, "A comparative study on content-based music genre classification," in *Proc. of the 26th Annual International ACM SIGIR Conference on Research and Development in Informaion Retrieval*, 2003, pp. 282–289.

[3] I. Mierswa and K. Morik, "Automatic feature extraction for classifying audio data.," *Machine Learning*, vol. 58, no. 2-3, pp. 127–149, 2005.

[4] D. Turnbull and C. Elkan, "Fast recognition of musical genres using rbf networks," *IEEE Trans. on Knowledge and Data Engineering*, vol. 17, no. 4, pp. 580–584, April 2005.

[5] A. Meng, P. Ahrendt, J. Larsen, and L. Hansen, "Temporal feature integration for music genre classification," *IEEE Trans. on Audio, Speech, and Language Processing*, vol. 15, no. 5, pp. 1654–1664, July 2007.

[6] Y. Song and C. Zhang, "Content-based information fusion for semi-supervised music genre classification," *IEEE Trans. on Multimedia*, vol. 10, no. 1, pp. 145–152, Jan. 2008.

[7] C.-H. Lee, J.-L. Shih, K.-M. Yu, and H.-S. Lin, "Automatic music genre classification based on modulation spectral analysis of spectral and cepstral features," *IEEE Trans. on Multimedia*, vol. 11, no. 4, pp. 670–682, June 2009.

[8] T. L. Nwe and H. Li, "Exploring vibrato-motivated acoustic features for singer identification," *IEEE Trans. on Audio, Speech, and Language Processing*, vol. 15, no. 2, pp. 519–530, Feb. 2007.

[9] J. Shen, J. Shepherd, B. Cui, and K. L. Tan, "A novel framework for efficient automated singer identification in large music databases," *ACM Trans. on Information Systems*, vol. 27, no. 3, pp. 1–31, May 2009.

[10] T. Li and M. Ogihara, "Toward intelligent music information retrieval," *IEEE Trans. on Multimedia*, vol. 8, no. 3, pp. 564–574, June 2006.

[11] 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.

[12] L. Mion and G. De Poli, "Score-independent audio features for description of music expression," *IEEE Trans. on Audio, Speech, and Language Processing*, vol. 16, no. 2, pp. 458–466, Feb. 2008.

[13] J. Wang, Y. Lee, Y. Chin, Y. Chen, and W. Hsieh, "Hierarchical dirichlet process mixture model for music emotion recognition," *IEEE Trans. on Affective Computing*, vol. PP, no. 99, pp. 1–1, 2015.

[14] G. Agostini, M. Longari, and E. Pollastri, "Musical instrument timbres classification with spectral features," *EURASIP Journal on Advances in Signal Processing*, vol. 2003, pp. 5–14, Jan. 2003.

[15] S. Essid, G. Richard, and B. David, "Musical instrument recognition by pairwise classification strategies," *IEEE Trans. on Audio, Speech, and Language Processing*, vol. 14, no. 4, pp. 1401–1412, July 2006.

[16] S. Essid, G. Richard, and B. David, "Instrument recognition in polyphonic music based on automatic taxonomies," *IEEE Trans. on Audio, Speech, and Language Processing*, vol. 14, no. 1, pp. 68–80, Jan. 2006.

[17] T. Kitahara, M. Goto, K. Komatani, T. Ogata, and H. G. Okuno, "Instrument identification in polyphonic music: Feature weighting to minimize influence of sound overlaps," *EURASIP Journal on Advances in Signal Processing*, vol. 2007, no. 1, pp. 155–155, 2007.

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

[19] L. Barrington, D. Turnbull, M. Yazdani, and G. Lanckriet, "Combining audio content and social context for semantic music discovery," in *Proc. of the 32th annual international ACM SIGIR conference on Research and development in information retrieval*, 2009.

[20] P. Åman and L. A. Liikkanen, "A survey of music recommendation aids," in *Proc. of Workshop on Music Recommendation And Discovery (WOMRAD)*, Sept. 2010.

[21] Y. Song, S. Dixon, and M. Pearce, "A survey of music recommendation systems and future perspectives," in *9th International Symposium on Computer Music Modelling and Retrieval (CMMR)*, 2012.

[22] J. L. Herlocker, J. A. Konstan, L. G. Terveen, John, and T. Riedl, "Evaluating collaborative filtering recommender systems," *ACM Trans. on Information Systems*, vol. 22, pp. 5–53, 2004.

[23] B. Logan and A. Salomon, "A content-based music similarity function," *Technical report, Compaq Cambridge Research Lab*, 2001.

[24] P. Chiliguano and G. Fazekas, "Hybrid music recommender using content-based and social information," in *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2016, pp. 2618–2622.

[25] K. Benzi, V. Kalofolias, X. Bresson, and P. Vandergheynst, "Song recommendation with non-negative matrix factorization and graph total variation," in *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2016, pp. 2439–2443.

[26] T. Kamishima, "Exploiting customer's preference -leading edge of user profiling technique- : Problems for collaborative filtering : Privacy, shilling attack, and variability of users' ratings," *Information Processing Society of Japan (IPSJ) Magazine*, vol. 48, no. 9, pp. 966–971, 2007.

[27] T. Kamishima, "Algorlthms for recommender systems （1）," *Trans. of the Japanese Society for Artificial Intelligence*, vol. 26, no. 6, pp. 826–837, 2007.

[28] T. Kamishima, "Algorlthms for recommender systems （2）," *Trans. of the Japanese Society for Artificial Intelligence*, vol. 23, no. 1, pp. 89–103, 2008.

[29] T. Kamishima, "Algorlthms for recommender systems （3）," *Trans. of the Japanese Society for Artificial Intelligence*, vol. 23, no. 2, pp. 248–263, 2008.

[30] G. Häubl and K. B. Murray, "Recommending or persuading? the impact of a shopping agent's algorithm on user behavior," in *Proc. of the 3rd ACM Conference on Electronic Commerce*, 2001, pp. 163–170.

[31] M. F. Alhamid, M. Rawashdeh, H. Dong, M. A. Hossain, and A. E. Saddik, "Exploring latent preferences for context-aware personalized recommendation systems," *IEEE Trans. on Human-Machine Systems*, vol. 46, no. 4, pp. 615–623, Aug. 2016.

[32] W. Zhang, G. Ding, L. Chen, C. Li, and C. Zhang, "Generating virtual ratings from chinese reviews to augment online recommendations," *ACM Trans. on Intell. Syst. Technol.*, vol. 4, no. 1, pp. 9:1–9:17, Feb. 2013.

[33] L. C. Wang, X. Y. Zeng, L. Koehl, and Y. Chen, "Intelligent fashion recommender system: Fuzzy logic in personalized garment design," *IEEE Trans. on Human-Machine Systems*, vol. 45, no. 1, pp. 95–109, Feb. 2015.

[34] B. Guo, D. Zhang, D. Yang, Z. Yu, and X. Zhou, "Enhancing memory recall via an intelligent social contact management system," *IEEE Trans. on Human-Machine Systems*, vol. 44, no. 1, pp. 78–91, Feb. 2014.

[35] H. Zhu, E. Chen, H. Xiong, K. Yu, H. Cao, and J. Tian, "Mining mobile user preferences for personalized context-aware recommendation," *ACM Trans. on Intell. Syst. Technol.*, vol. 5, no. 4, pp. 58:1–58:27, Dec. 2014.

[36] D. A. Adamos, S. I. Dimitriadis, and N. A. Laskaris, "Towards the bio-personalization of music recommendation systems: A single-sensor EEG biomarker of subjective music preference," *Information Sciences*, vol. 343?344, pp. 94–108, 2016.

[37] M. F. Alhamid, M. Rawashdeh, H. Al Osman, and A. El Saddik, "Leveraging biosignal and collaborative filtering for context-aware recommendation," in *Proc. of the 1st ACM International Workshop on Multimedia Indexing and Information Retrieval for Healthcare*, 2013, pp. 41–48.

[38] S. Shin, D. Jang, J. J. Lee, S. J. Jang, and J. H. Kim, "Mymusicshuffler: Mood-based music recommendation with the practical usage of brain-wave signals," in *Proc. of the 2014 IEEE International Conference on Consumer Electronics (ICCE)*, 2014, pp. 355–356.

[39] C. T. Lin, L. W. Ko, J. C. Chiou, J. R. Duann, R. S. Huang, S. F. Liang, T. W. Chiu, and T. P. Jung, "Noninvasive neural prostheses using mobile and wireless EEG," *Proc. of the IEEE*, vol. 96, no. 7, pp. 1167–1183, July 2008.

[40] Y. M. Chi and G. Cauwenberghs, "Wireless non-contact EEG/ECG electrodes for body sensor networks," in *Proc. of 2010 International Conference on Body Sensor Networks*, 2010, pp. 297–301.

[41] L.-D. Liao, C.-Y. Chen, I.-J. Wang, S.-F. Chen, S.-Y. Li, B.-W. Chen, J.-Y. Chang, and C.-T. Lin, "Gaming control using a wearable and wireless EEG-based brain-computer interface device with novel dry foam-based sensors," *Journal of NeuroEngineering and Rehabilitation*, vol. 9, no. 1, p. 5, 2012.

[42] E. Moser, M. Meyerspeer, F. P. S. Fischmeister, G. Grabner, H. Bauer, and S. Trattnig, "Windows on the human body in vivo high-field magnetic resonance research and applications in medicine and psychology," *Sensors*, vol. 10, no. 6, pp. 5724–5757, 2010.

[43] A. Gevins, J. Le, N. K. Martin, P. Brickett, J. Desmond, and B. Reutter, "High resolution EEG: 124-channel recording, spatial deblurring and MRI integration methods," *Electroencephalography and Clinical Neurophysiology*, vol. 90, no. 5, pp. 337–358, 1994.

[44] 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 Engineerring*, vol. 57, no. 7, pp. 1798–1806, 2010.

[45] 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.

[46] 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.

[47] 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.

[48] 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.

[49] W.-C. Lin, H.-W. Chiu, and C.-Y. Hsu, "Discovering EEG signals response to musical signal stimuli by time-frequency analysis and independent component analysis," in *Proc. of the IEEE 27th Annual International Conference of the Engineering in Medicine and Biology Society (EMBS)*, 2005, pp. 2765–2768.

[50] S. Koelstra, C. Muhl, M. Soleymani, J. S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "Deap: A database for emotion analysis ;using physiological signals," *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 18–31, Jan. 2012.

[51] H. Hotelling, "Relations between two sets of variates," *Biometrika*, vol. 28, no. 3, pp. 321–377, 1936.

[52] R. Sawata, T. Ogawa, and M. Haseyama, "Novel favorite music classification using EEG-based optimal audio features selected via KDLPCCA," in *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2016, pp. 759–763.

[53] Y.-R. Yeh, C.-H. Huang, and Y.-C. F. Wang, "Heterogeneous domain adaptation and classification by exploiting the correlation subspace," *IEEE Trans. on Image Process*, vol. 23, no. 5, pp. 2009–2018, 2014.

[54] M. Sugiyama, "Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis," *J. Mach. Learn. Res.*, vol. 8, pp. 1027–1061, May 2007.

[55] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, Sept. 1995.

[56] S. Akaho, "A kernel method for canonical correlation analysis," in *Proc. of the International Meeting of the Psychometric Society (IMPS)*, 2001.

[57] T. Sun and S. Chen, "Locality preserving CCA with applications to data visualization and pose estimation," *Image and Vision Computing*, vol. 25, no. 5, pp. 531–543, 2007.

[58] T.-K. Sun, S.-C. Chen, Z. Jin, and J.-Y. Yang, "Kernelized discriminative canonical correlation analysis," in *Proc. of International Conference on Wavelet Analysis and Pattern Recognition (ICWAPR)*, Nov. 2007, vol.3, pp. 1283–1287.

[59] 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.

[60] 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.

[61] J. Blitzer, D. Foster, and S. Kakade, "Domain adaptation with coupled subspaces," in *Proc. of the Conference on Artificial Intelligence and Statistics*, 2011.

[62] H. Ohkushi, T. Ogawa, and M. Haseyama, "Music recommendation according to human motion based on kernel CCA-based relationship," *EURASIP Journal on Advances in Signal Processing*, vol. 2011, p. 121, 2011.

[63] W. Li, L. Duan, D. Xu, and I. W. Tsang, "Learning with augmented features for supervised and semi-supervised heterogeneous domain adaptation," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 36, no. 6, pp. 1134–1148, June 2014.

[64] T. Ogawa and M. Haseyama, "2D semi-supervised CCA-based inpainting including new priority estimation," in *Proc. of IEEE International Conference on the Image Processing (ICIP)*, 2014, pp. 1837–1841.

[65] X. He and P. Niyogi, "Locality preserving projections," in *Proc. of Neural Information Processing Systems*, 2003, vol.16.

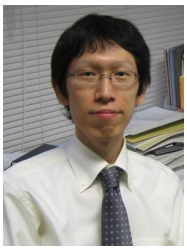
[66] T. Melzer, M. Reiter, and H. Bischof, "Appearance models based on kernel canonical correlation analysis," *Pattern Recognition*, vol. 36, no. 9, pp. 1961–1971, 2003.

[67] 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.

[68] 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.

[69] D. Sammler, M. Grigutsch, T. Fritz, and S. Koelsch, "Music and emotion: Electrophysiological correlates of the processing of pleasant and unpleasant music," *Psychophysiology*, vol. 44, no. 2, pp. 293–304, 2007.
[70] T. Baumgartner, M. Esslen, and L. Jancke, "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.
[71] S. Nakamura, N. Sadato, T. Oohashi, E. Nishina, Y. Fuwamoto, and Y. Yonekura, "Analysis of music-brain interaction with simultaneous measurement of regional cerebral blood flow and electroencephalogram beta rhythm in human subjects," *Neuroscience Letters*, vol. 275, no. 3, pp. 222–226, 1999.
[72] E. M. Whitham, K. J. Pope, S. P. Fitzgibbon, T. Lewis, C. R. Clark, S. Loveless, M. Broberg, A. Wallace, D. DeLosAngeles, P. Lillie, A. Hardy, R. Fronsko, A. Pulbrook, and J. O. Willoughby, "Scalp electrical recording during paralysis: Quantitative evidence that EEG frequencies above 20 hz are contaminated by EMG," *Clinical Neurophysiology*, vol. 118, no. 8, pp. 1877–1888, 2007.
[73] S. Yuval-Greenberg, O. Tomer, A. S. Keren, I. Nelken, and L. Y. Deouell, "Transient induced gamma-band response in EEG as a manifestation of miniature saccades," *Neuron*, vol. 58, no. 3, pp. 429–441, 2008.
[74] E. M. Whitham, T. Lewis, K. J. Pope, S. P. Fitzgibbon, C. R. Clark, S. Loveless, D. DeLosAngeles, A. K. Wallace, M. Broberg, and J. O. Willoughby, "Thinking activates EMG in scalp electrical recordings," *Clinical Neurophysiology*, vol. 119, no. 5, pp. 1166–1175, 2008.
[75] Y. Pan, C. Guan, J. Yu, K. K. Ang, and T. E. Chan, "Common frequency pattern for music preference identification using frontal EEG," in *Proc. of the 6th International IEEE/EMBS Conference on Neural Engineering (NER)*, Nov. 2013, pp. 505–508.
[76] G. Pfurtscheller and F. H. Lopes da Silva, "Event-related EEG/MEG synchronization and desynchronization: basic principles," *Clinical Neurophysiology*, vol. 110, no. 11, pp. 1842–1857, Nov. 1999.
[77] B. Graimann, J. Huggins, S. Levine, and G. Pfurtscheller, "Visualization of significant erd/ers patterns in multichannel EEG and ECoG data," *Clinical Neurophysiology*, vol. 113, no. 1, pp. 43–47, 2002.
[78] D. P. Allen and C. D. MacKinnon, "Time-frequency analysis of movement-related spectral power in EEG during repetitive movements: A comparison of methods," *Journal of Neuroscience Methods*, vol. 186, no. 1, pp. 107–115, 2010.
[79] 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.
[80] O. Lartillot and P. Toiviainen, "A matlab toolbox for musical feature extraction from audio," in *Proc. of the 10th International Conference on Digital Audio Effects*, 2007.
[81] 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.
[82] M. J. Alhaddad, M. I. Kamel, M. M. Makary, H. Hargas, and Y. M. Kadah, "Spectral subtraction denoising preprocessing block to improve P300-based brain-computer interfacing," *BioMedical Engineering OnLine*, vol. 13, no. 1, p. 36, 2014.
[83] A. Omidvarnia, G. Azemi, B. Boashash, J. O'Toole, P. Colditz, and S. Vanhatalo, "Measuring time-varying information flow in scalp EEG signals: Orthogonalized partial directed coherence," *IEEE Trans. on Biomedical Engineering*, vol. 61, no. 3, pp. 680–693, March 2014.
[84] 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.
[85] H. J. Baek, H. J. Lee, Y. G. Lim, and K. S. Park, "Conductive polymer foam surface improves the performance of a capacitive EEG electrode," *IEEE Trans. on Biomedical Engineering*, vol. 59, no. 12, pp. 3422–3431, Dec. 2012.
[86] T. Yamsa-ard and Y. Wongsawat, "The relationship between EEG and binaural beat stimulation in meditation," in *Proc. of the Biomedical Engineering International Conference (BMEiCON)*, Nov. 2014, pp. 1–4.
[87] W.-C. Lin, H.-W. Chiu, and C.-Y. Hsu, "Discovering EEG signals response to musical signal stimuli by time-frequency analysis and independent component analysis," in *Proc. of the IEEE 27th Annual International Conference of the Engineering in Medicine and Biology Society (EMBS)*, 2005, pp. 2765–2768.
[88] S. Koelsch, T. Fritz, V. D. Y. Cramon, K. Müller, and A. D. Friederici, "Investigating emotion with music: An fMRI study," *Human Brain Mapp.*, vol. 27, no. 3, pp. 239–250, 2006.
[89] D. Sammler, M. Grigutsch, T. Fritz, and S. Koelsch, "Music and emotion: Electrophysiological correlates of the processing of pleasant and unpleasant music," *Psychophysiol.*, vol. 44, pp. 293–304, 2007.
[90] L. I. Aftanas, N. V. Reva, A. A. Varlamov, S. V. Pavlov, and V. P. Makhnev, "Analysis of evoked EEG synchronization and desynchronization in conditions of emotional activation in humans: Temporal and topographic characteristics," *Neuroscience and Behavioral Physiology*, vol. 34, no. 8, pp. 859–867, 2004.
[91] J. Bachorik, M. Bangert, P. Loui, K. Larke, J. Berger, R. Rowe, and G. Schlaug, "Emotion in motion: Investigating the time-course of emotional judgments of musical stimuli," *Music Perception*, vol. 26, pp. 355–364, 2009.
[92] 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.
[93] G. Wu and E. Y. Chang, "Adaptive Feature-Space Conformal Transformation for Imbalanced Data Learning," in *Proc. of the Twentieth International Conference on Machine Learning*, 2003, pp. 816–823.
[94] 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.
[95] C.-W. Hsu, C.-C. Chang, and C.-J. Lin, "A practical guide to support vector classification," *Technical Report*, Department of Computer Science, 2003.
[96] S. Koelsch, B. Maess, T. Grossmann, and A. D. Friederici, "Electric brain responses reveal gender differences in music processing," *Neuroreport*, vol. 14, no. 5, pp. 709–713, 2003.
[97] S. A. Miles, R. A. Miranda, and M. T. Ullman, "Sex differences in music: A female advantage at recognizing familiar melodies," *Frontiers in Psychology*, vol. 278, no. 5, pp. 1–12, 2016.

Ryosuke Sawata (S'15–M'16) received his B.S. and M.S. degrees in Electronics and Information Engineering from Hokkaido University, Japan in 2014 and 2016, respectively. He is currently a researcher in the Sony Corporation. His research interests include biosignal processing, music information retrieval and acoustic signal processing. He is a member of the IEEE.



Takahiro Ogawa (S'03–M'08) received his B.S., M.S. and Ph.D. degrees in Electronics and Information Engineering from Hokkaido University, Japan in 2003, 2005 and 2007, respectively. He joined Graduate School of Information Science and Technology, Hokkaido University in 2008. He is currently an associate professor in the Graduate School of Information Science and Technology, Hokkaido University. His research interests are multimedia signal processing and its applications. He has been an Associate Editor of ITE Transactions on Media Technology and Applications. He is a member of the IEEE, ACM, EURASIP, IEICE, and ITE.

Miki Haseyama (S'88–M'91–SM'06) received her B.S., M.S. and Ph.D. degrees in Electronics from Hokkaido University, Japan in 1986, 1988 and 1993, respectively. She joined the Graduate School of Information Science and Technology, Hokkaido University as an associate professor in 1994. She was a visiting associate professor of Washington University, USA from 1995 to 1996. She is currently a professor in the Graduate School of Information Science and Technology, Hokkaido University. Her research interests include image and video processing and its development into semantic analysis. She has been a Vice-President of the Institute of Image Information and Television Engineers, Japan (ITE), an Editor-in-Chief of ITE Transactions on Media Technology and Applications, a Director, International Coordination and Publicity of The Institute of Electronics, Information and Communication Engineers (IEICE). She is a member of the IEEE, IEICE, Institute of Image Information and Television Engineers (ITE) and Acoustical Society of Japan (ASJ).