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

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

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

Recently, it has become obvious that electroencephalogram (EEG) features extracted from EEG signals of a human who listens to music have a close relationship to his/her music preference [12–18]. For example, Hadjidimitriou et al. reported discrimination between user’s EEG signals depending on music preference, i.e., whether the user listened to favorite or unfavorited musical pieces [17, 18]. Furthermore, observation of EEG signals has become easier, and the quality of observed signals has become better due to the development and miniaturization of an EEG sensor. For instance, the music inspiration from your subconsciousness (mico), which is a headphone with an embedded EEG sensor, produced by neurowear,2 can easily provide EEG signals from the user when the headphone is used like standard ones. Therefore, we consider that EEG features obtained from the user’s EEG signals during listening to music is utilizable and important to find a set of audio features related to user’s music preference. However, so far, such research has not been studied adequately as far as we know.

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

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

This work was partly supported by Grant-in-Aid for Scientific Research (B) 25280036 and Grant-in-Aid for Challenging Exploratory Research 15K12023, Japan Society for the Promotion of Science (JSPS).
Table 1. EEG features calculated by the proposed method: C denotes the number of channels of EEG signals and CP represents the number of symmetric electrode pairs placed on the scalp.

<table>
<thead>
<tr>
<th>DESCRIPTION</th>
<th>DIMENSION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zero Crossing Rate</td>
<td>C</td>
</tr>
<tr>
<td>Content Percentage of The Power Spectrum</td>
<td></td>
</tr>
<tr>
<td>Slow wave (1-7 Hz)</td>
<td>C</td>
</tr>
<tr>
<td>Fast wave (7-15 Hz)</td>
<td>C</td>
</tr>
<tr>
<td>Vertex</td>
<td></td>
</tr>
<tr>
<td>Slow wave (13-19 Hz)</td>
<td>C</td>
</tr>
<tr>
<td>Fast wave (19-30 Hz)</td>
<td>C</td>
</tr>
<tr>
<td>Power Spectrum of The Hemispheric Asymmetry [15]</td>
<td></td>
</tr>
<tr>
<td>Middle frequency (7-10 Hz)</td>
<td>2C</td>
</tr>
<tr>
<td>Fast frequency (13-19 Hz)</td>
<td>2C</td>
</tr>
<tr>
<td>Spectral Energy of θ wave (1-7 Hz)</td>
<td>2C</td>
</tr>
<tr>
<td>Spectral Energy of μ wave (7-10 Hz)</td>
<td>2C</td>
</tr>
<tr>
<td>Spectral Energy of γ wave (10-30 Hz)</td>
<td>2C</td>
</tr>
<tr>
<td>TOTAL</td>
<td>(2C + 16Cp)</td>
</tr>
</tbody>
</table>

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

In this section, we explain favorite music classification using EEG-based optimal audio features. In our method, we first extract user’s EEG features from EEG signals recorded while the user listens to musical pieces and audio features from corresponding musical pieces. Then we apply our novel CCA, i.e., KDLPCCA to obtained EEG and audio features in order to extract the relationship between them. Next, we select a set of audio features related to user’s music preference using KDLPCCA-based classifier since it is broadly used in some studies. In this paper, we experimentally employ SVM as a classifier since it is broadly used in some studies.

2.1. EEG-based Audio Feature Selection via KDLPCCA

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

First, we obtain two sets of \(N\) feature vectors \(X^E = [x_1^E, x_2^E, \ldots, x_n^E] \in \mathbb{R}^{d_E \times N}\) and \(X^A = [x_1^A, x_2^A, \ldots, x_n^A] \in \mathbb{R}^{d_A \times N}\) from EEG and audio signals, where \(d_E\) and \(d_A\) are the dimensions of the EEG and the audio feature vector, respectively. Note that the sets of EEG and audio features, used in this paper, were acquired in an authors’ previous work [24] and summarized in Tables 1 and 2. Since the space of the paper is limited, the details of calculation of EEG and audio feature are given there. Each obtained feature vector has a label representing whether it is obtained from "Favorite" or "Unfavorite" musical pieces.

Table 2. Audio features used in the proposed method

<table>
<thead>
<tr>
<th>CATEGORY</th>
<th>DESCRIPTION</th>
<th>STATISTICS</th>
<th>DIMENSION</th>
</tr>
</thead>
<tbody>
<tr>
<td>dynamics</td>
<td>Root Mean Square</td>
<td>Mean, Std</td>
<td>2</td>
</tr>
<tr>
<td>spectral</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>timbre</td>
<td>Zero Crossing Rate</td>
<td>Mean, Std</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rhythm</td>
<td>Tempo</td>
<td>Mean, Std</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>total</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2.1.1. Applying KDLPCCA

We firstly apply non-linear mapping to obtained feature vectors. Note that the dimension of audio feature vector needs to be preserved in order to apply the following feature selection. Hence, feature vectors of only EEG set, i.e., \(x_k^E (j = 1, 2, \ldots, N)\), are transformed into Hilbert space via non-linear maps \(\phi_E : x_k^E \rightarrow \phi_E(x_k^E) \in \mathbb{R}^{d_E} \). From the aforementioned mapped results, we obtain \(\Phi^E = \{\phi_E(x_k^E), \phi_E(x_k^A), \ldots, \phi_E(x_N^A)\} \in \mathbb{R}^{d_E \times N}\). Then we apply Discriminant Locality Preserving CCA (DLPCA) [25] to \(X^E\) and \(\Phi^E\). Specifically, we seek to maximize the following objective function by using the projections \(u_k^E \in \mathbb{R}^{d_E}\) and \(u_k^A \in \mathbb{R}^{d_A}\):

\[
(\hat{u}_k^E, \hat{u}_k^A) = \arg \max_{u_k^E, u_k^A} \frac{u_k^E^T C_{EE} u_k^E}{\sqrt{u_k^E^T C_{EE} u_k^E} \sqrt{u_k^A^T C_{AA} u_k^A}}
\]

where \(\zeta\) is a tunable parameter and

\[
C_E = \Phi^E H (S_E \circ S_A^H) H^T \Phi^E^T, \quad C_A = \Phi^A H (S_A \circ S_E^H) H^T \Phi^A^T.
\]

\[
C_E^\zeta = \Phi^E H (L_E^\zeta + L_A^\zeta) H^T \Phi^E, \quad C_A^\zeta = \Phi^A H (L_A^\zeta + L_E^\zeta) H^T \Phi^A
\]

where the symbol \(\circ\) denotes the Hadamard product. Furthermore, \(H = I - \frac{1}{N} I^N\) is a centering matrix, \(I \in \mathbb{R}^{N \times N}\) is an \(N \times N\) identity matrix, and \(I = [1, 1, \ldots, 1] \in \mathbb{R}^N\) is an \(N\)-dimensional vector. In Eq. (1), \(C_E \in \mathbb{R}^{d_E \times N}\) and \(C_A \in \mathbb{R}^{d_A \times N}\), defined as shown in Eq. (2), are respectively the within-class covariance matrix and the between-class covariance matrix, which they are derived by considering the class similarity matrices \(S_E = [S_{ij}^E]_{N \times N}\), \(S_A = [S_{ij}^A]_{N \times N}\), \(S_E = [S_{ij}^E]_{N \times N}\), and \(S_A = [S_{ij}^A]_{N \times N}\), defined as follows:

\[
S_{ij}^E = \begin{cases}
\exp(-\delta_{ij}/\tau_{ij}), & \text{label}(\phi_E(x_i^E)) = \text{label}(\phi_E(x_j^E)) \\
0, & \text{otherwise}
\end{cases}
\]

(4)

\[
S_{ij}^A = \begin{cases}
\exp(-\delta_{ij}/\tau_{ij}), & \text{label}(\phi_A(x_i^A)) \neq \text{label}(\phi_A(x_j^A)) \\
0, & \text{otherwise}
\end{cases}
\]

(5)

\[
S_{ij}^A = \begin{cases}
\exp(-\delta_{ij}/\tau_{ij}), & \text{label}(x_i^A) = \text{label}(x_j^A) \\
0, & \text{otherwise}
\end{cases}
\]

(6)

\[
S_{ij}^A = \begin{cases}
\exp(-\delta_{ij}/\tau_{ij}), & \text{label}(x_i^A) \neq \text{label}(x_j^A) \\
0, & \text{otherwise}
\end{cases}
\]

(7)

where \(\delta_{ij}^E = ||\phi_E(x_i^E) - \phi_E(x_j^E)||^2, \quad \tau_{ij} = \frac{1}{N(N-1)} \sum_{k \neq i} \sum_{k \neq j} \delta_{ij}^E, \quad \delta_{ij}^A = ||x_i^A - x_j^A||^2, \quad \tau_A = \frac{1}{N(N-1)} \sum_{k \neq i} \sum_{k \neq j} \delta_{ij}^A\). Furthermore, "label(\bullet)" is...
the \( \bullet \)'s class label. By using these class similarity matrices, we can also derive Laplacian matrices \( L^x \in \mathbb{R}^{N \times N} \), \( L^s \in \mathbb{R}^{N \times N} \), \( L^s \in \mathbb{R}^{N \times N} \) and \( L^a \in \mathbb{R}^{N \times N} \) as defined in Eq. (3) as follows:

\[
\begin{align*}
L^x &= D^x - S^x \circ S^x, \\
L^s &= D^s - S^s \circ S^s, \\
L^a &= D^a - S^a \circ S^a, \\
L^a &= D^a - S^a \circ S^a,
\end{align*}
\]

(8)

where \( D^x = \text{diag}[\sum(S_{i,i}^x)^2, \sum(S_{i,i}^x)^2, \ldots, \sum(S_{i,i}^x)^2] \), \( D^s = \text{diag}[\sum(S_{i,i}^s)^2, \sum(S_{i,i}^s)^2, \ldots, \sum(S_{i,i}^s)^2] \), \( D^a = \text{diag}[\sum(S_{i,i}^a)^2, \sum(S_{i,i}^a)^2, \ldots, \sum(S_{i,i}^a)^2] \).

Solving Eq. (1) corresponds to the maximization of the correlation considering the class information and the local structures of input training sets of EEG and audio features.

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

\[
\frac{u^x \text{T}(C_u - \zeta C_s)u^a}{\sqrt{u^x \text{T}C_u u^x} \sqrt{\alpha^T C_s \alpha}} = \frac{u^x \text{T}(\Phi^TH(C_u - \zeta C_s)\Phi)u^a}{\sqrt{\alpha^T C_s \alpha}} \sqrt{\text{T}C_u \text{C}u^a} = \alpha^T K^a(X^a \circ S^a)X^T u^a,
\]

where

\[
\begin{align*}
\tilde{C}^a &= HK^a H(L^x + \tilde{L}^x)HK^a H + \xi HK^a H,
\end{align*}
\]

and \( \alpha \in \mathbb{R}^N \) is a coefficient vector. Furthermore, \( K^a = \Phi^TH \Phi \in \mathbb{R}^{N \times N} \) is gram matrix whose \( (i, j) \)th component is \( k^a(x^a_i, x^a_j) \) and \( \xi \) is a regularization parameter. Then \( \delta^a_{ij} \) in Eqs. (4) and (5) becomes calculable as \( \delta^a_{ij} = ||\delta^a_i(x^a_i) - \phi(x^a_i)||^2 = (K^a)_i - 2(K^a)_i + (K^a)_i \), where \( (K^a)_i \) denotes the \( (i, \ j) \)th component of \( K^a \). Consequently, we can obtain \( \alpha \) and \( u^a \) as the optimal solutions by solving the following Lagrange function:

\[
\begin{align*}
\mathcal{L}(\alpha, u^a) &= \alpha^T K^a(X^a \circ S^a)X^T u^a - \frac{\lambda^e_1}{2}(\alpha^T C_s \alpha - 1) - \frac{\lambda^e_2}{2}(u^a^T C_s u^a - 1),
\end{align*}
\]

(12)

where \( \lambda^e_1 = \lambda^e_2 \), and they are equivalent to the optimal solution of Eq. (1).

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

### 2.1.2. Feature Selection Using KDLPCA-based cross-loadings

In order to select audio features considering the relationship between EEG and audio features, class information and local structures of those features, we newly derive KDLPCA-based cross-loadings \( r \in \mathbb{R}^{d_a} \) based on Eq. (12) defined as follows:

\[
r = X^a \circ (S^r \circ S^a) - \zeta (S^r \circ S^a)^T K^a \alpha.
\]

(13)

Cross-loadings measure how well a feature loads on the canonical variate from the opposing feature set instead of its own. Specifically, an \( ith \) component \( r_i \) of \( r \) denotes the degree of relevance between KDLPCA-based canonical variates of EEG and corresponding \( ith \) original audio feature. Thus, we employ these cross-loadings as novel criteria to select audio features related to the user’s music preference. Specifically, we select audio features according to the value of the cross-loadings in descending order.

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

### 2.2. Favorite Music Classification

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

\[
f(x^{\text{test}}) = \text{sgn}(w, x^{\text{test}} + b).
\]

(14)

Given training music database \( D = \{(x^{\text{train}}_i, y_i)\}_{i=1}^{N_{\text{train}}} \), \( w \) and \( b \) can be obtained by solving the SVM formulation [23] as follows:

\[
\min_{\|w\|_2^2 + C} \sum_{i=1}^{N_{\text{train}}} \eta_i,
\]

s.t. \( y_i(w, x^{\text{train}}_i) + b \geq 1, \eta_i \geq 0, \forall (x^{\text{train}}_i, y_i) \in D \),

where \( x^{\text{train}}_i \) is obtained from the \( ith \) piece of music database, \( \eta_i \) is a slack variable, and \( y_i \in \{+1, -1\} \) denotes a correct class label, i.e., “Favorite” or “Unfavorite”.

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

### 3. EXPERIMENTAL RESULTS

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

---

\(^3\)The database has 12 genres: Pops, Rock, Jazz, Latin, Classic, March, etc. We use 5 musical pieces for each genre, i.e., the total number of musical pieces used in our experiment is 60.
for KDLPCCA, where the kernel width is chosen according to each subject adaptively since the optimal $k$ is different for each subject as shown in Figs. 2(a)-(e). Meanwhile, our method, i.e., KDLPCCA-based method does not depend on $k$. Nonetheless, the performance of our method is relatively high as shown in Figs. 2(a)-(e). In addition, our method’s average accuracy for all subjects outperforms all other methods as shown in Fig. 2(f). Therefore, our method realizes robust favorite music classification successfully.

4. CONCLUSIONS

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

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

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

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

Next, we describe how to collect EEG signals in our experiment. EEG signals were collected from five healthy subjects, and their average age was about 23 years. We recorded EEG signals from 12 channels (Fp1, Fp2, F7, F8, C3, C4, P3, P4, O1, O2, T3 and T4) decided based on [17, 31] according to the international 10-20 system. The subjects were instructed to keep their eyes closed and to relax and remain seated during listening to music.

Since EEG signals are weak, we amplified the signals by utilizing band-pass filter to recorded EEG signals in order to avoid artifacts, and we set the filter bandwidth to 0.04–100 Hz. The subjects were instructed to keep their eyes closed and to relax and remain seated during listening to music.

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

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


