EEG-Based Emotion Recognition in Music Listening

Yuan-Pin Lin, Chi-Hong Wang, Tzyy-Ping Jung*, *Senior Member, IEEE*, Tien-Lin Wu, Shyh-Kang Jeng, Jeng-Ren Duann, *Member, IEEE*, and Jyh-Horng Chen, *Member, IEEE*

***Abstract*—Ongoing brain activity can be recorded as electroencephalograph (EEG) to discover the links between emotional states and brain activity. This study applied machine-learning algorithms to categorize EEG dynamics according to subject self-reported emotional states during music listening. A framework was proposed to optimize EEG-based emotion recognition by systematically 1) seeking emotion-specific EEG features and 2) exploring the efficacy of the classifiers. Support vector machine was employed to classify four emotional states (joy, anger, sadness, and pleasure) and obtained an averaged classification accuracy of 82.29% $\pm$ 3.06% across 26 subjects. Further, this study identified 30 subject-independent features that were most relevant to emotional processing across subjects and explored the feasibility of using fewer electrodes to characterize the EEG dynamics during music listening. The identified features were primarily derived from electrodes placed near the frontal and the parietal lobes, consistent with many of the findings in the literature. This study might lead to a practical system for noninvasive assessment of the emotional states in practical or clinical applications.**

***Index Terms*—EEG, emotion, machine learning, music.**

I. Introduction

THE ULTIMATE objective of bioinspired multimedia research is to access multimedia content from users' biosignals through interpreting the inherent responses during multimedia appreciation. For example, a recent work in brainwave–music interface [1] built-up sonification rules to map EEG characteristics to musical structures (note, intensity, and pitch). Unlike brainwave–music interface, the ultimate goal of this study is to build a more immersive multimedia environment based on listener' appreciation/emotion, as measured by EEG. In both applications, the ability to interpret user's multimedia-induced perception and emotional experience is very crucial.

Manuscript received December 17, 2009; revised February 23, 2010; accepted April 4, 2010. Date of publication May 3, 2010; date of current version June 16, 2010. This work was supported in part by the Taiwan National Science Council under Grant NSC97-2917-I-002-119. *Asterisk indicates corresponding authors.*

Y.-P. Lin is with the Department of Electrical Engineering, National Taiwan University, Taipei 10617, Taiwan (e-mail: yplin0115@gmail.com).

C.-H. Wang is with the Department of Neurology, Cardinal Tien Hospital, Yung-Ho Branch, Taipei 23445, Taiwan (e-mail: isami.chwang@gmail.com).

*T.-P. Jung is with the Swartz Center for Computational Neuroscience, University of California, San Diego, CA 92093-0961 USA (e-mail: jung@sccn.ucsd.edu).

T.-L. Wu and S.-K. Jeng are with the Department of Electrical Engineering, National Taiwan University, Taipei 10617, Taiwan (e-mail: terrywu0311@hotmail.com; skjeng@ew.ee.ntu.edu.tw).

J.-R. Duann is with the Institute for Neural Computation, University of California, San Diego, CA 92093-0523 USA, and also with the Biomedical Engineering Research and Development Center, China Medical University Hospital, Taichung 40447, Taiwan (e-mail: jengren00@gmail.com).

J.-H. Chen is with the Department of Electrical Engineering, National Taiwan University, Taipei 10617, Taiwan (e-mail: jhchen@ntu.edu.tw).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TBME.2010.2048568

Many methods for estimating human emotion have been proposed in the past. The conventional methods basically utilize audio and visual attributes to model human emotional responses, such as speech, facial expressions, and body gestures. More recently, accessing physiological responses has gained increasing attention in characterizing the emotional states [2]–[5]. Biosignals used in these studies were recorded from autonomic nervous system (ANS) in the periphery, such as ECG, skin conductance (SC), electromyography (EMG), respiration, pulse, etc. As compared with audio- and/or visual-based methods, the responses of biosignals tend to provide more detailed and complex information as an indicator for estimating emotional states [3].

In addition to periphery biosignals, signals captured from the brain in central nervous system (CNS) have been proved to provide informative characteristics in responses to the emotional states. The ongoing brain activity recorded using EEG provides noninvasive measurement with temporal resolution in milliseconds. EEG has been used in cognitive neuroscience to investigate the regulation and processing of emotion for the past decades. Power spectra of the EEG were often assessed in several distinct frequency bands, such as delta (δ: 1–3 Hz), theta (θ: 4–7 Hz), alpha (α: 8–13 Hz), beta (β: 14–30 Hz), and gamma (γ: 31–50 Hz) [6], to examine their relationship with the emotional states. One of the common indicators of emotional states is the alpha-power asymmetry derived from the spectral differences between a symmetric electrode pair at the anterior areas of the brain [7]–[9]. Other spectral changes and brain regions were also reported, which are associated to emotional responses, such as the alpha-power changes at right parietal lobe [9], [10], the theta-power changes at right parietal lobe [11], the frontal midline (Fm) theta power [12], the beta-power asymmetry at the parietal region [13], and the gamma spectral changes at the right parietal regions [14]. Although emotion is one of complex and less-understood cognitive functions generated in the brain and associated with several brain oscillations in combinations [15], the aforementioned evidences proved the feasibility of using EEG to characterize emotional states. Therefore, compared to periphery biosignals, the EEG might provide more insights into emotional processes and responses.

Using EEG to assess emotional states is still in its infancy, compared to the works using audio–visual based methods. Ishino and Hagiwara [16] proposed a system that estimated subjective feeling using neural networks to categorize emotional states based on EEG features. They reported an average accuracy range from 54.5% to 67.7% for each of four emotional states. Takahashi [17] proposed an emotion-recognition system using multimodal signals (EEG, pulse, and SC). The experimental results showed that the recognition rate using a support vector machine (SVM) reached an accuracy of 41.7% for five emotions. Chanel *et al.* [18] showed that arousal assessment of emotion could be obtained with a maximum accuracy of 58% for three emotion classes. The classification was performed using the Naïve Bayes classifier applied to six EEG features derived from specific frequency bands at particular electrode locations. Heraz *et al.* [19] established an agent to predict emotional states during learning. The best classification in the study was an accuracy of 82.27% for distinguishing eight emotional states, using k-nearest neighbors as a classifier and the amplitudes of four EEG components as features. Chanel *et al.* [20] reported an average accuracy of 63% by using EEG time-frequency information as features and SVM as a classifier to characterize EEG signals into three emotional states. Ko *et al.* [21] demonstrated the feasibility of using EEG relative power changes and Bayesian network to predict the possibility of user's emotional states. Also, Zhang and Lee [22] proposed an emotion understanding system that classified users' status into two emotional states with the accuracy of $73.0\% \pm 0.33\%$ during image viewing. The system employed asymmetrical characteristics at the frontal lobe as features and SVM as a classifier. However, most of works focused only on EEG spectral power changes in few specific frequency bands or at specific scalp locations. No study has yet been conducted to systematically explore the correspondence between emotional states and EEG spectral changes across the whole brain and relate the findings to those previously reported in emotion literature.

The objective of this study is to systematically uncover the association between the EEG dynamics and emotions by 1) searching emotion-specific features of the EEG and 2) testing the efficacy of different classifiers. To this end, this study will explore a wide range of features across multiple subjects and establish an EEG-based emotion-recognition scheme in the next sections.

II. Data Collection and Experiment Procedure

EEG data in this study were collected from 26 healthy subjects (16 males and 10 females; age 24.40 ± 2.53) during music listening. Most of subjects were undergraduate or graduate students from College of Electrical Engineering and Computer Science or College of Engineering at National Taiwan University. They had minimal formal musical education and could thus be considered as nonmusicians. A 32-channel EEG module (Neuroscan, Inc.) arranged according to international 10–20 system was used. All leads were referenced to linked mastoids (average of A1 and A2), and a ground electrode was located in the forehead. The sampling rate and filter bandwidth were set to 500 Hz and $1 \sim 100$ Hz, respectively. An additional 60 Hz notch filter was employed to avoid the power-line contamination. All electrode impedances were kept below 10 kΩ for the EEG. Electrooculogram (EOG) activity was also recorded to facilitate subsequent EEG artifact rejection.

Subjects were instructed to keep their eyes closed and remain seated in the music-listening experiment. This study examined four basic emotional states following a 2-D valence–arousal emotion model [23], including joy (positive valence and high arousal), anger (negative valence and high arousal), sadness (negative valence and low arousal), and pleasure (positive valence and low arousal). Sixteen excerpts from Oscar's film soundtracks were selected as stimuli, according to the consensus tagging reported from hundreds of subjects [24]. Each was edited into a 30-s music excerpt. Four of 16 music excerpts were randomly selected without replacement to form a four-run experiment. A 15-s silent rest was inserted between music excerpts. After each run, the subjects were requested to report the emotional states (joy, anger, sadness, and pleasure) to each music excerpt based on what they felt via a tool FEELTRACE [25] for labeling on a 2-D emotion model. Each experiment thus consisted of 16 30-s emotion-specific EEG segments for further analysis, whereas the given self-reported emotional states were used to verify EEG-based emotion classification.

III. Data Classification

A. Feature Extraction

The recorded EEG data were first preprocessed to remove serious and obvious motion artifacts through visual inspection, and the artifact-free data were then divided into 16 30-s segments for each individual. Since the features under study were based on the spectral power changes, a 512-point short-time Fourier transform (STFT) with a nonoverlapped Hanning window of 1 s was applied to each of 30 channels of the EEG data to compute the spectrogram. The resultant spectral time series was averaged into five frequency bands, including delta (δ: 1–3 Hz), theta (θ: 4–7 Hz), alpha (α: 8–13 Hz), beta (β: 14–30 Hz), and gamma (γ: 31–50 Hz). The spectral time series for each subject thus consisted of around 480 sample points. In order to find an optimal set of emotion-specific features from a wide range of feature candidates, two major factors were tested: 1) the types of features and 2) the frequency bands of the EEG. Several feature categories were systematically tested in this study. First, individual spectral power from 30 scalp electrodes were used as the features, including Fp1, Fp2, F7, F3, Fz, F4, F8, FT7, FC3, FCz, FC4, FT8, T7, C3, Cz, C4, T8, TP7, CP3, CPz, CP4, TP8, P7, P3, Pz, P4, P8, O1, Oz, and O2. This feature type was named PSD30 (power spectrum density of all 30 channels) in the following sections. Next, the spectral power of the hemispheric asymmetry index was also adopted and extended from our previous study [26], [27]. Throughout the whole brain, there were 12 asymmetry indexes derived from 12 symmetric electrode pairs, namely Fp1–Fp2, F7–F8, F3–F4, FT7–FT8, FC3–FC4, T7–T8, P7–P8, C3–C4, TP7–TP8, CP3–CP4, P3–P4, and O1–O2. The asymmetry indexes were calculated either by power subtraction (e.g., power of C3 $-$ power of C4) or division (e.g., power of C3/power of C4) and labeled as differential asymmetry of 12 electrode pairs (DASM12) and rational asymmetry of 12 electrode pairs (RASM12), respectively. Lastly, the individual spectra of these 12 symmetric electrode pairs (24 channels) were also used as the features for emotion classification, named power spectrum density of 24 channels (PSD24). The PSD24 was part of PSD30 without the electrodes along the midline (Fz, FCz, Cz, CPz, Pz, and Oz). Table I summarizes the four feature types used in this study. Before feeding data to classifiers, the feature vectors were normalized to the range from 0 to 1. In addition, to test the feasibility of automatic classification of

TABLE I
NUMBER OF FEATURES USING DASM12, RASM12, PSD24, AND PSD30 IN DIFFERENT EEG FREQUENCY BANDS

Feature Type	EEG Frequency Band					
	Delta	Theta	Alpha	Beta	Gamma	ALL
DASM12	12	12	12	12	12	60
RASM12	12	12	12	12	12	60
PSD24	24	24	24	24	24	120
PSD30	30	30	30	30	30	150

The condition ALL represents the combination of five EEG frequency bands.

EEG segments, each EEG segment was tagged with the corresponding emotional label according to the subject's self-report.

B. Feature Classification

This study employed and evaluated two classifiers, multilayer perceptron (MLP) and SVM, for EEG classification. These classifiers have been separately applied previously to some of the aforementioned features (DASM12 and PSD24) [26], [27]. This study systematically compared the effects of all the four feature types on the classification performance.

The MLP used in this study consisted of an input layer, a hidden layer with a sigmoid function representing neural excitation, and an output layer. The number of neurons in the input layer and hidden layers varied according to the feature type used, whereas the number of neurons in the output layer was four, each corresponded to one of the four emotional states. The number of neurons in the hidden layer was empirically assigned based on the half of summation of neurons in the input and output layers. For example, when the feature-type DASM12 was used as the input to the MLP, the number of neurons of the input layer and hidden layer were 12 and 8, respectively. The EEG feature vector and the corresponding emotional label were used to adjust the weight coefficients within the network layers using a back-propagation algorithm. After the training procedure converged, the optimized MLP estimated an emotion label for each EEG segment. This study employed Weka [28], a collection of machine-learning algorithms intended for data mining, to perform the MLP classification.

Next, this study employed SVM to classify the emotion label for each EEG segment. SVM is one of the most popular supervised learning algorithms for solving the classification problems. The basic idea is to project input data onto a higher dimensional feature space via a kernel transfer function, which is easier to be separated than that in the original feature space. Depending on input data, the iterative learning process of SVM would eventually converge into optimal hyperplanes with maximal margins between each class. These hyperplanes would be the decision boundaries for distinguishing different data clusters. This study used LIBSVM software [29] to build the SVM classifier and employed radial basis function (RBF) kernel to nonlinearly map data onto a higher dimension space.

In the experiments, the number of sample points from each subject was around 480 points (16 30-s EEG segments $\times$ around 30 points per segment) derived from artifact-free EEG signals. A ten times of 10-fold cross-validation scheme with randomization was applied to dataset from each subject in order to increase the reliability of the recognition results. In a 10-fold cross validation, whole EEG dataset was divided into ten subsets. The MLP and SVM were trained with nine subsets of feature vectors, whereas the remaining subset was used for testing. This procedure was repeated ten times with each subset having an equal chance of being the testing data. The entire 10-fold cross validation was then repeated ten times with different subset splits. The accuracy was evaluated by the ratio of correctly classified number of samples and the total number of samples. After validation processing, the average subject-dependent performance using different feature types as inputs were evaluated.

C. Feature Selection

Feature selection was a necessary process before performing any data classification and clustering. The objective of feature selection is to extract a subset of features by removing redundant features and maintaining the informative features. Further, since 32-channel EEG module was used to acquire the brain activity, the feature selection also tested the feasibility of using fewer electrodes for practical applications. Thus, the feature selection seems particularly important not only to improve the computational efficiency, but also expand the applicability of EEG-based human-centered system in real-world applications. This study adopted F-score index, one of statistical methods to measure the ratio of between- and within-class variance, for sorting each feature in descending order accounting for discrimination between different EEG patterns, i.e., the larger the F-score, the greater the discrimination power. The F-score of the ith feature is defined as follows [30]:

$$F(i) = \frac{\sum_{l=1}^{g} n_l \left(\overline{x}_{l,i} - \overline{x}_i\right)\left(\overline{x}_{l,i} - \overline{x}_i\right)'}{\sum_{l=1}^{g}\sum_{k=1}^{n_l}\left(x_{l,k,i} - \overline{x}_i\right)\left(x_{l,k,i} - \overline{x}_i\right)'}$$

where $\overline{x}_i$ and $\overline{x}_{l,i}$ are the average of the ith feature of the entire data set and class l data set ($l = 1 \sim g$, $g = 4$ for four emotion labels), respectively, $x_{l,k,i}$ is the ith feature of the kth of the class l instance, and n_l is the number of instances of class l. The number of the features varied depending on the types of feature used. Further, in order to investigate the number of features retained, the leave-N-feature-out scheme was also employed for iteratively conducting the classification by removing a subset of features based on the F-score rank list.

IV. CLASSIFICATION RESULT

To better understand the association between EEG activities and the emotional responses, several factors have been intensively investigated, which are as follows:

1) the types of features;
2) the frequency bands of the EEG;
3) the types of classifiers;
4) the number of features;
5) the number of electrodes.

Table II shows the averaged classification performance of MLP using four feature types DASM12, RASM12, PSD24, and

TABLE II
AVERAGE MLP RESULTS (STANDARD DEVIATION) USING DASM12, RASM12, PSD24, AND PSD30

Feature Type	EEG Frequency Band					
	Delta	Theta	Alpha	Beta	Gamma	ALL
DASM12	**63.93 (5.91)** *	**63.67 (5.34)** *	**64.07 (6.42)**	55.71 (7.53)	53.24 (7.02)	**81.52 (3.71)**
RASM12	48.54 (5.70)	50.69 (5.39)	55.40 (6.95)	48.21 (7.03)	44.82 (6.38)	65.33 (5.52)
PSD24	49.20 (6.07)	52.10 (6.46)	57.79 (7.70)	53.20 (7.25)	54.46 (6.55)	75.66 (4.41)
PSD30	52.12 (5.72)	55.61 (6.49)	61.89 (7.84)	**57.92 (7.09)**	**58.04 (6.50)**	79.54 (4.21)

The condition ALL refers to a combination of five EEG frequency bands, and numbers in bold represents the best performance using the frequency band. The significant difference was tested under each frequency band (* $p < 0.05$).

TABLE III
AVERAGE SVM RESULTS (STANDARD DEVIATION) USING DASM12, RASM12, PSD24, AND PSD30

Feature Type	EEG Frequency Band					
	Delta	Theta	Alpha	Beta	Gamma	ALL
DASM12	**69.91 (6.55)** *	**68.27 (5.29)** *	**66.94 (6.41)** *	**58.83 (8.02)**	57.35 (7.37)	**82.29 (3.06)** *
RASM12	50.91 (5.62)	51.39 (5.94)	56.95 (6.79)	50.29 (6.99)	47.61 (7.23)	65.81 (5.09)
PSD24	51.02 (6.17)	53.27 (6.95)	54.61 (7.44)	55.42 (6.57)	56.80 (6.58)	69.54 (5.10)
PSD30	53.38 (5.79)	55.61 (6.68)	56.64 (7.03)	58.71 (6.31)	**59.54 (6.16)**	71.15 (4.88)

The condition ALL refers to a combination of five EEG frequency bands, and numbers in bold represents the best performance using the frequency band. The significant difference was tested under each frequency band (* $p < 0.05$).

PSD30 across different EEG frequency bands. The classification performance of using DASM12 was evidently better than those based on other feature types under conditions (significant difference shown in the delta and theta bands, $p < 0.05$), except in the cases using beta and gamma power. A maximum classification accuracy of 81.52% $\pm$ 3.71% was obtained using ALL frequency bands (but no significant difference compared to PSD30, $p > 0.05$).

Table III shows the averaged classification performance of SVM using DASM12, RASM12, PSD24, and PSD30 across different EEG frequency bands. Again, DASM12 gave best classification performance (significant difference shown in delta, theta, and alpha, $p < 0.05$), except in the case using gamma power. A maximum classification accuracy of 82.29% $\pm$ 3.06% was obtained from the condition ALL with significant difference ($p < 0.05$). When comparing the classification results obtained by MLP and SVM, applied to DASM12, it was noted that SVM outperformed MLP by 2%$\sim$4% (significant improvement was shown in the delta, theta, and gamma, $p < 0.05$), whereas in condition ALL, SVM improved the classification performance from 81.52% $\pm$ 3.71% to 82.29% $\pm$ 3.06% (but not statistically significant, $p > 0.05$).

Next, since the best performance was obtained using DASM12 across all frequencies (condition ALL), F-score index was further applied to this feature type to sort the feature across frequency bands. The leave-N-feature-out scheme was then used to iteratively remove a group of ranked features and examine its effects on the classification performance. Fig. 1 shows the average results across subjects obtained by iteratively removing N F-score-ranked features out at a time, e.g., in the case of leaving-5-features-out, each point in the figure corresponds to retain 55 of 60 features in the classification process with the first point representing the removal of top-F-score-ranked features ranked from first to fifth and the second data point representing the removal of features ranked from sixth to tenth, etc. Fig. 1(b) plots the average results across subjects obtained by iteratively removing N F-score-ranked features out at a time, but the removed N features were randomly selected. First, the classification performance obtained from SVM decreased as the number of features increased from 5 to 20. Second, the classification accuracy decreased appreciably as top-F-score-ranked features were removed as the lower accuracy was seen on the left of Fig. 1(a). On the contrary, no appreciable differences in accuracy were shown in Fig. 1(b) when N of 60 features were randomly drawn and removed from the inputs. These results suggested that the top-F-score-ranked features were more discriminative than the lower ranked ones. Fig. 2 shows the feature space along (a) top and (b) last two F-score-ranked features. As can be seen, Fig. 2(a) is more structural and the data points corresponding to anger (triangles) were largely separable from the rest of the points, whereas the data points of different emotional states were highly overlapped in Fig. 2(b).

A natural question is if the optimal EEG features for emotion recognition were common across subjects (i.e., subject-independent). In this study, the subject-independent feature set was evaluated by summarizing the accumulation of F-score value of each feature across subjects. Fig. 3 shows those classification results obtained by top-N F-score-ranked subject-dependent features, subject-independent features, and the number of electrodes required for deriving the top-N subject-independent features. As expected, the classification accuracy in general declined as the number of input attributes of DASM12 decreased. Interestingly, the classification performance using subject-independent features was comparable to that using subject-dependent features, except for the case where only five of 60 features were used. As an example, the averaged classification accuracy of 74.10% $\pm$ 5.85% was obtained by applying SVM to the top-30 subject-independent features, compared to

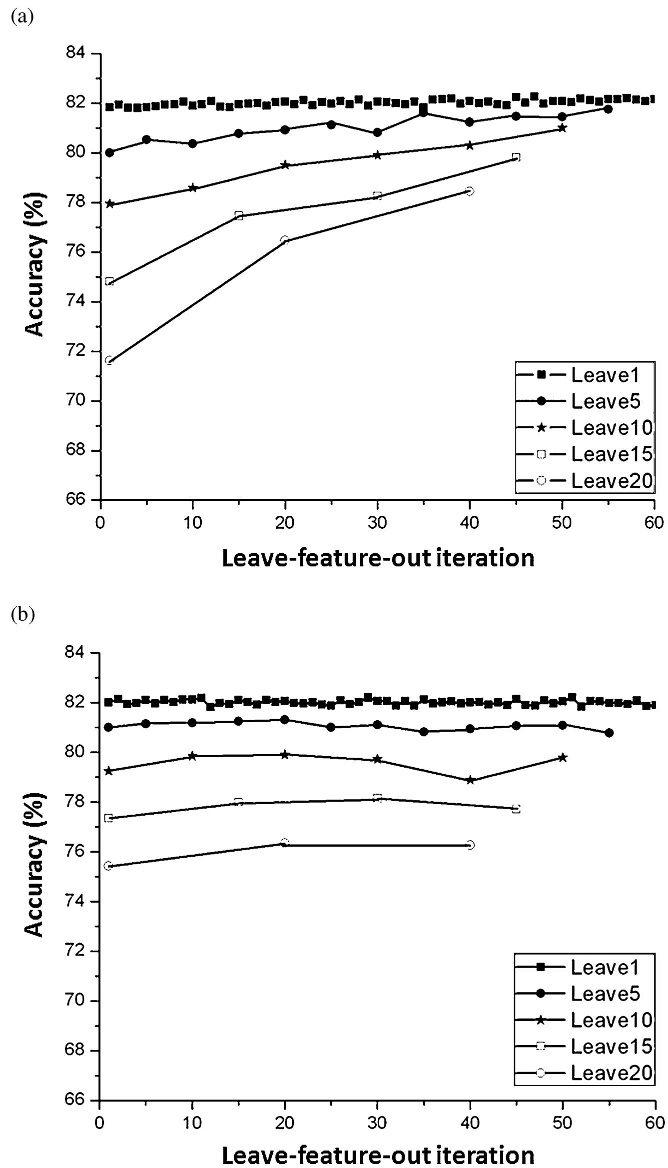


Fig. 1. Average performance obtained by iteratively removing N feature out based on (a) F-score sorting and (b) random selection. Each iteration use same number of features for classification, for example, $60 - 5 = 55$ features were used in each iteration in the case of leave 5, where the first point represents the removal of the first- to fifth-ranked (or randomly selected) features and the second data point represents the removal of the sixth to tenth features, etc.

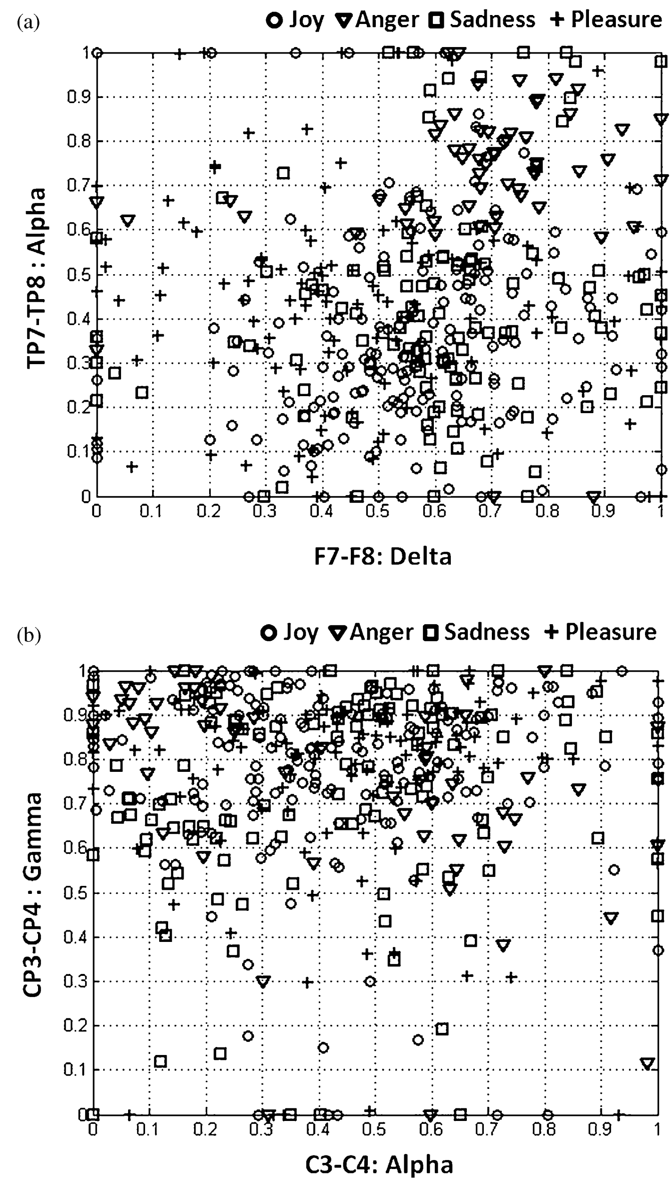


Fig. 2. Comparison of 2-D feature scatter plot from a sample subject along (a) top and (b) last two of F-score-ranked features.

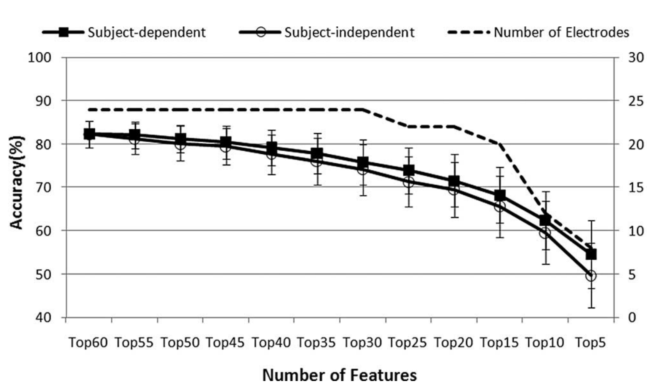


Fig. 3. Comparison of average results using top-N subject-dependent, subject-independent features, and the number of electrodes for subject-independent features, where top N was defined as the range of the F-score from the first to the Nth attributes.

the accuracy of 75.75% ± 5.24% obtained by using top-30 subject-dependent features. However, though the number of attributes was reduced from 60 to 30, the electrodes required to derive these top-30 features remained the same (24). Nevertheless, using only 30 of 60 attributes would considerably reduce the computational complexity.

Finally, Fig. 4 accesses the importance of the inclusion of particular electrode pairs by plotting the degree of use of each electrode in the top-30 subject-independent DASM12 features. As can be seen, the features derived from the frontal and parietal lobes were used more frequently than other regions, indicating these electrodes provided more discriminative information than other sites.

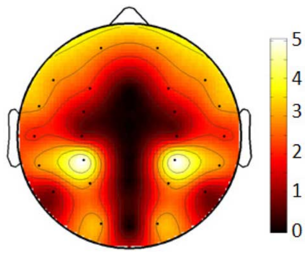


Fig. 4. Degree of use of each electrode in the top-30 subject-independent features. The degree of use is color coded, according to the color bar on the right. (The symmetry pattern is due to the fact that the spectral differences were derived from symmetrical pairs.)

In short, the frontal and parietal electrode pairs were most informative about the emotional states. By combining EEG spectral estimation and SVM, it is feasible to identify four emotional states (joy, anger, sadness, and pleasure) of participants during music listening. The maximum classification accuracy of 82.29% $\pm$ 3.06% could be obtained by applying SVM to DASM12.

V. Discussion

This study demonstrated the feasibility of using EEG dynamics to recognize emotional states in music listening. Several important issues were explored.

A. EEG Feature Types

The effects of emotional processing have been found with different temporal dynamics of the EEG during listening to different music excerpts [12], indicating EEG pattern would evolve over time during music listening. Thus, this study aimed to characterize the EEG dynamics accompanying emotion processing with second-by-second temporal resolution. Four different types of EEG features, DASM12, RASM12, PSD24 and PSD30, were derived from EEG recordings at different frequency bands. The results of this study (see Tables II and III) showed that the differential asymmetry of hemispheric EEG power spectra (DASM12) provided better classification accuracy than the rational asymmetry of hemispheric EEG power spectra (RASM12). Second, although DASM12 and PSD24 were recorded and derived from the same set of electrodes, the DASM12 features significantly improved classification performance. This result is highly in line with that hemispheric power asymmetry useful for the discrimination of mental tasks or similar work, as shown previously [22], [31]. Lastly, the classification accuracy using DASM12 outperformed that using PSD30, despite the fact that the feature dimensions of DASM12 were considerably lower than that of PSD30 (60: 12 electrode pairs $\times$ 5 frequency bands versus 150: 30 electrodes $\times$ 5 frequency bands).

B. Compare to Related Work

The best emotion classification (82.29% $\pm$ 3.06%) was obtained by SVM with a ten times of 10-fold cross-validation scheme based on DASM12. The use of STFT with a nonoverlap 1-s window, as opposed to our previous study using an overlapped window [27], made the results of this study more convincing, since the training and testing datasets were totally disjoint.

TABLE IV
Average SVM Confusion Matrix across Subjects Using DASM12

Input	Output			
	Joy (%)	Anger (%)	Sadness (%)	Pleasure (%)
Joy	86.15	1.29	5.10	7.46
Anger	10.35	74.11	6.69	8.84
Sadness	9.65	1.69	79.59	9.08
Pleasure	8.94	1.34	6.13	83.59

Table IV summarizes the average confusion matrix obtained by SVM applied to DASM12 across subjects. The best average accuracy for four emotional states was obtained for joy (86.15%), followed by pleasure, sadness, and anger with accuracy of 83.59%, 79.59%, and 74.11%, respectively. It is hard to compare the obtained accuracy of individual emotional states with previous literature, since the number of targeted emotional states varied from study to study. Therefore, the overall classification accuracy of the emotional states is compared next. Ishino and Hagiwara [16] proposed an emotion-estimation system for recognizing one of the four defined emotional states with the average accuracy ranging from 54.5% to 67.7% on single subject's dataset. Takahashi [17] reported an averaged recognition rate of 41.7% for distinguishing five emotional states in a film-induced emotional dataset from 12 subjects. Chanel *et al.* [18] obtained an average accuracy of 58% for distinguishing three emotional classes using image-arousal emotional dataset of four participants. In 2007, Heraz proposed a system to classify learner's status into one of the eight emotional states with a best accuracy of 82.27% on image-induced emotional dataset from 17 subjects. Further, recently Chanel *et al.* [20] reported another study on subject self-elicited emotional dataset of ten subjects and obtained a mean accuracy of 63% for three emotional classes. Zhang and Lee [22] presented a best result of 73.0% $\pm$ 0.33% for predicting two emotional states on image-induced emotional dataset from ten subjects. Although, the classification performance of this study (82.29% $\pm$ 3.06%) on 26 subjects was clearly better than those of previous works, it is however too premature to conclude the proposed method is superior to others as a variety of factors might affect the classification results, including, yet not limited to, experimental paradigms and conditions, stimulus types, and the number of induced emotions.

C. Subject-Independent Features

This study also explored the features that are most relevant to emotional process with F-score. Table V lists the top-30 F-score-ranked features across 26 subjects. In general, many features derived from the EEG sensors placed near the frontal lobe (see Fig. 4) consist with previous studies reported that the frontal lobe played a key role in emotional processing [32]. Sutton and Davidson [33] also showed that the prefrontal cortex played an important role in maintaining affective

TABLE V
TOP-30 FEATURE SELECTION RESULTS USING ACCUMULATED F-SCORE CRITERION

Rank	Electrode pair	Component	Brain lobe	Rank	Electrode pair	Component	Brain lobe
1	TP7-TP8	Gamma	Temporal	16	FT7-FT8	Beta	Frontal
2	FT7-FT8	Theta	Frontal	17	F3-F4	Gamma	Frontal
3	T7-T8	Delta	Temporal	18	F3-F4	Delta	Frontal
4	TP7-TP8	Delta	Temporal	19	TP7-TP8	Alpha	Temporal
5	F7-F8	Beta	Frontal	20	FP1-FP2	Delta	Frontal
6	O1-O2	Beta	Occipital	21	FP1-FP2	Gamma	Frontal
7	F7-F8	Delta	Frontal	22	P3-P4	Delta	Parietal
8	T7-T8	Theta	Temporal	23	O1-O2	Theta	Occipital
9	P3-P4	Theta	Parietal	24	CP3-CP4	Beta	Central
10	P7-P8	Theta	Parietal	25	P3-P4	Gamma	Parietal
11	FT7-FT8	Delta	Frontal	26	FC3-FC4	Delta	Frontal
12	C3-C4	Delta	Central	27	O1-O2	Gamma	Occipital
13	FP1-FP2	Alpha	Frontal	28	CP3-CP4	Gamma	Central
14	CP3-CP4	Theta	Central	29	CP3-CP4	Delta	Central
15	F7-F8	Theta	Frontal	30	CP3-CP4	Alpha	Central

representations, supporting the alpha-power asymmetry at (Fp1–Fp2) in the current study. The beta asymmetry at CP3–CP4 was also comparable with the finding that the parietal beta asymmetry at P3–P4 pair play a role in motivation and emotion [13]. Further, the involvement of the theta asymmetry at the frontal (F7–F8) and parietal (P3–P4 and P7–P8) sites were consistent with their roles in analyzing the emotional arousal during affective-pictures stimuli [11]. Moreover, the gamma-power asymmetry found in the parietal (P3–P4) region was consistent with [11] and [14], which claimed the feature served as a powerful tool for studying cortical activation during different level of emotional arousal induced by image stimuli. However, commonly reported alpha asymmetry at the F3–F4 pair related to valence emotion [7], [8] was missing from the top-30 list. This study also found additional features located at other brain regions and frequency bands related to emotional process. As noted in [34], there was still other EEG spectral power in different bands that may provide additional information about emotion, which was not reflected in the alpha activity (based on the predicted inverse relation with metabolism). Our results suggested that these distributed spectral asymmetries might provide meaningful information about emotional responses.

With respect to music stimulation in the human brain, many studies primarily focused on music structures such as timber, mode, tempo, rhythm, pitch, and melody. Analysis of the EEG during music perception suggested simultaneous and homogeneous activity in the multiple cortical regions [35]. For example, an enhancement of the delta-band activity has been shown widely distributed in the brain regions while listening to different music pieces [36]. Accordingly, there exists a considerable amount of dynamic changes of EEG patterns, which not only resulted from emotional responses, but also associated with music perception. Emotion in music, however, will be conveyed through the structure and rendering of the music itself [37]. A recent functional MRI (fMRI) study [38] showed that the manipulation of two major musical structures (mode and tempo) resulting in the variation of emotional perception eliciting the engagement of the brain structures (orbitofrontal and cingulate cortices) known to intervene in emotion processing. Consequently, we should note that EEG power changes resulted from the confounding factors, such as music perception and emotional processing, and cannot be easily dissociated from each other in our study. Nevertheless, both could contribute to characterize EEG power changes associated with the arousal and valence emotion dimensions with high classification performance.

In addition, Bhattacharya and Petsche in a music perception study [36] reported that only musicians retrieved extensive repertoire of musical patterns from their long-term musical memory, which was accompanied by an enhanced gamma-band synchrony, whereas delta-band synchrony over distributed cortical areas was significantly increased in nonmusicians. The results of this study however found several emotion-related EEG features in the gamma band from nonmusicians. But, the current study did not record sufficient EEG data from musician, and thus, could not compare the EEG dynamics during music appreciation between musicians and nonmusicians.

D. Electrode Reduction

It is worth noting that the top-30 subject-independent features might not be generally optimal for all individuals, they however provided some information about the common areas that are involved in emotional processing. Further, an acceptable accuracy could be achieved by removing 30 (50%) of 60 features of DASM12, greatly reducing the computational cost. However, the number of involved electrodes did not decrease appreciably (remained to be 12 electrode pairs). A closer look at Table V found three lightly used features: FC3–FC4 (delta), P7–P8 (theta) and C3–C4 (delta). Excluding these three features would reduce the number of required electrodes from 24 to 18 (nine electrode pairs) at the expense of a slight decrease in the classification accuracy (from 74.10% $\pm$ 5.85% to 72.75% $\pm$ 5.69%).

VI. Conclusion

This study has conducted a systematic EEG feature extraction and classification in order to assess the association between EEG dynamics and music-induced emotional states. The results of this study showed that DASM12, a spectral power asymmetry across multiple frequency bands, was a sensitive metric for characterizing brain dynamics in response to emotional states (joy, angry, sadness, and pleasure). A group of features extracted from the frontal and parietal lobes have been identified to provide discriminative information associated with emotion processing, which were relatively insensitive to subject-variability. The involvement of these features was largely consistent with previous literature. A machine-learning approach to classify four music-induced emotional states was proposed and tested in this study, which might provide a different viewpoint and new insights into music listening and emotion responses.

The future work includes a further evaluation of the specific link between EEG dynamics, emotional responses, and music structures to dissociate the brain responses to the music perception, music appreciation, as well as music-induced emotions. We expect that further understanding the different stages of how the brain processes music information will make an impact on the realization of novel EEG-inspired multimedia applications, where the contents of multimedia will be meaningfully inspired by users' feedback.

References

[1] D. Wu, C. Y. Li, and D. Z. Yao, "Scale-free music of the brain," *PLoS One*, vol. 4, no. 6, p. e5915, Jun. 15, 2009.

[2] J. Kim and E. Andre, "Emotion recognition using physiological and speech signal in short-term observation," in *Proc. Percept. Interactive Technol.*, 2006, vol. 4021, pp. 53–64.

[3] J. Kim and E. Andre, "Emotion recognition based on physiological changes in music listening," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 12, pp. 2067–2083, Dec. 2008.

[4] K. H. Kim, S. W. Bang, and S. R. Kim, "Emotion recognition system using short-term monitoring of physiological signals," *Med. Biol. Eng. Comput.*, vol. 42, no. 3, pp. 419–427, May 2004.

[5] R. W. Picard, E. Vyzas, and J. Healey, "Toward machine emotional intelligence: Analysis of affective physiological state," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 10, pp. 1175–1191, Oct. 2001.

[6] D. Mantini, M. G. Perrucci, C. Del Gratta, G. L. Romani, and M. Corbetta, "Electrophysiological signatures of resting state networks in the human brain," *Proc. Nat. Acad. Sci. USA*, vol. 104, no. 32, pp. 13170–13175, Aug. 2007.

[7] J. J. B. Allen, J. A. Coan, and M. Nazarian, "Issues and assumptions on the road from raw signals to metrics of frontal EEG asymmetry in emotion," *Biol. Psychol.*, vol. 67, no. 1/2, pp. 183–218, Oct. 2004.

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

[9] W. Heller, "Neuropsychological mechanisms of individual differences in emotion, personality and arousal," *Neuropsychology*, vol. 7, pp. 476–489, 1993.

[10] M. Sarlo, G. Buodo, S. Poli, and D. Palomba, "Changes in EEG alpha power to different disgust elicitors: The specificity of mutilations," *Neurosci. Lett.*, vol. 382, no. 3, pp. 291–296, Jul. 2005.

[11] 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," *Neurosci. Behav. Physiol.*, vol. 34, no. 8, pp. 859–867, 2004.

[12] 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, Mar. 2007.

[13] D. J. L. Schutter, P. Putman, E. Hermans, and J. van Honk, "Parietal electroencephalogram beta asymmetry and selective attention to angry facial expressions in healthy human subjects," *Neurosci. Lett.*, vol. 314, no. 1/2, pp. 13–16, Nov. 2001.

[14] M. Balconi and C. Lucchiari, "Consciousness and arousal effects on emotional face processing as revealed by brain oscillations. A gamma band analysis," *Int. J. Psychophysiol.*, vol. 67, no. 1, pp. 41–46, Jan. 2008.

[15] E. Basar, C. Basar-Eroglu, S. Karakas, and M. Schurmann, "Oscillatory brain theory: A new trend in neuroscience—The role of oscillatory processes in sensory and cognitive functions," *IEEE Eng. Med. Biol. Mag.*, vol. 18, no. 3, pp. 56–66, May/Jun. 1999.

[16] K. Ishino and M. Hagiwara, "A feeling estimation system using a simple electroencephalograph," in *Proc. IEEE Int. Conf. Syst., Man Cybern.*, 2003, vol. 5, pp. 4204–4209.

[17] K. Takahashi, "Remarks on emotion recognition from bio-potential signals," in *Proc. 2nd Int. Conf. Auton. Robots Agents*, Dec. 13–15, 2004, pp. 186–191.

[18] G. Chanel, J. Kronegg, D. Grandjean, and T. Pun, "Emotion assessment: Arousal evaluation using EEG's and peripheral physiological signals," *Multimedia Content Representation, Classification Secur.*, vol. 4105, pp. 530–537, 2006.

[19] A. Heraz, R. Razaki, and C. Frasson, "Using machine learning to predict learner emotional state from brainwaves," in *Proc. 7th IEEE Int. Conf. Adv. Learning Technol.*, 2007, pp. 853–857.

[20] G. Chanel, J. J. M. Kierkels, M. Soleymani, and T. Pun, "Short-term emotion assessment in a recall paradigm," *Int. J. Human-Comput. Stud.*, vol. 67, no. 8, pp. 607–627, Aug. 2009.

[21] K. E. Ko, H. C. Yang, and K. B. Sim, "Emotion recognition using EEG signals with relative power values and bayesian network," *Int. J. Control Autom. Syst.*, vol. 7, no. 5, pp. 865–870, Oct. 2009.

[22] Q. Zhang and M. H. Lee, "Analysis of positive and negative emotions in natural scene using brain activity and GIST," *Neurocomputing*, vol. 72, no. 4–6, pp. 1302–1306, Jan. 2009.

[23] J. A. Russell, "A circumplex model of affect," *J. Pers. Soc. Psychol.*, vol. 39, no. 6, pp. 1161–1178, 1980.

[24] T. L. Wu and S. K. Jeng, "Probabilistic estimation of a novel music emotion model," in *Proc. 14th Int. Multimedia Model. Conf.*, Kyoto, Japan, 2008, pp. 487–497.

[25] R. Cowie, E. Douglas-Cowie, S. Savvidou, E. McMahon, M. Sawey, and M. Schröder, "'FEELTRACE': An instrument for recording perceived emotion in real time," in *Proc. ISCA Workshop Speech Emotion*, 2000, pp. 19–24.

[26] Y. P. Lin, C. H. Wang, T. L. Wu, S. K. Jeng, and J. H. Chen, "Multilayer perceptron for EEG signal classification during listening to emotional music," in *Proc. IEEE Int. Region 10 Conf.*, Taipei, Taiwan, 2007, pp. 1–3.

[27] Y. P. Lin, C. H. Wang, T. L. Wu, S. K. Jeng, and J. H. Chen, "Support vector machine for EEG signal classification during listening to emotional music," in *Proc. IEEE Int. Workshop Multimedia Signal Process.*, Queensland, Australia, 2008, pp. 127–130.

[28] Weka, *Data Mining: Practical Machine Learning Tools and Techniques*, 2nd ed. San Francisco, CA: Morgan Kaufmann, 2005.

[29] C. C. Chang and C. J. Lin. (2001). LIBSVM: A library for support vector machines [Online]. Available: Software available at http://www.csie.ntu.edu.tw/∼cjlin/libsvm

[30] Y. W. Chen and C. J. Lin, "Combining SVMs with various feature selection strategies," in *Feature Extraction, Foundations and Applications*. New York: Springer-Verlag, 2006.

[31] R. Palaniappan, "Utilizing gamma band to improve mental task based brain-computer interface design," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 14, no. 3, pp. 299–303, Sep. 2006.

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

[33] S. K. Sutton and R. J. Davidson, "Prefrontal brain electrical asymmetry predicts the evaluation of affective stimuli," *Neuropsychologia*, vol. 38, no. 13, pp. 1723–1733, 2000.

[34] R. J. Davidson, "What does the prefrontal cortex "Do" in affect: Perspectives on frontal EEG asymmetry research," *Biol. Psychol.*, vol. 67, no. 1/2, pp. 219–233, Oct. 2004.

[35] P. E. Andrade and J. Bhattacharya, "Brain tuned to music," *J. R. Soc. Med.*, vol. 96, no. 6, pp. 284–287, Jun. 2003.

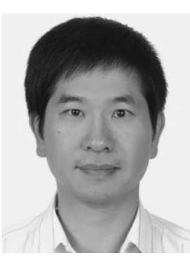
[36] J. Bhattacharya and H. Petsche, "Phase synchrony analysis of EEG during music perception reveals changes in functional connectivity due to musical expertise," *Signal Process.*, vol. 85, no. 11, pp. 2161–2177, Nov. 2005.
[37] C. D. Tsang, L. J. Trainor, D. L. Santesso, S. L. Tasker, and L. A. Schmidt, "Frontal EEG responses as a function of affective musical features," *Biol. Found. Music*, vol. 930, pp. 439–442, 2001.
[38] S. Khalfa, D. Schon, J. L. Anton, and C. Liegeois-Chauvel, "Brain regions involved in the recognition of happiness and sadness in music," *Neuroreport*, vol. 16, no. 18, pp. 1981–1984, Dec. 2005.

Yuan-Pin Lin received the B.S degree in biomedical engineering from Chung Yuan Christian University, Chung Li, Taiwan, in July 2003, and the M.S. degree in the electrical engineering from the National Taiwan University (NTU), Taipei, Taiwan, in July 2005, where he is currently working toward the Ph.D. degree in electrical engineering.

He is currently also with the Swartz Center for Computational Neuroscience, University of California, San Diego. His research interests include biomedical signal processing, machine learning, and human–machine interaction.

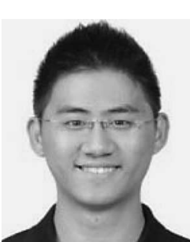
Chi-Hong Wang received the M.D. degree and the M.S. degree in electric engineering from the National Taiwan University, Taipei, Taiwan, in 1996 and 2005, respectively.

He is currently an Attending Physician in the Department of Neurology, Cardinal Tien Hospital, Yung-Ho branch, Taipei. His research interests include functional MRI analysis and biomedical signal processing.



Tzyy-Ping Jung (S'91–M'92–SM'06) received the B.S. degree in electronics engineering from the National Chiao Tung University, Hsinchu, Taiwan, in 1984, and the M.S. and Ph.D. degrees in electrical engineering from The Ohio State University, Columbus, in 1989 and 1993, respectively.

He was a Research Associate in the Computational Neurobiology Laboratory, The Salk Institute, San Diego, CA. He is currently a Research Scientist at the Institute for Neural Computation and an Associate Director of the Swartz Center for Computational Neuroscience, University of California, San Diego. He is also a Professor in the Department of Computer Science, National Chiao Tung University. His research interests include areas of biomedical signal processing, cognitive neuroscience, machine learning, time-frequency analysis of human electroencephalogram, functional neuroimaging, and brain–computer interfaces and interactions.



Tien-Lin Wu received the B.S. degree in electrical engineering from the National Taiwan University, Taipei, Taiwan, in 2005, where he is currently working toward the Ph.D. degree from the Department of Electrical Engineering.

His current research interests include multimedia information retrieval and analysis, human-centric computing, and affective computing.

Shyh-Kang Jeng received the B.S.E.E. and Ph.D. degrees from the National Taiwan University (NTU), Taipei, Taiwan, in 1979 and 1983, respectively.

In 1981, he joined the faculty of the Department of Electrical Engineering, NTU, where he is currently a Professor. From 1985 to 1993, he was a Visiting Research Associate Professor and a Visiting Research Professor at the University of Illinois, Urbana-Champaign. In 1999, he was with the Center for Computer Research in Music and Acoustics, Stanford University, Stanford, CA, for half of a year. His research interest includes numerical electromagnetics, ultrawide-band wireless system, music signal processing, music information retrieval, computational neuroscience, and electromagnetic scattering analysis.

Jeng-Ren Duann (S'90–A'98–M'04) received the B.S. and M.S. degrees in biomedical engineering and the Ph.D. degree in physics from Chung Yuan Christian University, Chung Li, Taiwan, in 1990, 1992, and 1999, respectively.

He was a Research Associate in the Computational Neurobiology Laboratory, The Salk Institute for Biological Studies, La Jolla, CA. Since 2004, he has been an Assistant Project Scientist at the Institute for Neural Computation, University of California, San Diego. He is currently an Associate Director of the Biomedical Engineering Research and Development Center, China Medical University Hospital, Taichung, Taiwan. He is the author of FMRLAB, a freely downloadable MATLAB toolbox for functional neuroimaging data analysis using independent component analysis. His research interests include biomedical signal and image processing, biosystem simulation and modeling, structural and functional human brain mapping and their applications in cognitive neuroscience, and functional cardiac imaging.

Jyh-Horng Chen (S'89–M'91) received the B.S. degree in electrical engineering from the National Taiwan University (NTU), Taipei, Taiwan, in 1982, the M.S. degree in medical engineering from the National Yang-Ming University, Taipei, in 1986, and the Ph.D. degree in the intercampus Bioengineering Program from the University of California, Berkeley and San Francisco, in 1991.

In 1991, he joined the faculty of Department of Electrical Engineering, NTU, as an Associate Professor, where he has been a Professor since 2000 and the Chair of Institute Biomedical Engineering since 2002. His current research interests include general medical imaging systems design, sensory-aid design, biological signal detection, man–machine interface system, and medical informatics.

Dr. Chen is a Member of the Administration Committee of IEEE/Engineering in Medicine and Biology Society, the International Society for Magnetic Resonance in Medicine, and the Society of Molecular Imaging.