Toward an EEG-Based Recognition of Music Liking Using Time-Frequency Analysis

Article *in* IEEE transactions on bio-medical engineering · September 2012

DOI: 10.1109/TBME.2012.2217495 · Source: PubMed

2 authors:

Stelios K. Hadjidimitriou
Aristotle University of Thessaloniki
49 PUBLICATIONS **356** CITATIONS

SEE PROFILE

Leontios Hadjileontiadis
Aristotle University of Thessaloniki
328 PUBLICATIONS **5,122** CITATIONS

SEE PROFILE

Toward an EEG-Based Recognition of Music Liking Using Time-Frequency Analysis

Stelios K. Hadjidimitriou, *Student Member, IEEE*, and Leontios J. Hadjileontiadis*, *Senior Member, IEEE*

***Abstract*—Affective phenomena, as reflected through brain activity, could constitute an effective index for the detection of music preference. In this vein, this paper focuses on the discrimination between subjects' electroencephalogram (EEG) responses to self-assessed liked or disliked music, acquired during an experimental procedure, by evaluating different feature extraction approaches and classifiers to this end. Feature extraction is based on time–frequency (TF) analysis by implementing three TF techniques, i.e., spectrogram, Zhao–Atlas–Marks distribution and Hilbert–Huang spectrum (HHS). Feature estimation also accounts for physiological parameters that relate to EEG frequency bands, reference states, time intervals, and hemispheric asymmetries. Classification is performed by employing four classifiers, i.e., support vector machines, k-nearest neighbors (k-NN), quadratic and Mahalanobis distance-based discriminant analyses. According to the experimental results across nine subjects, best classification accuracy $\{$86.52 ($\pm$0.76)%$\}$ was achieved using k-NN and HHS-based feature vectors (*FV*s) representing a bilateral average activity, referred to a resting period, in *beta* (13–30 Hz) and *gamma* (30–49 Hz) bands. Activity in these bands may point to a connection between music preference and emotional arousal phenomena. Furthermore, HHS-based *FV*s were found to be robust against noise corruption. The outcomes of this study provide early evidence and pave the way for the development of a generalized brain computer interface for music preference recognition.**

***Index Terms*—Electroencephalogram (EEG), machine learning, music liking/disliking, time–frequency (TF) analysis.**

I. Introduction

BIOINSPIRED human–computer interaction is expected to be the next breakthrough in the field of multimedia systems. In this direction, vast amount of research has been conducted toward the development of effective human–computer interfaces based on biological signals. For example, the emerging field of affective computing fosters the recognition, processing, and simulation of human affects by computer-based systems through physiological data. In this vein, the ultimate goal of the research presented in this study is to pave the way for the development of a brain–computer interface (BCI) for music preference recognition, by exploring the potential of electroencephalogram (EEG) responses.

The preference toward a musical genre or specific musical pieces forms a distinctive characteristic of our personality, which has a significant impact on our everyday habits and social interactions. This inclination is woven by numerous factors that have emerged gradually during human evolution [1]. Although many of the factors that drive preferences toward some pieces over others are subjective and vary across cultures, affective responses when listening to liked (or disliked) music tend to be common. Even unfamiliar music can elicit such responses from listeners [2]. In the psychology-based literature music appraisal is mainly interpreted in terms of affective experiences, such as emotional resonance, aesthetic awe, and thrill [3]–[5]. Nevertheless, this affective response is somehow reflected over the human brain.

In the field of neurophysiology, evidence of brain activation due to emotion-evocative music is mainly reported. The vast amount of research implicates the theory of valence and arousal in order to model different emotional states [6]. Positive valence corresponds to positive emotions and negative valence to negative emotions, whereas arousal constitutes a measure of excitation. A series of studies, by employing neuroimaging techniques, mainly functional magnetic resonance imaging, have revealed activation of cortical and subcortical brain structures relating to emotion processing, during pleasant or unpleasant music listening (ML) [7]–[9]. Another series of studies used EEG signals to examine affective responses to music. Altenmüller *et al.* [10] found that pleasant music causes left frontal brain activation, while unpleasant music causes right and slightly bilateral frontal activation. Similar evidence has been produced by the study of Schmidt and Trainor [11], i.e., asymmetric activation during the listening of positively valenced music (left frontal activity) and negatively valenced music (right frontal activity). Another EEG-based study showed that consonant (pleasant) music causes greater midline brain activity compared to dissonant music (unpleasant) [12]. Finally, as far as EEG-based emotion recognition is concerned, the only characteristic example can be found in the work of Lin *et al.* [13], where spectrogram (SPG)-based feature extraction was used, in order to classify distinct emotions due to ML. Classification results, using support vector machines (SVM), showed that spectral hemispheric asymmetry, derived from all EEG frequency bands (fbs), can lead to an accuracy of 82.29 $\pm$ 3.06% for the correct recognition of four emotions. However, the aforementioned electrophysiological evidence arises from the listening of positively/negatively

Manuscript received May 3, 2012; revised July 26, 2012; accepted August 30, 2012. Date of publication September 27, 2012; date of current version November 22, 2012. *Asterisk indicates corresponding author.*

S. K. Hadjidimitriou is with the Department of Electrical and Computer Engineering, Aristotle University of Thessaloniki, GR-54124 Thessaloniki, Greece (e-mail: stelios@psyche.ee.auth.gr).

*L. J. Hadjileontiadis is with the Department of Electrical and Computer Engineering, Aristotle University of Thessaloniki, GR-54124 Thessaloniki, Greece (e-mail: leontios@auth.gr).

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.2012.2217495

valenced music that cannot be directly mapped to liked/disliked music. For instance, depending on the listeners' mood, sad-sounding musical excerpts can be occasionally preferred instead of happy-sounding ones [14].

In general, due to their noninvasive recording procedure and temporal resolution, EEG signals have been widely used in order to study brain activity relating to affective responses. Evidence of such activity is reported in the majority of EEG fbs, i.e., *theta* (4–8 Hz), *alpha* (8–13 Hz), *beta* (13–30 Hz), and *gamma* (30–49 Hz) [15], [16]. Frontal midline *theta* power modulation is suggested to reflect affective processing during consonant or dissonant music [12]. The *alpha*-power asymmetry on the prefrontal cortex has been proposed as an index for the discrimination between positively and negatively valenced emotions [11], [15]. *Beta* activity has been associated with emotional arousal modulation [17] and moreover, asymmetric activity in this band is linked to the emotional dimensions of approach or withdrawal [18]. Finally, *gamma* band is mainly related to arousal effects [19], [20].

In this framework, this study focuses on music preference and the user-independent classification of brain responses to liked and disliked music, using EEG data. For this purpose, subjects participated in an experiment during which they listened to and rated musical excerpts according to their preference, while their electrophysiological activity was recorded from electrode sites across the brain. The subsequent classification procedure involved feature extraction methods based on time–frequency (TF) analyses. The latter were selected as a means of exploiting the energy-related characteristics of EEG signals that may vary both in time and frequency. Epigrammatically, the main objective of the study is to implement and compare different feature extraction approaches and classifiers, in order to establish a methodological base for the efficient recognition of music liking. Toward such endeavor, this study systematically investigates the recognition of music preference based on electrophysiological data and employs TF signal processing techniques and machine learning to this end.

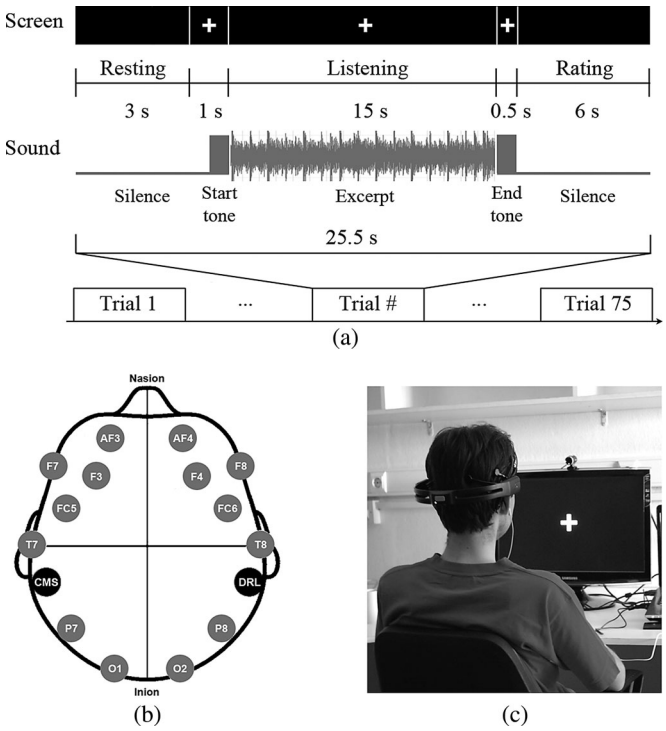


Fig. 1. (a) Trial structure adopted in the experimental protocol. (b) Electrode positions, according to the 10–20 system, of the Emotiv EPOC device used for EEG acquisition. (c) Experimental configuration.

II. Materials and Methods

A. Experiment and Data Acquisition

Nine subjects participated in the experiment (two females and seven males; age 23.22 $\pm$ 1.72 years). All of them were undergraduate students of the Department of Electrical and Computer Engineering, Aristotle University of Thessaloniki, Greece, with no special music education background. Subjects were right handed and they did not report any hearing impairments.

During the experimental procedure, each subject listened to 60 musical excerpts from four common musical genres, i.e., rock pop, electronic, jazz, and classical (15 excerpts per genre) and 15 excerpts of broadband noise. Noise was used to establish a reference state (RS), as it was expected, in general, to be extremely unpleasant, thus not liked. Vocal music was not included, in order to avoid brain activation due to lyrics processing. All excerpts were 15-s long and they were presented pseudorandomly in terms of type (music or noise) and musical genre. At the end of each excerpt, subjects were asked to impulsively rate their liking on a five point scale (1: do not like at all; 2: do not like; 3: undecided; 4: like; and 5: like very much). The total number of trials was equal to the number of acoustic excerpts, i.e., 75 per subject. The structure of each trial is shown in Fig. 1(a). In the beginning, a 3-s interval of relaxation preceded, followed by the projection of a white fixation cross on a black background until the end of each excerpt. A 500-ms 440-Hz-tone marked the beginning and the end of the excerpts that were presented with 500-ms fade in and fade out effects, in order to avoid startling. At the end of each excerpt, a rating form was presented to the subjects for 6 s. The interval of 9.5 s between the end of a musical excerpt and the beginning of the next one falls into in the range of 3–15-s-long interstimulus intervals that were adopted in related studies [8], [12], [13], [16]. However, it cannot be clearly stated that an interval in the order of seconds is long enough for "emotional resetting" to take place, if that is even possible. From a psychological aspect, emotional events are suggested to last only for seconds, but enduring emotional states, seen as a prolonged triggering of emotional episodes, can even last for hours [21].

EEG recordings were conducted using the Emotiv EPOC 14-channel EEG wireless recording headset (Emotiv Systems, Inc., San Francisco, CA). The electrode scheme was arranged according to the international 10–20 system and included active electrodes at AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, and AF4 positions, referenced to the common mode sense (CMS—left mastoid)/driven right leg (DRL—right mastoid) ground [Fig. 1(b)]. EEG data are acquired with an internal

sampling frequency of 2048 Hz by the recording device, and, subsequently, bandpass filtered in the range of 0.16–85 Hz using hardware filters. Two digital notch filters at 50 and 60 Hz are further applied, and the output data are downsampled to 128 Hz. The portability and easy-to-wear characteristics of the recording device favor its use in a future BCI system. As there was no EEG electrode cap used, special care was taken in order for the device to be adapted in an appropriate manner on each subject's head. The experimental procedure took place in a laboratory environment, under dim lighting conditions, to avoid any visual nuisance. Subjects sat comfortably on a chair with a distance of 1 m from a computer screen [Fig. 1(c)]. They were instructed to remain as still as possible, in order to reduce muscular artifacts, and to focus their sight on the cross, whenever it was presented, in order to minimize occipital artifacts. The acoustic stimuli were provided to them by in-ear headphones on a volume level that varied according to each subject's preference, to ensure comfort.

All acoustic stimuli were prepared in Adobe Audition 3.0 (Adobe Systems, Inc., San Jose, CA). Experimental protocol was developed using Microsoft Visual Studio 2010 (Microsoft Corporation, Redmond, WA).

B. Feature Extraction

After the completion of data acquisition, all EEG signals were filtered offline in the frequency range of 1–49 Hz (IIR Butterworth 6th order filter with zero-phase shift). The focus was placed upon the five EEG fbs, i.e., *delta* (1–4 Hz), *theta* (4–8 Hz), *alpha* (8–13 Hz), *beta* (13–30 Hz), and *gamma* (30–49 Hz), thus, features were estimated for each of these bands. Initially, for each trial and recording channel the TF representation was obtained. Subsequent feature estimation was based on the event-related synchronization/desynchronization (ERS/ ERD) theory [22], in order for the features to reflect quantities relevant to brain activation. In general, each feature F was computed as

$$F = \frac{A - R}{R} \tag{1}$$

where A represents the quantity estimated during ML and R represents the quantity estimated from a RS. A, for each trial j and channel i, was computed as the average of the TF values in each of the five EEG fbs and in time

$$A^{\mathrm{fb},w_k}(i,j) = \frac{1}{N_{w_k}} \sum_t \left(\frac{1}{N_{\mathrm{fb}}} \sum_f \mathrm{TF}^{\mathrm{ML}}_{i,j}[t,f] \right) \tag{2}$$

where $[t, f]$ represents the discrete (time, frequency) point in the TF plane, $\mathrm{TF}^{\mathrm{ML}}$ is the obtained TF representation during ML, and N_w, N_{fb} denote the number of data points in the time window ($w_k, k = 1{,}2$) and the number of data points in the fb of interest, respectively. A was computed in two time intervals (w_k): 1) the total duration (w_1) of ML (15 s), and 2) a time window of 8– 15 s (w_2) after each excerpt onset, when an emotional response was expected according to [23]. The TF plane partition for the estimation of (2) is depicted in Fig. 2. Probing further, two RSs were considered, the first corresponding to the resting period (RP) and the second one corresponding to the listening of noise, along with the case of no referencing (NoR) for the sake of completion:

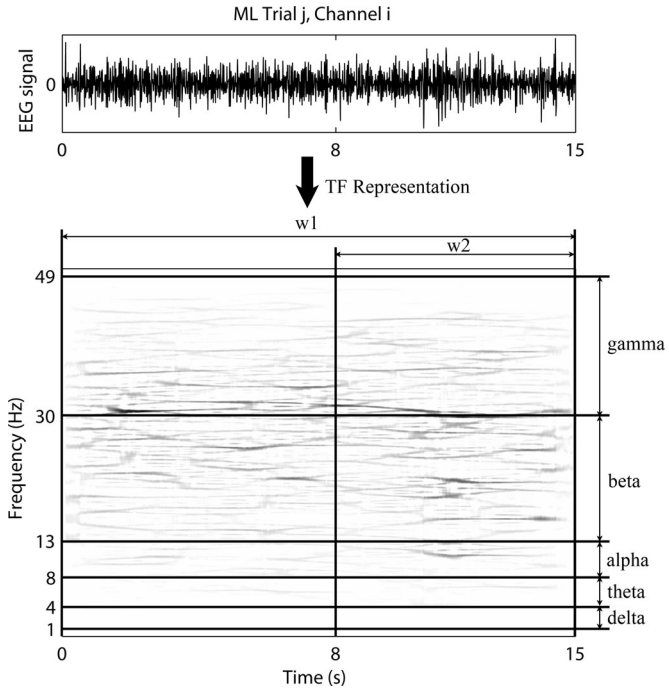


Fig. 2. TF plane partition for the estimation of A^{fb,w_k}, in (2), from the TF representation of the EEG signal corresponding to channel i and trial j of ML.

1) *RP RS:* For the RP RS, R was estimated by implementing (2) in the last 2 s of the resting interval, preceding each excerpt presentation:

$$R^{\mathrm{fb}}(i,j) = \frac{1}{N_{[2-3\ s]}} \sum_t \left(\frac{1}{N_{\mathrm{fb}}} \sum_f \mathrm{TF}^{\mathrm{RP}}_{i,j}[t,f] \right) \tag{3}$$

where $\mathrm{TF}^{\mathrm{RP}}$ is the obtained TF representation during RP, and $N_{[2-3\ s]}$ denotes the number of data points in the time interval of 2–3 s. The selection of the resting interval complies with the practices adopted in studies investigating ERD/ERS patterns [24], [25]. The first second was omitted to avoid any resting brain activity, as it succeeded subjects' rating. Thus, (1) can be written as

$$F^{\mathrm{fb},w_k}_{\mathrm{RP}}(i,j) = \frac{A^{\mathrm{fb},w_k}(i,j) - R^{\mathrm{fb}}(i,j)}{R^{\mathrm{fb}}(i,j)}. \tag{4}$$

2) *NL RS:* In the case of NL RS, R was computed by averaging the TF values estimated in the 0–15 s interval (w_1) over all trials of noise listening for all subjects, i.e.,

$$\bar{R}^{\mathrm{fb}}(i) = \frac{1}{N} \sum_j \left(\frac{1}{N_{w_1}} \sum_t \left(\frac{1}{N_{\mathrm{fb}}} \sum_f \mathrm{TF}^{\mathrm{NL}}_{i,j}[t,f] \right) \right) \tag{5}$$

where $\mathrm{TF}^{\mathrm{NL}}$ is the obtained TF representation during NL and N is the total number of NL trials, i.e., 135 (15 trials $\times$ 9 subjects). Thus, in the case of NL referencing, (1) can

be written as

$$F_{\mathrm{NL}}^{\mathrm{fb},w_k}(i,j) = \frac{A^{\mathrm{fb},w_k}(i,j) - \bar{R}^{\mathrm{fb}}(i)}{\bar{R}^{\mathrm{fb}}(i)}. \quad (6)$$

Averaging over all trials was adopted due to the expectation that NL would have evoked a common response among subjects.

3) *NoR:* In this case, features were estimated without referencing, i.e., $F_{\mathrm{NoR}}^{\mathrm{fb},w_k}(i,j) = A^{\mathrm{fb},w_k}(i,j)$.

For each EEG fb and time window w_k, feature vector *FV* corresponding to each trial j was constructed as

$$\mathrm{FV}_{\mathrm{RS}}^{\mathrm{fb},w_k}(j) = \{F_{\mathrm{RS}}^{\mathrm{fb},w_k}(1,j), \ldots, F_{\mathrm{RS}}^{\mathrm{fb},w_k}(14,j)\} \quad (7)$$

where RS denotes each referencing approach (RP, NL, and NoR). So far, *FV*s were formed without taking into consideration asymmetric activations that occur during emotional responses (NoAs case). In order to take into account such activations, two additional types of features—and consequently *FV*s—were considered. The first type represents a differential asymmetry index (DAs case), computed by subtracting the estimated features F for each of the seven symmetric channel pairs, i.e., AF3–AF4, F7–F8, F3–F4, FC5–FC6, T7–T8, P7–P8, and O1–O2. The second type represents a rational asymmetry index (RAs case), computed by the division of the estimated features F for each of the symmetric channel pairs. Thus, *FV*s in the case of DAs (*dFV*) and RAs (*rFV*) can be expressed as

$$d\mathrm{FV}_{\mathrm{RS}}^{\mathrm{fb},w_k}(j) = \{F_{\mathrm{RS}}^{\mathrm{fb},w_k}(i,j) - F_{\mathrm{RS}}^{\mathrm{fb},w_k}(i*,j)\}$$
$$r\mathrm{FV}_{\mathrm{RS}}^{\mathrm{fb},w_k}(j) = \left\{ \frac{F_{\mathrm{RS}}^{\mathrm{fb},w_k}(i,j)}{F_{\mathrm{RS}}^{\mathrm{fb},w_k}(i*,j)} \right\} \quad i = 1, \ldots, 7 \quad (8)$$

where i^* denotes the symmetric channel of the ith channel.

A schematic representation of the feature extraction and *FV* construction procedure is shown in Fig. 3. To sum up, the number of *FV* sets, derived from the aforementioned procedure, was 270 [5 EEG frequency bands (*delta*, *theta*, *alpha*, *beta*, *gamma*) × 2 time windows (0–15 s, 8–15 s) × 3 referencing states (RP, NL, NoR) × 3 types of asymmetry (NoAs, DAs, RAs) × 3 TF techniques]. Additional combinations of the best performing *FV*s, derived from different EEG frequency bands, were also formed and subjected to the classification procedure. A detailed description of the combined *FV*s is provided in Section III for the sake of clarity.

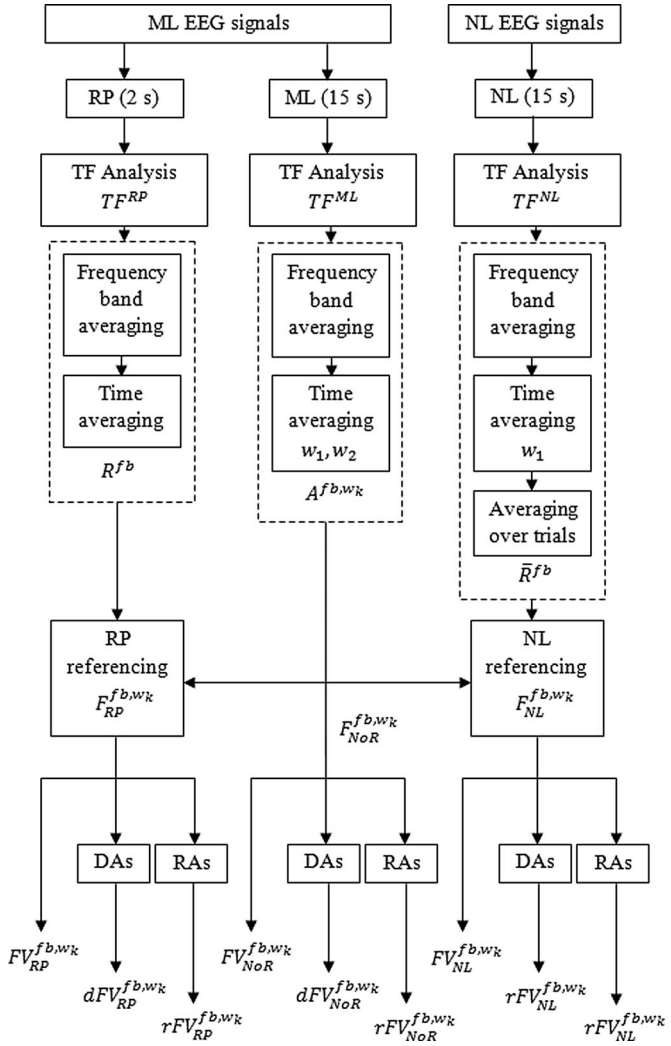


Fig. 3. Schematic representation of the feature extraction and *FV* construction procedure.

C. TF Techniques

Three TF techniques were used in the context of the feature extraction methodology described in the previous section. An epitomized description of these techniques, along with their implementation issues, follows.

1) SPG: This technique falls into the category of atomic decompositions where the TF distribution is acquired through a linear decomposition to elementary components, namely atoms, and it is based on the short-time Fourier transform (STFT). In particular, for a time series $X(t)$, the SPG-based TF representation is obtained by the squared modulus of the STFT of the series:

$$\mathrm{SPG}(t,f) = \left| \int_{-\infty}^{+\infty} X(\tau) h^*(\tau - t) e^{-j2\pi f \tau} d\tau \right|^2 \quad (9)$$

where $h^*(\tau - t)$ represents the short-time analysis window.

Here, the SPG technique was implemented in a similar way as in [12]. A 128-point fast fourier transform (FFT) with a nonoverlapped Hamming window of 1 s was used to this end.

2) Zhao–Atlas–Marks (ZAM) Distribution: The ZAM distribution [26] belongs to the category of quadratic TF representations, and especially, to the group of reduced interference distributions (RIDs). RIDs are members of the Cohen's class and thus, for a time series $X(t)$, they can be described by the general expression

$$\mathrm{RID}_X(t,f:\Phi) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \Phi(\xi\tau) A_X(\xi,\tau) e^{-j2\pi(f\tau + \xi t)} d\xi d\tau \quad (10)$$

where t and f denote time and frequency, respectively, while τ and ξ denote the delay and the doppler, respectively, in the ambiguity plane. $A_X(\xi,\tau)$ represents the ambiguity function which is associated with the Wigner–Ville distribution via a 2-D Fourier transform [27]. $\Phi(\xi\tau)$ is the, so-called, parameterization or kernel function. The ZAM distribution is derived by choosing the kernel function as

$$\Phi(\xi,\tau) = h(\tau)|\tau|\frac{\sin(\pi\xi\tau)}{\pi\xi\tau} \tag{11}$$

where $h(\tau)$ is a window function that leads to smoothing along the frequency axis. Thus, the following expression can be obtained which defines the ZAM distribution

$$\mathrm{ZAM}_X(t,f) = \int_{-\infty}^{+\infty}\left[h(\tau)\int_{t-|\tau|/2}^{t+|\tau|/2} X(s+t/2)X^*(s-t/2)ds\right]e^{-j2\pi f\tau}d\tau. \tag{12}$$

ZAM distribution was selected among RIDs due to its advantage of significantly reducing cross terms between signal components, through its cone-shaped kernel function (11) [26]. In this study, the ZAM-based TF representation was computed under a $N\times N$ TF resolution; N denotes the number of samples of the signal. Smoothing was performed using Hamming windows of N/10-samples and N/4-samples for time and frequency, respectively.

3) Hilbert–Huang Spectrum (HHS): The estimation of the HHS consists of the empirical mode decomposition (EMD) of a signal and the subsequent application of the Hilbert transform in order for a TF representation to be obtained [28]. Consider a time series $X(t)$; the EMD method decomposes $X(t)$, via a sifting process, in N intrinsic mode functions (IMFs). The sifting process is iterative and produces different versions of the signal based on its local extrema and zero crossings, till the final version becomes an IMF. At first, the upper and lower envelopes of the signal $X(t)$ of T duration are constructed based on its local maxima and minima, and their mean waveform is subtracted from $X(t)$, producing a signal $h_1(t)$. The latter is subjected to n iterations of the sifting process until it becomes an IMF. The stopping criterion for the sifting process mandates that the standard deviation (*SD*) between successive versions is smaller than a predetermined value, i.e., $\mathrm{SD} = \frac{\sum_{t=0}^{T}|h_{1n}(t)-h_{1(n-1)}(t)|^2}{\sum_{t=0}^{T}|h^2_{1(n-1)}(t)|} < \varepsilon\,(0.2<\varepsilon<0.3)$, where $h_{1n}(t)$ is the nth version of the signal $h_1(t)$ during the nth iteration of the sifting process. If the criterion is met, $h_{1n}(t)$ constitutes the first IMF $[IMF_1(t)]$ and the whole process is repeated using $r_1(t) = X(t) - IMF_1(t)$ as the original signal, till all IMFs are extracted. The superimposition of the N IMFs yields the original signal $X(t)$, i.e.,

$$X(t) = \sum_{k=1}^{N}\mathrm{IMF}_k(t) + r_N(t). \tag{13}$$

In (13), $r_N(t)$ denotes the residue of the EMD process and it is either a monotonic function or a constant. By computing the Hilbert transform of each IMF $\mathrm{IMF}_k^H(t)$ a conjugate pair is formed, so that an analytic signal can be written as

$$Z_k(t) = \mathrm{IMF}_k(t) + j\mathrm{IMF}_k^H(t) = A_k(t)e^{j\theta_k(t)} \tag{14}$$

where

$$A_k(t) = \sqrt{\mathrm{IMF}_k(t)^2 + \mathrm{IMF}_k^H(t)^2},$$
$$\theta_k(t) = \arctan\left(\frac{\mathrm{IMF}_k(t)}{\mathrm{IMF}_k^H(t)}\right). \tag{15}$$

In (15), $\theta_k(t)$ represents the instantaneous phase of $\mathrm{IMF}_k(t)$. Due to their generating process, IMFs constitute narrow band components and, therefore, a meaningful instantaneous frequency can be estimated by the derivative of $\theta_k(t)$

$$f_k(t) = \frac{1}{2\pi}\frac{d\theta_k}{dt}. \tag{16}$$

Based on the aforementioned, the original signal $X(t)$ can be expressed as

$$X(t) = \sum_{k=1}^{N}A_k(t)e^{j2\pi\int f_k(t)dt} \tag{17}$$

where $r_N(t)$ is left out due to its monotonic nature that denotes a longer trend [28]. Equation (17) enables us to represent the amplitude and the instantaneous frequency as functions of time in a 3-D plot, in which the amplitude can be contoured on the TF plane. In addition, the 3-D plot of the squared amplitude $A_k^2(t)$, for all IMFs, with respect to time and instantaneous frequency $f_k(t)$, forms the energy HHS-based TF representation used in this study.

Data analysis and subsequent classification was conducted using MATLAB R2009b (Mathworks, Inc., Natick, MA).

D. Classification

The classification approach adopted in this study was user independent, i.e., classification was performed on the complete dataset, created from all subjects' EEG responses. *FV*s were grouped with respect to two classes, namely "Like" and "Dislike." Class "Like" comprised of the *FV*s corresponding to the trials in which subjects rated their liking for the musical excerpts as 5 (like very much) or 4 (like). Consequently, class "Dislike" comprised of the *FV*s corresponding to the trials in which subjects rated their liking for the musical excerpts as 2 (do not like) or 1 (do not like at all). Four widely used classifiers were employed, i.e., quadratic discriminant analysis (QDA) [29], Mahalanobis [30], *k*-NN [31], and SVM [32]. For the *k*-NN classifier, the Euclidean distance was selected as the distance metric and the number of nearest neighbors was set to 4 ($= k$) after testing. Moreover, the Gaussian radial basis function kernel was used for the SVM classifier. The soft margin parameter and the scaling factor of the kernel were selected using cross-validation checking. Polynomial-based kernels were also considered at an initial stage; however, the polynomial mapping approach was abandoned, as the performance of the corresponding SVM classifiers was considerably lower compared to the Gaussian-based ones.

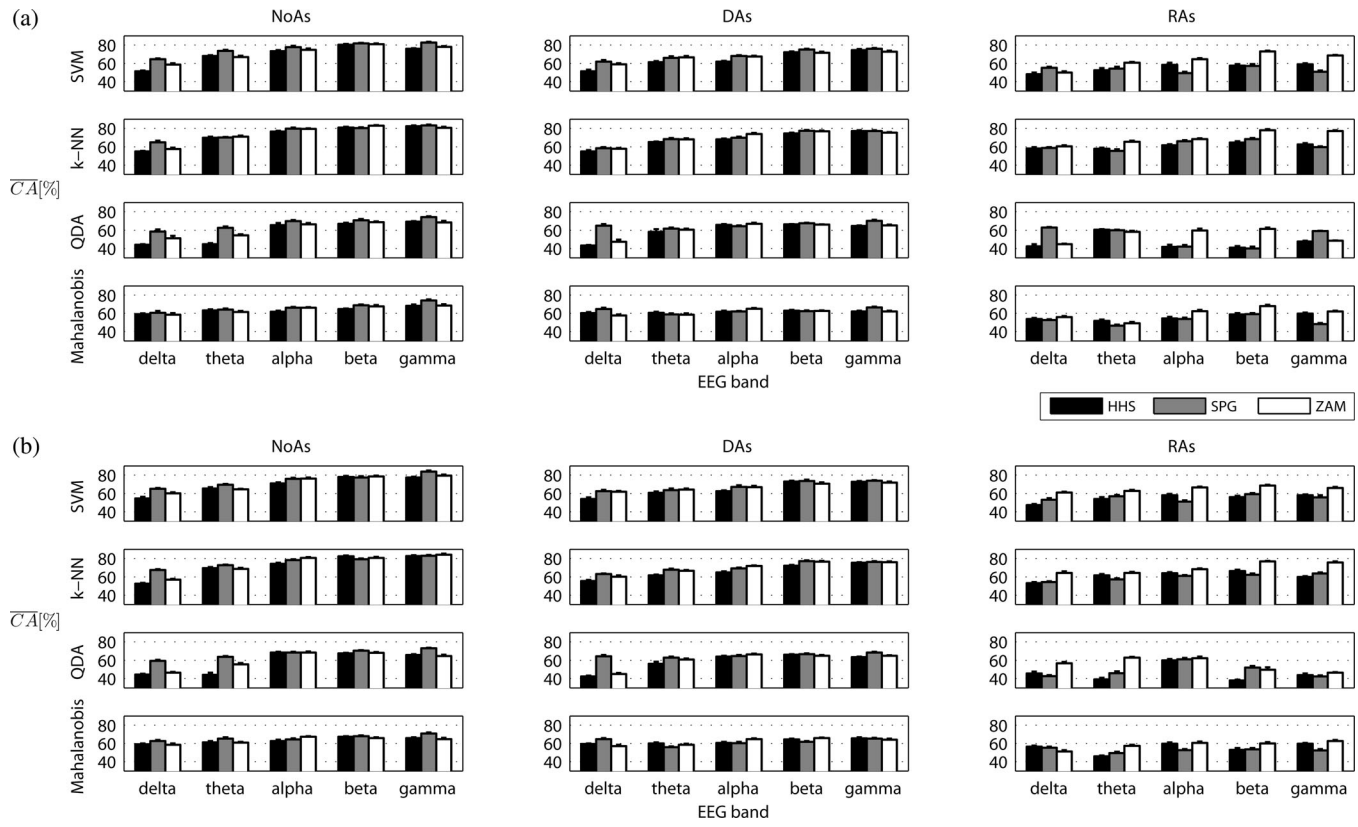


Fig. 4. Classification results in the case of the RP referencing approach with respect to each asymmetry type (NoAs, DAs, RAs), classifier (SVM, k-NN, QDA, Mahalanobis), EEG frequency band (*delta*, *theta*, *alpha*, *beta*, *gamma*), and TF technique (SPG, ZAM, HHS) for time windows (a) w_1 and (b) w_2. $\overline{CA}$s [%] are presented as bars along with the *SD* of *CA* (small scores above bars).

In order to use the entire dataset for training and testing the classifiers, a ten-fold cross-validation scheme was adopted [33]. During this scheme, the dataset is divided randomly in ten equal (or approximately equal) subsets and for each fold nine subsets are used for training and one subset for testing. The procedure is repeated ten times (ten-folds) in order for all subsets to be used as testing data. Here, each cross-validation scheme was furthermore repeated ten times, i.e., 100 repetitions of classification, in order to acquire more robust classification results. Classification performance was evaluated through the classification accuracy rate (*CA*), i.e., the percentage of correctly classified test *FV*s over the total number of test *FV*s. In particular, the average *CA* ($\overline{CA}$) and the *SD* of *CA* were estimated over ten runs of ten-fold cross validation for each set of *FV*s and each classifier.

Furthermore, classification under additive noise was also considered. The goal of the noise test was to compare the robustness of the three TF techniques (SPG, ZAM, HHS) with respect to the classification performance of the corresponding noisy features. In particular, zero-mean Gaussian noise was added to the EEG signals under five signal-to-noise ratios (SNRs), i.e., 15, 10, 5, and -3 dB. Subsequently, for each of the SNR cases, *FV*s that yielded the best classification rate under the noise-free condition, were recomputed using each of the three TF techniques and classified using the best classifier.

III. Results

A. Experimental Results

The dataset, subjected to the classification procedure, comprised of 219 instances of class "Like" and of 143 instances of class "Dislike." Trials corresponding to "undecided" rating were 178 and they were omitted from the feature extraction procedure and the subsequent classification. The imbalance in the dataset is reasonable and it is a direct consequence of the highly subjective nature of music preference rating. Nevertheless, in this study, the difference in the number of instances between the two classes was not considered significant enough to bias the classification results. Additionally, as it was expected, subjects rated their liking for the noise excerpt as 1 (do not like at all) in the 95.5% of cases and as 2 (do not like) in the remaining 4.5%.

Down to the outcomes of the classification procedure, Fig. 4 displays the classification results for the RP referencing approach for both time windows w_1 [Fig. 4(a)] and w_2 [Fig. 4(b)], with respect to each asymmetry type (NoAs, DAs, RAs), for all TF techniques and classifiers. $\overline{CA}$s are presented as bars, along with the *SD* of *CA* (small scores over bars) over ten runs of ten-fold cross validation. *CA*s in the text are reported in the format of $\overline{CA}$ ($\pm$*SD*)%. In the case of w_1, it can be seen that NoAs *FV*s led to higher *CA*s, as compared to the DAs and RAs *FV*s. Additionally, SVM and k-NN performed better

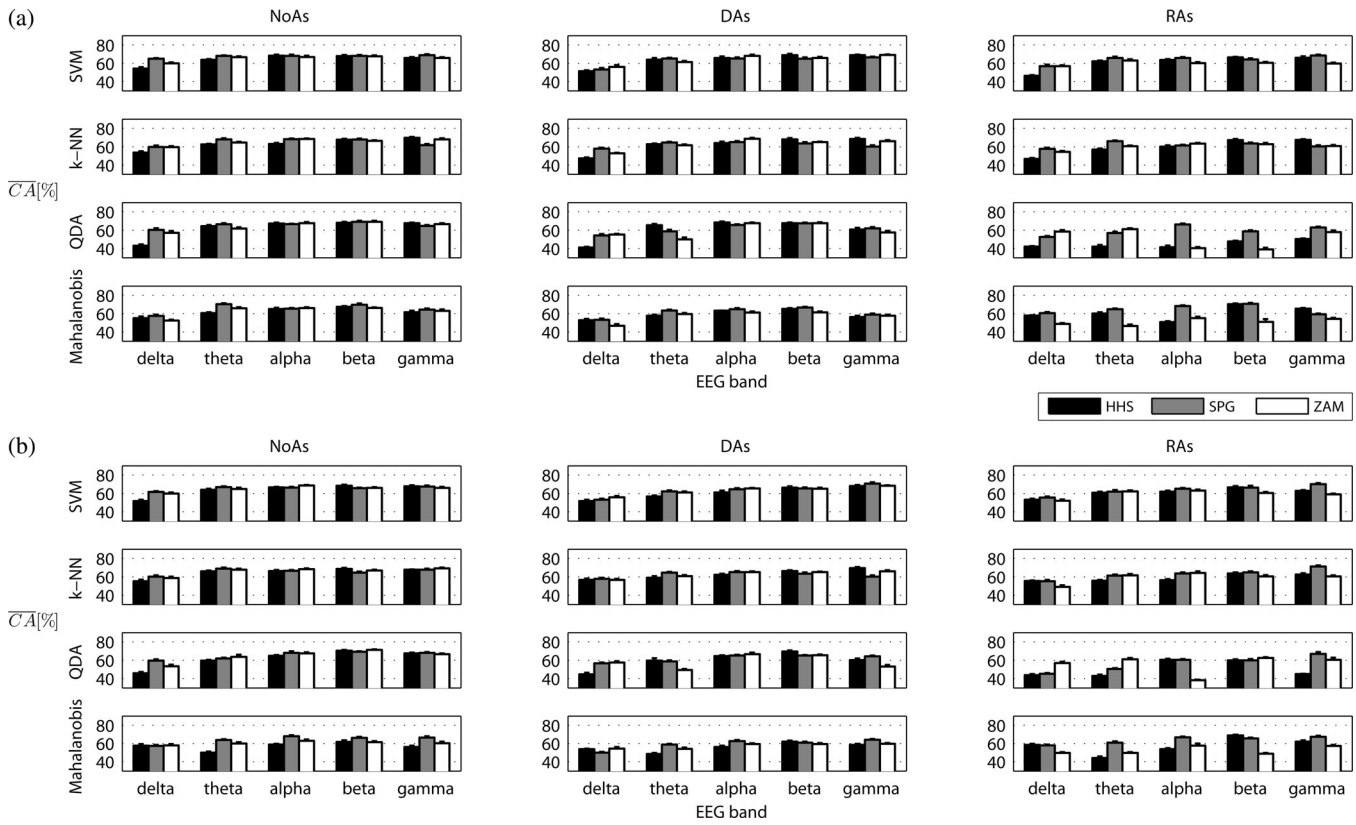


Fig. 5. Classification results in the case of the NL referencing approach with respect to each asymmetry type (NoAs, DAs, RAs), classifier (SVM, k-NN, QDA, Mahalanobis), EEG frequency band (*delta*, *theta*, *alpha*, *beta*, *gamma*), and TF technique (SPG, ZAM, HHS) for time windows (a) w_1 and (b) w_2. $\overline{CA}$s [%] are presented as bars along with the *SD* of *CA* (small scores above bars).

than discriminant analysis-based classifiers, i.e., QDA and Mahalanobis. Moreover, it can be deduced that, in general, *FV*s from *beta* and *gamma* bands led to higher *CA*s, followed by those from *alpha* and *theta* bands, while *FV*s from *delta* band led to the worst classification performance. In particular, best *CA* was achieved through the k-NN classifier by employing the NoAs *FV* type, estimated in *gamma* band using the SPG technique $\{83.15\ (\pm 0.93)\%\}$. Similar classification performance was achieved using the HHS $\{82.20\ (\pm 0\ .87)\%\}$ and ZAM $\{80.66\ (\pm\ 0.96)\%\}$ techniques. NoAs *FV*s derived from *beta* band and classified by k-NN also led to high *CA*s {HHS: 80.85 ($\pm$ 0.93)%; SPG: 80.28 ($\pm$0.92)%; ZAM: 82.85 ($\pm$0.73)%}. In the case of time window w_2, classification results were similar to those obtained in the case of w_1. Best *CA* was achieved through the k-NN classifier by employing the NoAs *FV* type, estimated in *gamma* band using the ZAM technique $\{84.09\ (\pm 1.08)\%\}$. Similar, but slightly lower, $\overline{CA}$s were achieved for the same EEG band, using SPG $\{83.76\ (\pm 1.16)\%\}$ and HHS $\{82.57\ (\pm 0.75)\%\}$ techniques. As far as other frequency bands are concerned, *beta* band yielded the second best *CA* {HHS: 82.34 ($\pm$0.73)%; ZAM: 80.50 ($\pm$1.12)%; SPG: 79.12 ($\pm$0.97)%} using also the NoAs-based *FV*s and k-NN.

Probing further, Fig. 5 shows the classification results for the NL referencing approach in the same format as in Fig. 4. In the case of w_1, best *CA* was achieved using the RAs *FV*s from the *beta* band, estimated through SPG and classified by Mahalanobis distance-based classifier $\{70.69\ (\pm 0.89)\%\}$. For the w_2 time window, best classification performance was achieved using the NoAs *FV*s from *gamma* band, estimated through ZAM, and classified by QDA $\{71.27\ (\pm 0.55)\%\}$. In general, $\overline{CA}$ values were lower than 72% for all types of *FV*s and classifiers, for both w_1 and w_2. In Fig. 6 the classification results for the NoR referencing approach are displayed in the same format as in Figs. 4 and 5. In the case of w_1, best $\overline{CA}$ was achieved using the NoAs *FV*s from *beta* band, estimated through SPG and classified by Mahalanobis distance-based classifier $\{70.80\ (\pm 0.69)\%\}$, while in the case of w_2, best classification performance was achieved using the NoAs *FV*s from *beta* band, estimated through ZAM, and classified using QDA $\{71.88\ (\pm 0.73)\%\}$. As in the case of the NL referencing approach, $\overline{CA}$ values were lower than 72% for all types of *FV*s and classifiers, for both w_1 and w_2. From the aforementioned, it can be derived that the RP referencing approach, and particularly the NoAs *FV* types, led to better classification results, as compared to the other two approaches (NL and NoR).

So far, best classification results were achieved following the RP referencing approach and the NoAs *FV* type. As far as individual EEG bands are concerned, *FV*s from the *beta* and *gamma* band yielded the highest *CA*s ($> 80\%$). Based on this evidence, and in order to explore the combined classification

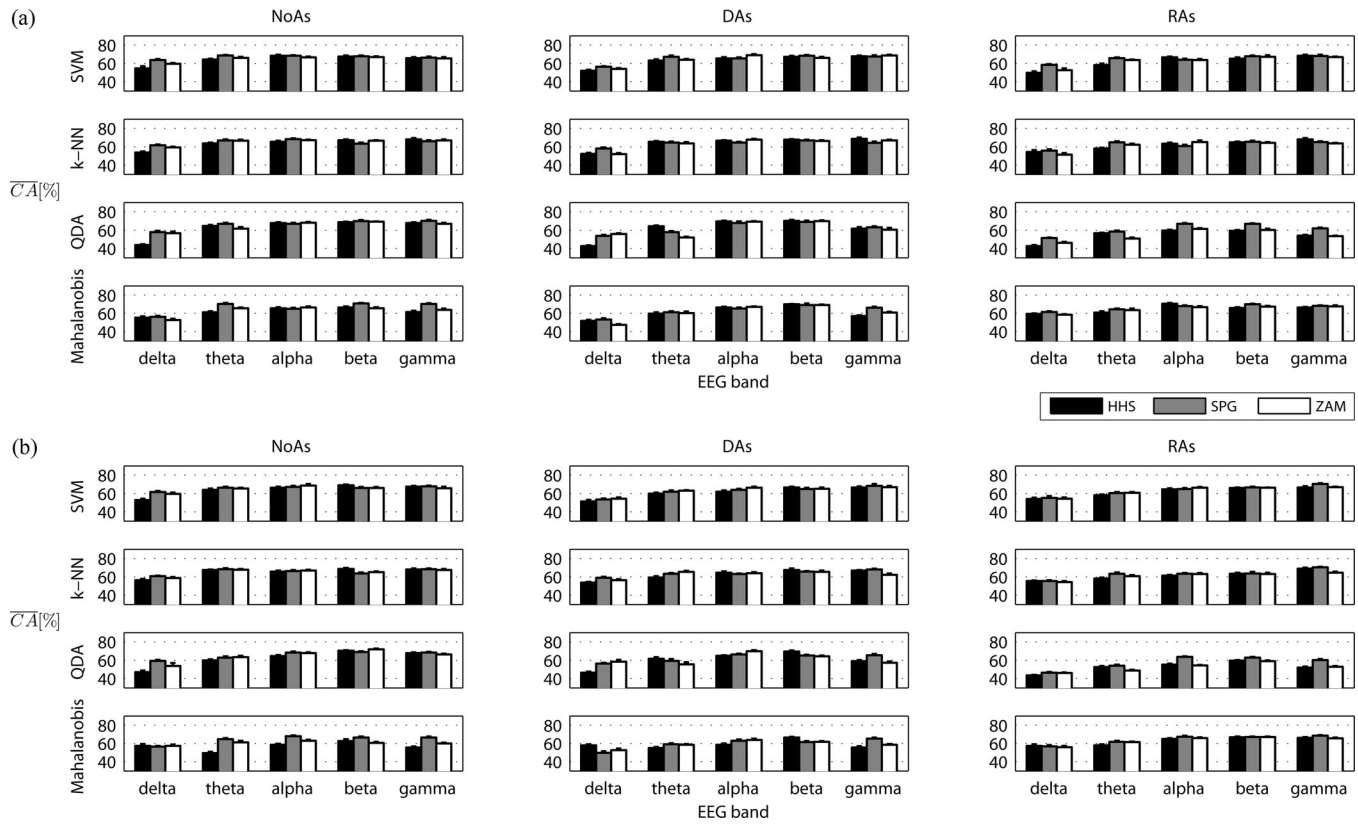


Fig. 6. Classification results in the case of the NoR referencing approach with respect to each asymmetry type (NoAs, DAs, RAs), classifier (SVM, k-NN, QDA, Mahalanobis), EEG frequency band (*delta*, *theta*, *alpha*, *beta*, *gamma*), and TF technique (SPG, ZAM, HHS) for time windows (a) w_1 and (b) w_2. $\overline{CA}$s [%] are presented as bars along with the *SD* of *CA* (small scores above bars).

performance of these two bands, *FV*s, constructed as the concatenation of the RP-NoAs *FV*s corresponding to *beta* and *gamma* band, were subjected to classification. For the sake of comparison, additional *FV*s were constructed as the concatenation of the RP-NoAs *FV*s derived from all EEG bands and tested in terms of classification performance. In the case of the *beta* and *gamma* band combination, best *CA* was achieved using the HHS technique in time window w_1 and the k-NN classifier $\{86.52\ (\pm 0.76)\%\}$. The corresponding rates using the ZAM and SPG technique were 85.41 ($\pm$0.32)% and 84.78 ($\pm$0.59)%, respectively. With regard to the *FV*s derived from all EEG bands, best classification was achieved also in w_1 using the SPG technique and k-NN classifier $\{83.43\ (\pm 1.03)\%\}$. Corresponding *CA* rates achieved using the HHS and ZAM techniques were 75.48 ($\pm$0.93)% and 81.49 ($\pm$0.92)%, respectively. Other combinations of features from different EEG bands were also considered, however, the classification performance of the latter was similar to the *CA* acquired using the combination of features from all bands ($\sim$ 83%). For an overall view, Table I displays the best classification results acquired from each EEG frequency band, along with the TF techniques and time windows with which they were achieved, in the case of RP-NoAs setup and k-NN classifier.

In the case of RP-NoAs and in order to verify the superiority of the *beta/gamma*-HHS-w_1 setup, a series of paired-samples t-tests were conducted on the *CA*s produced from each run and each fold of the cross-validation scheme, i.e., 100 values of *CA* for each case. The comparison was done between the best performing setup and the setups that also led to a satisfactory *CA* ($> 80\%$). The *beta/gamma*-HHS-w_1 setup statistically outperformed its counterparts in terms of TF technique, i.e., the *beta/gamma*-ZAM-w_1 ($t(99) = 2.559$, $p < 0.05$) and the *beta/gamma*-SPG-w_1 ($t(99) = 5.260$, $p = 0.001$) setups. Moreover, the *beta/gamma*-HHS-w_1 approach yielded significantly higher *CA*s as compared to the *beta*-ZAM-w_1 ($t(99) = 13.561$, $p < 0.001$), the *gamma*-ZAM-w_2 ($t(99) = 2.548$, $p < 0.05$), and

TABLE I
BEST $\overline{CA}$ ($\pm$*SD*)% RATES FOR EACH EEG FREQUENCY BAND (AND COMBINATIONS) FOR THE RP-NOAS SETUP AND k-NN CLASSIFIER, ALONG WITH THE CORRESPONDING TF TECHNIQUES AND TIME WINDOWS

Band	$\overline{CA}$ ($\pm SD$) %	TF technique	Time window
delta	67.29 (±1.01)	SPG	w_2
theta	73.65 (±0.99)	SPG	w_1
alpha	79.59 (±1.13)	SPG	w_1
beta	82.85 (±0.73)	ZAM	w_1
gamma	84.09 (±1.08)	ZAM	w_2
beta/ gamma	86.52 (±0.76)	HHS	w_1
all bands	83.43 (±1.03)	SPG	w_1

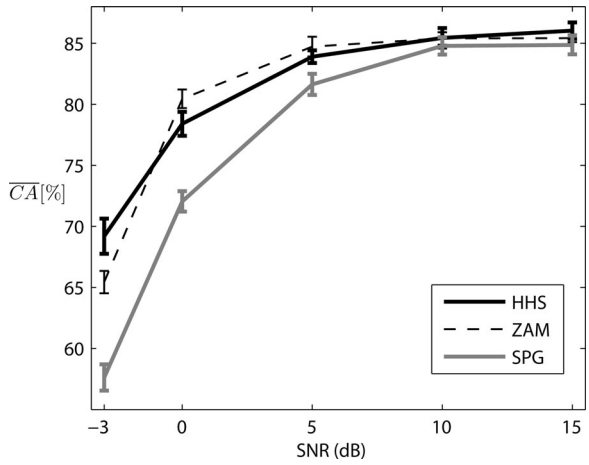


Fig. 7. Average classifications accuracies ($\overline{CA}$s [%]) under additive noise corruption (SNRs: 15, 10, 5, 0, -3 dB), using the k-NN classifier and the concatenated *beta* and *gamma* band RP-NoAs *FV*s, estimated using each of the three TF techniques (SPG, ZAM, HHS) in time window w_1. Vertical lines represent the *SD* of *CA*.

the all bands-SPG-w_1 ($t(99) = 4.627$, $p = 0.001$) setups, that are reported in Table I.

B. Noise Test Results

Fig. 7 shows the results of the classification procedure under additive noise. Classification was performed using the k-NN classifier on the best performing *FV* type, i.e., the concatenation of RP-NoAs *FV*s derived from the *beta* and *gamma* band, in time window w_1. As it was expected, classification performance deteriorates as the SNR decreases. However, the performance of the HHS and ZAM-based *FV*s is considerably better as compared to the SPG-based *FV*s, under 5-dB SNR or lower. The performance of the HHS-based *FV* set is comparable to the ZAM-based set except for the case of severe noise (SNR $= -3$ dB), where HHS-based *FV*s performed considerably better.

IV. Discussion

In this study, several *FV* types and classifiers were evaluated, in order to discriminate between EEG responses to liked and disliked music. *FV*s were constructed using three TF techniques by taking into consideration four parameters: RS, time window, asymmetry type, and EEG frequency band. Best classification $\{86.52\ (\pm 0.76)\%\}$ was achieved through the k-NN classifier by employing the fusion of *FV*s derived from *beta* and *gamma* bands and estimated in time window w_1, using the RP-NoAs approach and HHS TF analysis. Implications that arise from the acquired results are discussed later, in the light of both physiological and technical parameters.

A. Reference Approach and Time Intervals

In general, *FV*s estimated using the RP referencing approach led to better classification performance as compared to those estimated by the NL and NoR ones. The latter might have failed to yield satisfactory results due to the fact that ERS/ERD phenomena, in order to be meaningful, mandate that the EEG spectral activity during stimulation must be computed as a relative quantity to the spectral activity during a reference period [22]. Such conception reflects a transition from a stimulus-free period to the period of stimulation. The reference period should precede some seconds before the stimulus onset and be fairly stable in terms of EEG characteristics across trials and subjects. The latter apply for the RP referencing approach, as such stability is more likely to be presented during a stimulus-free RP and, moreover, for each trial, the 2-s-long resting interval, which was used in feature estimation, preceded the listening of the musical excerpt. On the other hand, these conditions do not apply in the case of the NL referencing state, as noise excerpts were presented pseudorandomly, and not in pair with musical excerpts. In addition, although, NL was expected to evoke a common and unpleasant response among listeners, the sporadic repetition of the noise excerpts in the course of the experimental procedure, might have reduced the attention of certain subjects to them, thus, not leading to a stable EEG response.

As far as the time window selection is concerned, better classification results were generally acquired from *FV*s estimated in the total duration of ML (0–15 s), i.e., time window w_1. Although, the selection of the narrower time window w_2 led to comparable classification results, brain responses earlier than 8 s, corresponding to appraisal judgments, might have occurred; hence, slightly lower performance of *FV*s estimated in this time window (8–15 s) is produced. Such judgments are highly influenced by the familiarity with each musical excerpt and thus, their accompanying electrophysiological correlates might differ in onset times and even duration. In [23], it was found that the average time for a listener to start making emotional judgments for musical pieces is 8.31 s. However, this latency might not be directly associated with the time that a subject needs to make appraisal judgments. In this study, time course issues that relate to familiarity or specific musical genre preference were not investigated and are considered as a subject of future research.

B. TF Analysis and Classification Perspective

Down to the TF analysis, classification performance achieved with each of the three techniques was comparable, with HHS yielding the best *CA* rate, followed by ZAM. The marginal preeminence of HHS may be attributed to its adaptive decomposition of each signal that addresses the nonstationary and nonlinear nature of processes like the EEG and to the high-frequency resolution derived from the instantaneous frequency estimation [28]. EEG signals constitute electric oscillatory patterns that are formed from the superimposition of common and invariant frequency modes, i.e., *delta*, *theta*, *alpha*, *beta,* and *gamma* [34]. Fluctuations in the latter discrete modes are indicative of brain responses to stimuli, thus, their efficient detection is crucial in order to profoundly describe the related electrophysiological phenomena. In general, HHS is an intuitive method that exploits the endogenous structural characteristics of each signal, in order to decompose it, instead of employing an arbitrary set of basis functions like in the cases of SPG and ZAM. Intuitively, EEG-based intrinsic modes may carry information about EEG processes that expand, in terms of frequency content, in more

than one of the established frequency bands. In this study, activity in the *beta* and *gamma* bands was mainly included in the first two extracted *IMF*s that represent signal changes of higher frequencies. On the other hand, ZAM TF analysis reduces cross terms between signal components through its cone-shaped kernel function [26], but its implementation requires vast computational time, especially for high TF resolutions. Additionally, although, SPG is widely used in the literature and its implementation is time efficient, it also constitutes a linear technique that is halted by TF resolution constraints. Nevertheless, the feature extraction approach, adopted in this study, did not depend significantly on time resolution as the time windows, in which energy-based features were estimated, were relatively large and this fact might have contributed to the comparable performance of the three TF techniques. An eminent advantage of the HHS and ZAM techniques over the SPG is revealed through the noise test, as the two former methods performed better under noise corruption, while SPG-based classification performance considerably deteriorated under severe noise (5-dB SNR or lower) (Fig. 7).

Finally, as far as the different classifiers are concerned, k-NN and SVM yielded higher *CA* rates, as compared to the discriminant analysis-based classifiers, i.e., QDA and Mahalanobis, with k-NN performing marginally better than SVM. The superiority of k-NN and SVM methods over the discriminant analysis-based ones might lie in the ability of the former classifiers to discriminate more efficiently between groups of *FV*s of high dimensionality through the sophisticated formation of separating hyperplanes (SVM) [32] or the estimation of multidimensional distance metrics which do not rely on prior knowledge of the data probability density (k-NN) [31]. As far as the comparison between those two is concerned, a possible explanation of the marginal preeminence of the k-NN classifier over SVM might lie in the scaling of *FV*s, however, an elaborate investigation of the precise reasons is considerably arduous and beyond the scopes of this study. In addition, it must be noted that k-NN classification required substantially lower computational time as compared to SVM.

C. Asymmetry and Familiarity

Down to the asymmetry type, *FV*s representing an overall brain activation led to better classification results than asymmetry-based *FV*s (DAs and RAs). Hemispheric lateralization has been proposed as an effective index for the discrimination between discrete emotional states, and especially between positive and negative emotions [15]. Such emotion-related activity has been previously reported in music-based studies [10], [11], [13], but it was also found to be absent during the listening of pleasant (unpleasant) consonant (dissonant) music [12]. In order to define the cause of this discrepancy, methodological parameters, such as the EEG reference montage [35], [36], are under debate, but there are no solid arguments so far [12], [35]. It must be noted that, in this study, two alternative reference approaches, i.e., the common average EEG reference and the reference of EEG recordings to infinity [37], were also considered. The latter method was implemented as proposed in [38]. However, the corresponding classification results were lower. The NoAs setup constantly led to better classification performance as compared to the DAs and RAs setups, nonetheless, *CA*s were lower by 6.43% on average in the case of the common average and by 4.92% in the case of infinity reference. Although, further scrutiny of this matter is beyond the scopes of this study, the fact that asymmetry-based *FV*s did not lead to the best classification performance, might imply that a direct mapping of positive/negative appraisal judgments to music-induced positive/negative-valenced emotions is not always plausible. The latter implication complies with evidence suggesting that, occasionally, negatively valenced music tends to be liked by listeners. Indeed, in a recent study, negative correlations between liking for music video clips and the valence/arousal of induced emotions reveal that subjects may more than often display a preference for "sad" audiovisual stimuli [39].

In this study, it must be noted that the role of familiarity was not scrutinized. According to the literature, the level of familiarity can affect both the time course of affective responses [23], as well as the intensity of such responses [40]. However, the level of familiarity does not prevent listeners to manifest an *impulsive* appraisal response (positive or negative) over a musical piece that is either familiar or unfamiliar to them. During the experimental procedure of this study, in the cases when the listeners remained undecided on whether they liked the musical excerpt or not, the rating form that was presented to them after the ending of each excerpt included the choice of "undecided" to choose. EEG trials that corresponded to such ratings were excluded from further analysis. Furthermore, feature estimation involved frequency- and time averaging over frequency bands and relatively long time intervals, which leads to an integration of the subjects' electrophysiological responses both in frequency and time, minimizing, in this way, the effect of familiarity. For example, the effect of any phasic responses (subsecond to second [41] in duration) due to high degree of familiarity is expected to be less influential to the extracted features, which mainly represent the tonic EEG responses (several seconds in duration). Nevertheless, a more precise time detection of appraisal responses is expected to increase the classification accuracy, and for this reason, this matter is under current investigation by the authors.

D. Neurophysiological Correlates

Probing further, during ML, brain responses are formed by the combined presence of processes that may relate to music perception, memory retrieval, and affection. However, as it was mentioned before, music appraisal is mainly interpreted in the context of affective responses; thus, in this study, electrophysiological evidence that led to the discrimination of liking/disliking judgments is considered to be more likely related to aspects of such responses. Best classification was achieved through the activity of *beta* and *gamma* band, which have been both suggested to reflect emotional arousal phenomena [17], [19], [20]. Therefore, arousal could be linked to music preference, as appraisal judgments might be seen as an expression of excitation (or lack of it) over a liked (or disliked) musical piece or as a

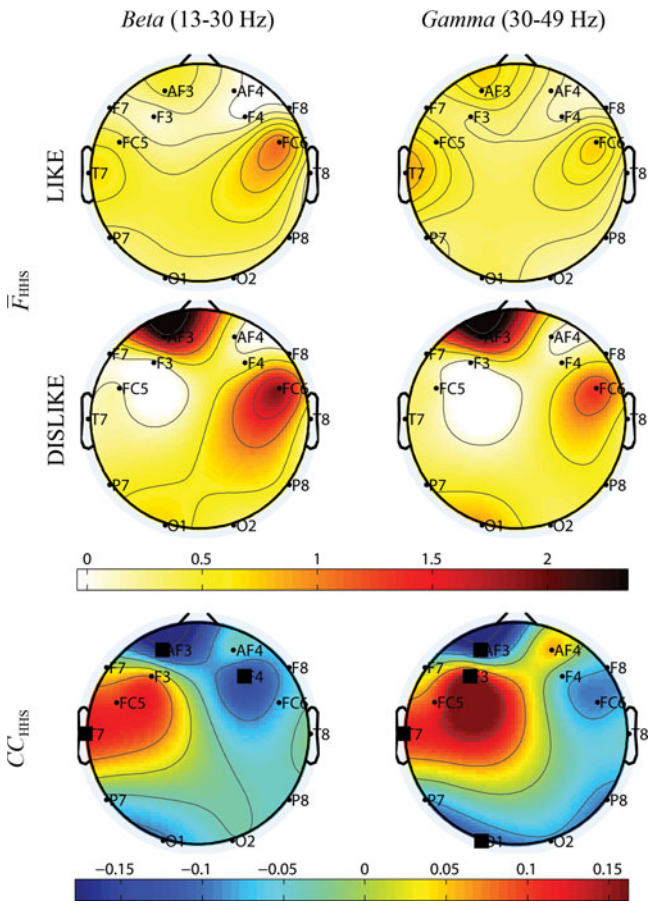


Fig. 8. Topographical maps of the average features ($\bar{F}_{\mathrm{HHS}}$) across trials, computed using the RP-HHS-w_1 approach in the *beta* and *gamma* EEG bands for each class (Like and Dislike) (two top panels), and corresponding brain maps of the correlations (CC_{HSS}) between features of all trials with appraisal ratings (lower panel); squares indicate the cases of statistically significant correlation ($p < 0.05$).

modulating factor of the intensity of the listener's emotional state, regardless of its valence. The role of arousal is corroborated by a recent study [42] in which both subjective ratings and physiological measures revealed that, not only arousal constitutes a demonstration of music appraisal, but also it could determine music preference, although, a direct association of self-reported arousal with emotional arousal as displayed through physiological signals was not established. Evidence concerning the association of different EEG bands to the formation of music appraisal judgments is scarce. In a study investigating the recognition of affective states evoked by music video clips, activity in *theta* and *alpha* bands was correlated with the valence dimension and liking/disliking ratings [43]. There, however, a maximum accuracy of 49.4% was achieved for the EEG-based discrimination between liking and disliking judgments over the video clips. In a more recent study using a similar experimental paradigm, activity in the *beta* band, as displayed through power spectral density, was associated with both arousal and liking, especially in the left temporal region, while *gamma* band was associated with high valence/arousal and low liking ratings [39].

In this study, by taking into consideration the best performing *FV*s and in order to illustrate such correlates, the average features ($\bar{F}_{\mathrm{HHS}}$), across trials, of the *beta* and *gamma* EEG bands, as computed by the RP-HHS-w_1 approach, are topographically plotted with respect to each electrode site in Fig. 8 (two top panels). Additionally, the correlation coefficients (CC_{HHS}) between appraisal ratings ("Like": 1, "Dislike": 0) and features of all trials of each electrode site are also topographically depicted in Fig. 8 (lower panel). From Fig. 8 (two top panels: case of "Like" versus "Dislike"), an increase of the average relative energy during the period of ML as compared to the RP is observed: 1) in left prefrontal areas (AF3) for both frequency bands, 2) in right fronto-central areas (F4, FC6—*beta*), and 3) in left occipital sites (O7— *gamma*). Moreover, a decrease in the average relative energy is observed in right frontal (F3—*gamma*) and temporal areas (T7— *beta* and *gamma*). Consequently, activity in the *gamma* band is negatively correlated with liking in right prefrontal and occipital areas and positively correlated with right frontal and temporal regions. Relative energy in the *beta* band is negatively correlated with liking in left prefrontal and right frontal areas, while it is positively correlated with left temporal-auditory regions. Neuroanatomically, frontal lobe is associated with emotional responses [16], while temporal lobe is linked to memory and auditory processing tasks [39]; hence, it seems that, during music preference phenomena, brain areas with different functionalities are recruited, possibly, in a combinatory way.

V. Conclusion and Future Work

This study proposed a methodological scheme for the EEG-based recognition of music preference. Bilateral average activity from *beta* and *gamma* band, when referred to the RP, led to the best discrimination between liking and disliking judgments. This evidence may point to a connection of music preference to emotional arousal phenomena. The performance of the three examined TF techniques was comparable, with a slight preeminence of the HHS that was further revealed through the robustness of this technique to noise corruption. So far, no related work that targets specifically to the EEG-based classification of music appraisal judgments is reported in the literature, in order for the acquired results to be compared. However, in this study, a large variety of *FV* types accounting for both physiological (RS, time window, and hemispheric asymmetry) and technical (TF techniques, classifiers) parameters were evaluated. Consequently, the best *CA* rate achieved (>85%) is considered satisfactory enough. Additionally, as the results arise from a user-independent logic, the proposed methodology can be transferred efficaciously to a generalized BCI system, avoiding, in this way, time-consuming user adaptation and training phases.

Future work toward a more efficient detection may focus on the latency of related brain responses. The systematical investigation of the time course of affective phenomena could lead to a system that would take into account additional parameters that influence music preference, such as familiarity. Additional investigation pertaining to feature selection and electrode reduction could improve the classification performance, while reducing computational time.

ACKNOWLEDGMENT

The authors would like to express their gratitude to the nine participants for their voluntary engagement in the experimental procedure. They would also like to thank Dr. V. Kimiskidis, Assistant Professor of Neurology, for his valuable help on the final shaping of this study.

REFERENCES

[1] E. Brattico, P. Brattico, and T. Jacobsen, "The origins of aesthetic enjoyment of music – A review of the literature," *Musicae Scientiae*, vol. 13, pp. 15–39, 2009.

[2] O. Ladinig and E. G. Schellenberg, "Liking for unfamiliar music: Effects of felt emotion and individual differences," *Psychol. Aesthetics, Creat., Arts*, vol. 6, no. 2, pp. 146–154, 2012.

[3] V. Konečni, "The aesthetic trinity: Awe, being moved, thrills," *Bull. Psychol. Arts*, vol. 5, no. 2, pp. 27–44, 2005.

[4] E. Schubert, "The influence of emotion, locus of emotion and familiarity upon preference in music," *Psychol. Music*, vol. 35, no. 3, pp. 499–515, 2007.

[5] P. Evans and E. Schubert, "Relationships between expressed and felt emotions in music," *Musicae Scientiae*, vol. 12, no. 1, pp. 75–99, 2008.

[6] J. A. Russell, "A circumflex model of affect," *J. Personality Soc. Psychol.*, vol. 39, pp. 1161–1178, 1980.

[7] A. J. Blood and R. J. Zattore, "Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion," *Proc. Natl. Acad. Sci. USA*, vol. 98, no. 20, pp. 11818–11823, 2001.

[8] 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.

[9] I. Peretz, "Towards a neurobiology of musical emotions," in *Handbook of Music and Emotion: Theory, Research, Applications*, P. N. Juslin and J. A. Sloboda, Eds. Oxford, U.K.: Oxford Univ. Press, 2010, pp. 99–126.

[10] E. Altenmüller, 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 lateralization patterns," *Neuropsychologia*, vol. 40, pp. 2242–2256, 2002.

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

[12] 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.

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

[14] P. G. Hunter and E. Glenn Schellenberg, "Music and emotion," in *Music Perception*, vol. 36, M. Riess Jones, R. R. Fay, and A. N. Popper, Eds. 1st Ed, Berlin, Germany: Springer, 2010, pp. 169–192.

[15] R. J. Davidson, "What does the prefrontal cortex "do" in affect: Perspectives on frontal EEG asymmetry research," *Biol. Psychol.*, vol. 67, pp. 219–233, 2004.

[16] 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.

[17] L. I. Aftanas, N. V. Reva, L. N. Savotina, and V. P. Makhnev, "Neurophysiological correlates of induced discrete emotions in humans: An individually oriented analysis," *Neurosci. Behav. Psychol.*, vol. 36, no. 2, pp. 119–130, 2006.

[18] D. J. L. G. 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, 2001.

[19] A. Keil, M. M. Müller, T. Gruber, C. Wienbruch, M. Stolarova, and T. Elbert, "Effects of emotional arousal in the cerebral hemispheres: A study of oscillatory brain activity and event-related potentials," *Clin. Neurophysiol.*, vol. 112, no. 11, pp. 2057–2068, 2001.

[20] 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, 2008.

[21] P. Ekman, "An argument for basic emotions," *Cogn. Emotion*, vol. 6, no. 3–4, pp. 169–200, 1992.

[22] G. Pfurtscheller and F. H. Lopes da Silva, "Event-related EEG/MEG synchronization and desynchronization: Basic principles," *Clin. Neurophysiol.*, vol. 110, pp. 1842–1857, 1999.

[23] J. P. 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, no. 4, pp. 355–364, 2009.

[24] B. Graimann, J. E. Huggins, S. P. Levine, and G. Pfurtscheller, "Visualization of significant ERD/ERS patterns in multichannel EEG and ECoG data," *Clin. Neurophysiol.*, vol. 113, no. 1, pp. 43–47, 2002.

[25] D. P. Allen and C. D. MacKinnon, "Time-frequency analysis of movement-related spectral power in EEG during repetitive movements: A comparison of methods," *J. Neurosci. Methods*, vol. 186, no. 1, pp. 107–115, 2010.

[26] Y. Zhao, L. E. Atlas, and R. J. Marks, "The use of cone-shaped kernels for generalized time-frequency representations of nonstationary signals," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 7, pp. 1084–1091, Jul. 1990.

[27] B. Boashash, "Theory of quadratic TFDs," in *Time Frequency Signal Analysis and Processing: A Comprehensive Reference*, B. Boashash, Ed. Oxford, U.K.: Elsevier, 2003, pp. 59–82.

[28] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. Liu, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proc. Roy. Soc. London Series A: Math., Phys. Eng. Sci.*, vol. 454, no. 1971, pp. 903–995, 1998.

[29] G. J. McLachlan, *Discriminant Analysis and Statistical Pattern Recognition*. New York: Wiley, 1992.

[30] P. C. Mahalanobis, "On the generalized distance in statistics," *Proc. Natl. Inst. Sci. USA*, vol. 2, no. 1, pp. 49–55, 1936.

[31] T. M. Cover and P. E. Hart, "Nearest neighbor pattern classification," *IEEE Trans. Inf. Theory*, vol. IT-13, no. 1, pp. 21–27, Jan. 1967.

[32] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge, U.K.: Cambridge Univ. Press, 2000.

[33] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Proc. 14th Int. Conf. Artif. Intell.*, Montreal, Canada, 1995, pp. 1137–1143.

[34] E. Baskar, C. Basar-Eroglu, S. Karakas, and M. Schürmann, "Oscillatory brain theory: A new trend in neuroscience," *IEEE Eng. Med. Biol.*, vol. 18, no. 3, pp. 56–66, May/Jun. 1999.

[35] D. Hagemann, "Individual differences in anterior EEG asymmetry: Methodological problems and solutions," *Biol. Psychol.*, vol. 67, no. 1–2, pp. 157–182, 2004.

[36] D. Yao, L. Wang, R. Oostenveld, K. D. Nielsen, L. Arendt-Nielsen, and A. C. Chen, "A comparative study of different references for EEG spectral mapping: The issue of the neutral reference and the use of the infinity reference," *Physiol. Meas.*, vol. 26, no. 3, pp. 173–185, 2005.

[37] D. Yao, "A method to standardize a reference of scalp EEG recordings to a point at infinity," *Physiol. Meas.*, vol. 22, pp. 693–711, 2001.

[38] R. A. Thuraisingham, "Analytical expressions for the transfer matrix to standardize scalp potentials to infinity reference," *J. Comput. Neurosci.*, vol. 31, pp. 609–613, 2011.

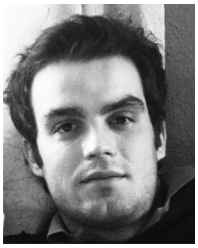
[39] E. Kroupi, A. Yazdani, and T. Ebrahimi, "EEG correlates of different emotional states elicited during watching music videos," in *Proc. 4th Int. Conf. Affective Comput. Intell. Interact.*, Memphis, TN, 2011, pp. 457–466.

[40] S. O. Ali and Z. F. Peynircioğlu, "Intensity of emotions conveyed and elicited by familiar and unfamiliar music," *Music Perception*, vol. 27, no. 3, pp. 177–182, 2010.

[41] R.-S. Huang, T.-P. Jung, A. Delorme, and S. Makeig, "Tonic and phasic electroencephalographic dynamics during continuous compensatory tracking," *Neuroimage*, vol. 39, no. 4, pp. 1896–1909, 2008.

[42] T. Schäfer and P. Sedlmeier, "Does the body move the soul? The impact of arousal on music preference," *Music Perception*, vol. 29, no. 1, pp. 37–50, 2011.

[43] S. Koelstra, A. Yazdani, M. Soleymani, C. Muhl, J.-S. Lee, A. Nijholt, T. Pun, T. Ebrahimi, and I. Patras, "Single trial classification of EEG and peripheral signals for recognition of emotions induced by music videos," in *Proc. Int. Conf. Brain Inf.*, Toronto, Canada, 2010, pp. 89–100.



Stelios K. Hadjidimitriou (S'10) was born in Karditsa, Greece, in 1984. He received the Dipl. degree in electrical and computer engineering from the Aristotle University of Thessaloniki, Thessaloniki, Greece, in 2008. He is currently working toward the Ph.D. degree in the Department of Electrical and Computer Engineering, Aristotle University of Thessaloniki, affiliated with the Signal Processing and Biomedical Technology Unit of the Telecommunications Laboratory.

During the summers of 2006 and 2007, he was with the Public Power Corporation S.A., Greece. His current research interests include advanced signal processing, biomedical engineering, and neural processes concerning music perception and cognition.

Mr. Hadjidimitriou is a member of the Technical Chamber of Greece.

Leontios J. Hadjileontiadis (S'87–M'98–SM'11) was born in Kastoria, Greece, in 1966. He received the Dipl. degree in electrical engineering in 1989 and the Ph.D. degree in electrical and computer engineering in 1997, both from the Aristotle University of Thessaloniki (AUTH), Thessaloniki, Greece. He also received the Diploma in Musicology from AUTH, in 2011, and the Ph.D. degree in music composition from the University of York, York, U.K., in 2004.

Since December 1999, he joined the Department of Electrical and Computer Engineering, AUTH, Greece, as a Faculty Member, where he is currently an Associate Professor, working on lung sounds, heart sounds, bowel sounds, ECG data compression, seismic data analysis, and crack detection in the Signal Processing and Biomedical Technology Unit of the Telecommunications Laboratory. He is also a Professor of Composition at the State Conservatory of Thessaloniki, Greece. His research interests include higher order statistics, alpha-stable distributions, higher order zero crossings, wavelets, polyspectra, fractals, neurofuzzy modeling for medical, mobile and digital signal processing applications.

Dr. Hadjileontiadis is a member of the Technical Chamber of Greece, the Higher Order Statistics Society, the International Lung Sounds Association, and the American College of Chest Physicians. He received the second award at the Best Paper Competition of the 9th Panhellenic Medical Conference on Thorax Diseases, Thessaloniki, in 1997. He was also an open finalist at the Student paper Competition (Whitaker Foundation) of the IEEE EMBS'97, Chicago, IL, a finalist at the Student Paper Competition (in memory of Dick Poortvliet) of the MEDICON'98, Lemesos, Cyprus, and the Young Scientist Award of the 24th International Lung Sounds Conference, Marburg, Germany, in 1999. From 2004 till present, he organized and served as a mentor to five-student teams that have ranked as third, second, and seventh worldwide, respectively, at the Imagine Cup Competition (Microsoft), Sao Paulo, Brazil (2004)/Yokohama, Japan (2005)/Seoul, Korea (2007)/New York, (2011)/Sydney, Australia (2012), with projects involving technology-based solutions for people with disabilities and pain management.