This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Emotion Recognition of Playing Musicians From EEG, ECG, and Acoustic Signals

Luca Turchet, *Senior Member, IEEE*, Barry O'Sullivan, Rupert Ortner, and Christoph Guger, *Member, IEEE*

***Abstract*—This article investigated the automatic recognition of felt and musically communicated emotions using electroencephalogram (EEG), electrocardiogram (ECG), and acoustic signals, which were recorded from eleven musicians instructed to perform music in order to communicate happiness, sadness, relaxation, and anger. Musicians' self-reports indicated that the emotions they musically expressed were highly consistent with those they actually felt. Results showed that the best classification performances, in a subject-dependent classification using a KNN classifier were achieved by using features derived from both the EEG and ECG (with an accuracy of 98.11%). Which was significantly more accurate than using ECG features alone, but was not significantly more accurate than using EEG features alone. The use of acoustic features alone or in combination with EEG and/or ECG features did not lead to better performances than those achieved with EEG plus ECG or EEG alone. Our results suggest that emotion detection of playing musicians, both felt and musically communicated, when coherent, can be classified in a more reliable way using physiological features than involving acoustic features. The reported machine learning results are a step toward the development of affective brain–computer interfaces capable of automatically inferring the emotions of a playing musician in real-time.**

***Index Terms*—Affective brain–computer interfacing, electrocardiogram (ECG), electroencephalogram (EEG), emotion recognition, music information retrieval (MIR).**

I. Introduction

MUSIC is a powerful method for emotional communication and is known to be capable of eliciting a wide range of emotional responses in listeners [1]. Emotions in music have been studied within various disciplines, including experimental psychology, neuroscience, and computer science. In this context, research has identified different categories of emotions: 1) perceived, i.e., the emotions identified by an individual when listening, without necessarily being affected physiologically; 2) felt, i.e., the emotional responses an individual experiences in body and mind when listening (these, it is worth noting, can be distinct from the perceived ones); 3) intended (or communicated), i.e., the emotions that the performer and/or composer aimed to convey [2].

Thus far, the vast majority of research in this space has been conducted on perceived and felt emotions, focusing on the listener and on the capability of music to induce emotional states. A listener's emotional response to a piece of music can be seen as a function of both the music itself and of the individual. On the one hand, researchers have investigated the relation of low- and high-level features in the musical signals (e.g., tempo, modality, genre) with perceived and induced emotions [3]. For this purpose, the field of music information retrieval (MIR) has developed several techniques to automatically recognize emotions from musical signals [4]. On the other hand, music-induced emotions are known to differ greatly between listeners and be the result of several factors, including the listener's own previous and current mental states or prior experiences [5]. Moreover, research has also shown that the emotions intended by musicians do not always coincide with those perceived or felt by listeners [6].

The investigation of perceived and felt emotions has involved self-reports by listeners, who are asked to judge their emotional experience on different scales relating to emotional models. One of the most widely utilized models is the circumplex model of affect proposed by Russel [7]. In this model, emotion is described using a 2-D space, where the dimensions are represented by arousal (excitement) and valence (pleasantness). Specifically, emotions are distributed on a plane divided by the two orthogonal axes of arousal (ranging from low to high) and valence (ranging from negative to positive), thus leading to four quadrants. However, self-reports may not reflect the actual felt emotions as they are subjective and possibly inaccurate. A more objective alternative approach is represented by physiological signals such as brain activity, heart rate, blood pressure, and skin conductance [8], [9], [10]. Notably, the study reported in [11] showed that a combination of both physiological measures of the listener and acoustic properties of the music may be used to effectively predict emotional responses to a piece of music.

However, the vast majority of studies involving musical stimuli in emotion research have been conducted on passive listeners. As of today, relatively few investigations have been conducted on communicated and felt emotions by musicians during their act of playing a musical instrument. The neurobiological mechanisms of intentional emotional communication by musicians are thus far not well understood. Moreover, methods to automatically

Manuscript received 19 March 2024; revised 17 May 2024; accepted 15 July 2024. This article was recommended by Associate Editor H. Zhou. *(Corresponding author: Luca Turchet.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by Ethical Committee of University of Trento and performed in line with the declaration is that of Helsinki (1964, revised in Fortaleza in 2013).

Luca Turchet and Barry O'Sullivan are with the Department of Information Engineering and Computer Science, University of Trento, 38122 Trento, Italy (e-mail: luca.turchet@unitn.it).

Rupert Ortner and Christoph Guger are with the G.TEC Medical Engineering GmbH, 4521 Schiedlberg, Austria.

Color versions of one or more figures in this article are available at https://doi.org/10.1109/THMS.2024.3430327.

Digital Object Identifier 10.1109/THMS.2024.3430327

recognize emotions communicated and/or felt by playing musicians from physiological signals have been scarcely addressed. Furthermore, to the authors' best knowledge, no research has been conducted on the joint use of physiological and acoustic signals for inferring emotions felt and musically expressed by playing musicians.

While pursuing research on automatic methods to recognize emotions both musically communicated and felt, it is important to be aware of the fact that the emotion intended to be communicated by musicians does not necessarily reflect their actual felt emotions. While this aspect is well known by researchers [12], [13], [14], to the best of our knowledge, the dichotomy between a musician's actual emotional state and his/her intended emotional communication has not been investigated using physiological and acoustic signals.

To bridge the research gaps described above, in this article, we investigate the recognition of felt and musically communicated emotions from both physiological and acoustic signals recorded from musicians who were instructed to perform music in order to communicate a given set of emotions. Specifically, we recorded the acoustic (ACO) signals from the musical instrument along with signals from electroencephalogram (EEG) and electrocardiogram (ECG). Based on the results reported in [11] regarding the improvement of classification accuracies of listeners' emotional responses to music using multimodal signals compared to unimodal ones, we hypothesized that the combined use of EEG, ECG, and acoustic signals would have led to better classification performances compared to the use of the signals individually or in combination (in the following, the combinations are indicated as: ECGACO = ECG plus ACO, EEGACO = EEG plus ACO, EEGECG = EEG plus ECG, and EEGACOECG = all three signals).

Our investigation was driven by the following research questions:

RQ1: Is it possible to automatically recognize felt and musically communicated emotions from playing musicians using physiological signals such as EEG and ECG?

RQ2: How the individual and combined use of physiological and acoustic signals impact the performances of emotion classification algorithms?

RQ3: What is the relation between subjective self-reports of emotions felt and communicated while playing expressive music, as well as between these subjective reports and objective measurements via physiological signals?

RQ4: How do subject-dependent classification performances compare to those of a subject-independent classification?

Answering these questions is important to understand what the best strategies are to automatically infer an emotional state from playing musicians. In particular, unraveling which sources of signal, individual or in combination, lead to better classification performances which would assist the definition of design guidelines for devices dedicated to these kinds of tasks. In particular, this line of inquiry is relevant to the creation of affective computing applications that can predict in real-time the emotional state of a musician and/or his/her intentional emotional communication. Such real-time understanding can then be repurposed for different kinds of services, not only for musicians but also for audiences. These include music therapy, training, and support for improvisation or enhancement of live performances, in particular in networked settings, an endeavor of the Internet of Musical Things field [15]. In this study, we perform initial steps toward this long-term vision, focusing on the offline analysis of the collected signals. Nevertheless, our results could be useful for guiding future real-time implementations.

Notably, our study focused on a classification-based problem rather than regression. Real-time emotion prediction based on previous signal states is also relevant to the creation of interactive affective devices for musical purposes, but this is not the object of our study.

II. Related Work

A. Measuring Emotions From EEG Signals

Much work has been conducted within the field of emotion recognition to find the best-performing electrodes for inferring affective states to reduce the number of electrodes needed for accurate detection. It is well known that both frontal-, parietal-, and central-based electrodes are strongly associated with emotion [16]. Within neuroscience, it is a well-established theory that the frontal lobe is the emotional control center of the brain [17]. Functional connectivity-based analysis looks at the relationships between various or all electrodes. This type of analysis is framed within theories that posit that cognitive functions are the result of functional relationships between anatomically separated parts of the brain [18]. Differential entropy (DE), which is considered to be a nonlinear dynamical system feature [19], has become more popular through EEG emotion-based scientific literature due to some very well-performing studies [20], [21]. DE is used as a complexity measure of continuous EEG signals aimed at calculating the level of vigilance. Derivative features of DE, such as differential asymmetry (DASM) and rational asymmetry (RASM), have also become more prominent in the literature, with many studies carrying out comparative analyses of differential entropy and its derivatives [20].

Both DASM and RASM look at contra-hemispheric electrode pairs, that is, the relationship between paired electrodes such as Fp1 and Fp2. DASM deals with the differential relationship between the electrode's pairs whereas RASM deals with the rational relationship between the electrode pairs. Several studies confirm that both DASM and RASM indices within various frequency bands are related to different affective responses [22]. An underutilized frequency band for the recognition of emotion is the Gamma band (30–140 Hz), which has been shown to be very impactful in emotion-based studies [23]. DASM can be computed using the differential entropy of these electrodes or it can be computed by getting the cumulative log powers of a specified period [22].

Another theory of emotion on which DASM and RASM are predicated upon is the emotion lateralization hypothesis. Some band powers are more useful than others for such measures. Analogous to this, the valence lateralization hypothesis posits that the left hemisphere is considered dominant in the expression of positive emotions and vice versa [24], [25]. The valence lateralization hypothesis or the right-hemisphere hypothesis is often debated [26], [27], with some studies casting doubt on the theories [28]. Nevertheless, many studies confirm its existence to varying degrees [25], [29], particularly for musical stimuli.

The majority of studies involving musical stimuli in emotion research using EEG signals have been conducted investigating passive listeners (see, e.g., [8], [9], and [30]). Only a handful of studies investigated neurophysiological correlates in active music playing, especially considering improvisation activities [31], [32], [33] or error detection [34] rather than actual emotional expressions. A study focusing on the emotions of playing musicians is reported in [35]. The authors investigated the spectral properties of EEG activity in ten piano players instructed to communicate a certain emotion through improvisation on a predefined simple music score. The emotional playing task was contrasted with a neutral playing task. The authors found that the tasks of emotional and neutral playing differed considerably with respect to the state of intended-to-transfer emotion arousal and valence levels. EEG activity differences were observed between distressed/excited and neutral/depressed/relaxed playing. Notably, in such a study, participants were instructed to rate how they thought their performance reflected the intended target emotion rather than their actual felt emotions.

B. Measuring Emotions From ECG Signals

Music has the ability to cause listeners to become more excited (or relaxed), which can lead to increases (or decreases) in heart rate. Such behavior can be detected from ECG signals and, subsequently, used for the classification of emotions. However, only a handful of studies have utilized ECG signals to identify listeners' emotional responses to music [8]. In [36], physiological ECG features were extracted from the time- and frequency-domain, and nonlinear analyses of ECG signals were used to find emotion-relevant features and to correlate them with emotional states. Positive/negative valence, high/low arousal, and four types of emotions (joy, tension, sadness, and peacefulness) were correctly recognized using least squares support vector machine recognizers, with accuracies of 82.78%, 72.91%, and 61.52%, respectively. However, to our best knowledge, the use of ECG in music playing has been largely overlooked.

C. Measuring Emotions From Acoustic Signals

Several studies in musical psychology have focused on the relations between emotions and specific musical attributes, uncovering various associations. For example, happiness is frequently related to pieces characterized by major modes, whereas sadness and anger are usually associated with minor modes [37]; complex, dissonant harmonies are often associated with emotions such as excitement, tension, or sadness, while simple, consonant harmonies with happiness, pleasantness, or relaxation [38]. Leveraging results from musical psychology research, the MIR research community has focused on the topic of music emotion recognition (MER), which aims to create systems able to automatically identify emotions present in musical signals [39], [40], [41]. A recent review of emotionally relevant audio features for MER is reported in [41], which covers both low-level (e.g., spectral features), perceptual (e.g., articulation), and high-level semantic features (e.g., genre).

To date, a variety of MER techniques have been developed (see, e.g., [38], [42], [43], [44], [45]). MER tasks have been primarily approached in two ways. The first consists of regressing a continuous emotional space, such as the arousal–valence one [7], and subsequently clustering such space to obtain a specific emotional vocabulary [46]. The second comprises the classification of a given musical excerpt into one or more emotions, thus becoming a multilabel classification problem with a fixed vocabulary [47]. As shown by the results of existing studies [6], [40], [45] and the Audio Mood Classification task of the 2007–2020 MIR Evaluation eXchange, state-of-the-art solutions for multilabel classifications are still unable to accurately solve simple problems such as the classification of four or five emotion classes.

D. Measuring Emotions From Multimodal Signals

From the review of the relevant literature, it emerges that combining various sources of physiological and acoustic signals for the accurate analysis of human emotions is yet to be explored in depth in music-related studies. One of the few available studies is reported in [11] for the case of emotions induced in listeners by music. The authors combined acoustic descriptors of the music with EEG measures of brain activity. Results showed that over 20% of the variance of the participant's music induced emotions could be predicted by their neural activity and the properties of the music. Moreover, the study showed that the combination of the features extracted from the two types of signals allowed for the prediction of music-induced emotions with significantly higher accuracies than either feature type alone.

To our best knowledge, no study has addressed the automatic recognition of emotions of playing musicians from multimodal sources combining physiological and acoustic signals.

III. Materials and Method

A. Participants

Eleven expert musicians (9 males, 2 females), aged between 20 and 39 (mean age = 29.9, standard deviation = 5.9) took part in the experiment. All participants were healthy adults who did not report having any mental health, mood, or psychiatric problems. The requirement of whether a musician is considered an expert for this study was either a university degree holder in music, past/present professional musician or having more than ten years of experience with their main instrument. Participants were from different nationalities (Spanish, Italian, Argentinian, French, Irish, and Venezuelan) and played different musical genres (rock, classical, flamenco, jazz, heavy metal, blues, folk, ambient, and pop). The musicians' main instruments used during the experiments were: acoustic guitar (eight musicians), electric guitar (one musician), piano (one musician), and handpan (one musician). All participants were right-handed. The reason for involving heterogeneous instruments and music genres was because we aimed to investigate the research questions without

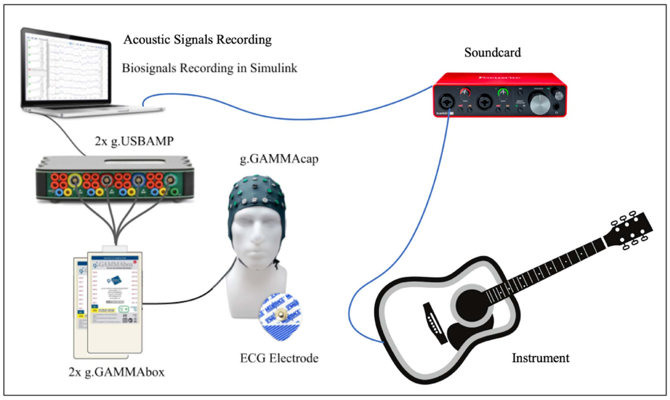

Fig. 1. Data acquisition system setup: Biosignal electrodes (EEG and ECG) are connected to two g.USBAMP biosignal amplifiers via two separate g.GAMMAbox's with a laptop recording the amplified data via USB in Simulink; acoustic signals were recorded on the same computer using a soundcard.

being bound to a specific instrument or genre and, therefore, to achieve more generalizable results.

B. Apparatus

The involved apparatus allowed for the synchronized recording of acoustic and physiological signals. The system setup is shown in Fig. 1 along with a picture of a participant using it.

1) Collected Acoustic Data: Different recording methods were adopted according to the characteristics of the musical instrument utilized by each musician. For instruments with direct input capabilities, such as those with XLR, jack, USB, or MIDI interfaces, musical signals were recorded using the direct input from the instrument to the audio interface (Steinberg UR22 MK2), as is preferred for higher quality audio recordings. For instruments without direct input capabilities, the audio recordings were performed using a Shure SM57 dynamic microphone connected to the audio interface. All pieces were recorded using the software Audacity and exported as WAV files encoded with a bit depth of 32-bits and a sampling rate of 44.1 kHz.

2) Collected Physiological Data: EEG data were recorded using a g.GAMMAcap[2] by g.tec Medical Engineering, a 64-channel cap with g.SCARABEO active electrodes, with two g.GAMMAsys reference active ear clip electrodes. The following 31 EEG channels were used: Fp1, Fp2, AFz, AF3, AF4, AF7, AF8, Fz, F3, F4, F7, F8, Cz, C3, C4, CP3, CP4, CP5, CP6, P1, P2, P3, P4, P5, P6, P7, P8, PO7, PO8, O1, and O2, of which 28 electrodes accounted for 14 contra-hemispheric pairs. AFz was used as a ground electrode and Cz was used as a rereference electrode. The right-side ear clip electrode was used as a reference electrode. Two g.GAMMAbox electrode connector boxes were used to connect the active electrodes to two g.USBamp biosignal amplifiers with a sampling frequency of 256 Hz. The two g.USBamps were connected to a laptop via USB cables.

ECG data were recorded using a single g.GAMMAclip active electrode clip connected directly to the g.GAMMAbox, sharing the same ground electrode with the EEG cap and placed on position V4 of the subjects. A modified version of g.HIsys, a Simulink-based online biosignal processing tool created by g.tec Medical Engineering, was used for the recordings.

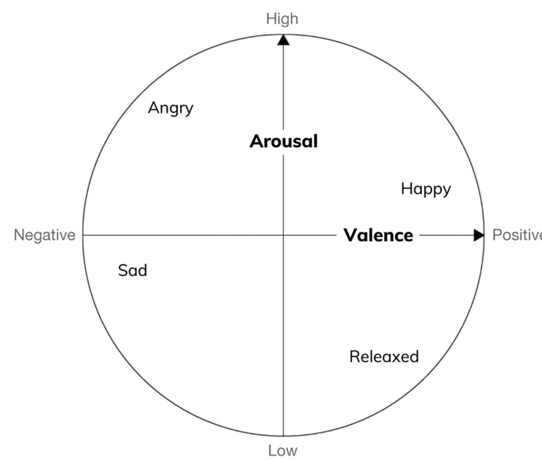

Fig. 2. Valence-Arousal circumplex showing the four quadrants representing the four emotions investigated in the study.

C. Procedure

Participants were given an information sheet and a consent form prior to their day of recording. They were asked to prepare at least one short piece (of about 2 min) for each of the four investigated emotions: happiness, sadness, anger, and relaxation. Such emotions were chosen for two reasons. First, because they have been investigated in several studies on emotional expression in music [48], and because they cover the four quadrants of the 2-D Arousal-Valence space (positive/high arousal, negative/high arousal, negative/low arousal, and positive/low arousal, see Fig. 2) [2]. Second, as they have been tested in previous machine listening setups [38], [49] (see Section II-C).

Participants were instructed to play in such a way as to musically communicate the emotions in question to a potential listener and that the audio recordings of their performance were to be evaluated by listeners at a later date. Specifically, musicians were given imagined scenarios to be used during playing to accentuate the communication of the emotion under question. The given imagined scenarios, inspired by those in the study reported in [50], were as follows.

1) Angry: *"Imagine that your neighbor is being obnoxiously loud and it is keeping you awake. After asking them to be*

quiet numerous times to no avail, you decide to let them know how you feel by playing music outside their door."

2) Happy: *"Imagine that, on a wonderful sunny day you have just found out that you have won the lottery, you decide to play some music before going to collect your winnings."*
3) Relaxed: *"Imagine you are on a tropical island relaxing under a palm tree with your favorite drink. You decide to play some music as you watch the clouds slowly float by."*
4) Sad: *"Imagine that you have just lost a dear friend and are playing a piece as their casket is being brought out of the church and into the cemetery."*

Notably, we did not instruct musicians to be in a specific emotional state where they could actually feel the indicated emotions, but just to communicate such emotions. This was due to the fact that one of our goals was to check whether the emotions felt matched the musically communicated emotions.

Following the setup of audio equipment, participants were asked to sit down on a chair as the biosignal sensors were placed on their bodies. Such sensors were an EEG cap and an ECG adhesive electrode. After this setup, participants underwent a practice trial to familiarize themselves with the system and the task to be accomplished. Subsequently, each musician was asked to play at least one piece from each of the four specified emotions. The order of the emotions was randomized across participants. A total of 56 pieces were recorded (an average of five pieces for each musician).[1]

After each piece was played, musicians were asked to fill out a brief questionnaire, which consisted of the following questions: 1) the objective valence and arousal ratings of the music just played; 2) the valence and arousal they felt while playing the piece. Both were assessed on a 5-point self-assessment Manikin (SAM), which is a nonverbal pictorial assessment technique established in research about felt or communicated emotions [51]. The use of the two questionnaires was devised to take into account that musicians, while playing, may express emotions according to or in response to the intentions of the content they are playing. This does not necessarily reflect their actual felt emotions. This aspect was made clear to the participants, who were instructed to reflect in their ratings any difference between the emotion they felt during playing and the emotion they expressed by playing.

In between each trial, a recording of a soft storm (a neutral stimulus from the expanded version of the international affective digitized sounds [52]) was played for five minutes to accelerate the dissipation of prior identified emotions. After the experiment concluded, participants were asked to fill out a questionnaire about their demographic information such as age, gender, handedness, and musical expertise. This information was collected to assess the possible differences in neuronal activity between these demographics, particularly while considering the lateralization hypotheses [53], [54].

The procedure, approved by the local ethics committee, was in accordance with the relevant ethical standards of the Declaration of Helsinki (1964, revised in Fortaleza in 2013), and compliant with the EU GDPR. Each subject took, on average, two hours to complete the experiment.

IV. Data Analysis

All audio and biosignal recordings (lasting about 2 min) were cut to equally last 1 minute and 30 seconds, removing any initial data preceding the moment in which the musician actually started playing. All analyses were performed on consecutive segments of 30 s each, meaning each trial was split into three. From the recorded multimodal dataset, we extracted acoustic features from each of the pieces of music played by the participants, as well as the physiological features from the participant's EEG and ECG signals. We then attempted to identify subsets of these features that could be used to reliably predict a participant's intentional emotional communication and felt emotion. Such a process was performed for the individual types of signals and combinations thereof.

Before proceeding to the analysis, a check was conducted on the SAM scales ranked by participants for each session. No inconsistency was found between the arousal and valence rankings and the actual emotions musicians were supposed to express by playing. Furthermore, the SAM scores for the emotions felt while playing and those for the emotions communicated by the played music were very consistent and even identical in the vast majority of the cases (see Fig. 3). This occurred for all subjects, none were excluded. An analysis with the Pearson's correlation test showed a significantly strong correlation between rankings of felt and communicated valence (r(56) $= 0.92$, $p < 0.001$) as well as felt and communicated arousal (r(56) $= 0.92$, $p < 0.001$). Since there was no mismatch whatsoever between the communicated and felt emotions, these were treated jointly in the analysis. Accordingly, we only investigated classifications for the four emotional categories rather than for arousal and valence labels.

Two distinct methods of classification were used: 1) subject-dependent classification, whereby a unique classifier is trained for each subject; and 2) subject-independent classification, whereby all subject data (except the subject in question) is used for training the classifier and the subject in question's data is used for testing. These two methods were investigated to assess the specificity and the generalizability of our results and approach.

A. Preprocessing of EEG Data

No hardware filters were applied and thus all preprocessing was done after recording. The majority of the preprocessing steps were executed in the EEGLAB toolbox for MATLAB [55]. First, a notch filter (47–53 Hz) and a temporary FIR bandpass filter (1–30 Hz) were applied. Bad pairs of channels were removed through the use of the clean_rawdata tool, an average of 2.98 ± 1.39 channels were removed from each subject. Data were then manually partitioned into trials based on trial onset data collected via the audio interface. Subsequently, non-EEG channels were removed and separate biosignal files were created. To deal with bad segments of data, the artifact subspace reconstruction (ASR) algorithm was applied to the EEG data to correct bad data segments. A 0.5 s Hanning window with a 50% overlap was

[1] [Online]. Available: https://doi.org/10.5281/zenodo.10396364

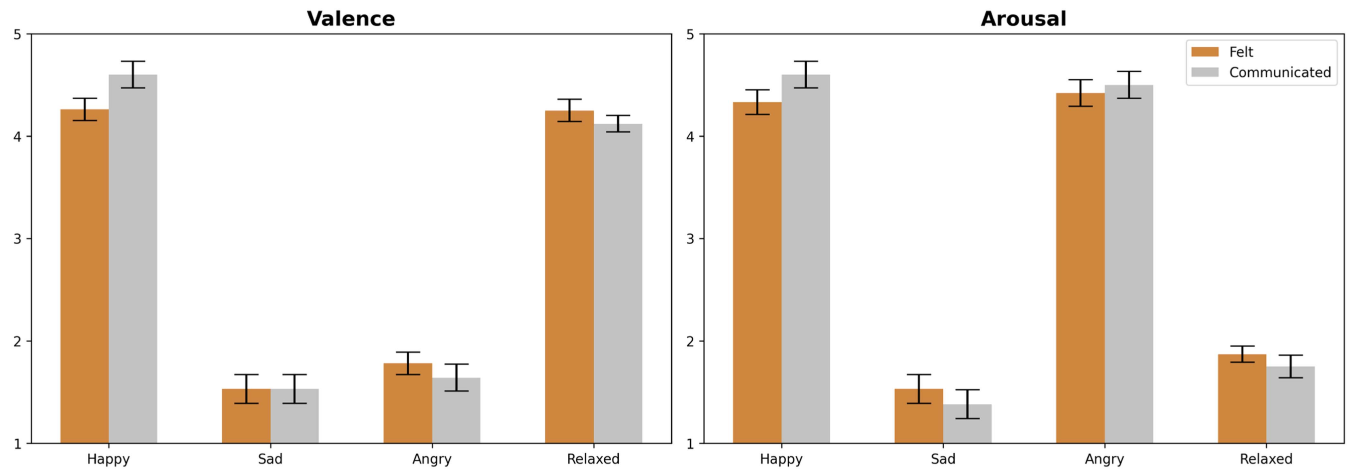

Fig. 3. Mean and standard error of the SAM rankings for both felt and expressed emotions for each of the investigated emotions.

used to decompose the data using PCA, signals with a standard deviation of 10 above or below the RMS variance were corrected based on the baseline data recorded during the neutral stimulus of each trial. In order to reduce muscle artifact noise from the movement associated with playing an instrument, independent component analysis (ICA) was performed. Using the EEGLAB ICLabel tool, all components with less than an 80% probability of being a brain-related component were rejected.

B. Feature Extraction

Concerning EEG, the features chosen for the analysis were RASM and DASM, in line with previous studies in the field. From RASM, a rational asymmetry index can be determined, and various statistical features can be calculated. The indices are calculated as follows:

$$\mathrm{RASM} = (X_{\mathrm{left}}) / (X_{\mathrm{right}})$$
$$\mathrm{DASM} = \log_{10}(\mathrm{TCP}X_{\mathrm{left}}) - \log_{10}(\mathrm{TCP}X_{\mathrm{right}})$$

where X_{left} and X_{right} are the electrode pair, and $\mathrm{TCP}X_{\mathrm{left}}$ and $\mathrm{TCP}X_{\mathrm{right}}$ are the total cumulative power for the electrode pair.

As shown in various RASM- and DASM-based studies, the frequency band for which the asymmetry index is calculated is very important as to what emotional state is to be detected [20], [22]. As such, for RASM, three distinct frequency bands were used and two for DASM. Cz is used as a rereference electrode and as such its root mean squared value is subtracted from the electrode pairs in question. A total of 14 electrode pairs were recorded. The features extracted using Simulink are summarized in Table I.

Regarding ECG, the extracted features were mean heart rate, the standard deviation of RR intervals (SDRR), root mean squared of RR intervals (RMSRR), the number of successive pairs of RR intervals that vary by more than 50 ms (NN50), and the proportion of NN50 divided by the total amount of RR (pNN50).

Regarding the acoustic signals, we used the Essentia library for MIR [56]. Specifically, we computed all features available, which include spectral, time-domain, rhythm, tonal, and high-level descriptors.

TABLE I
EXTRACTED EEG FEATURES: FREQUENCY BAND, MEASURE, AND AMOUNT

Feature extraction method	Frequency bands	Features extracted	Number
RASM index	Low Alpha (8 – 10 Hz)	Kurtosis of ratio	14
		Number of peaks	14
		Max ratio	14
		Median ratio	14
		Mean ratio	14
RASM index	Alpha (8 – 12 Hz)	Kurtosis of ratio	14
		Number of peaks	14
		Max ratio	14
		Median ratio	14
		Mean ratio	14
RASM index	All (1 – 47 Hz)	Kurtosis of ratio	14
		Number of peaks	14
		Max ratio	14
		Median ratio	14
		Mean ratio	14
DASM	Alpha (8 – 12 Hz)	Differential Cumulative	14
DASM	All (1 – 47 Hz)	Differential Cumulative	14
		Total	**238**

After feature extraction, emotion class labels (happy, sad, angry, and relaxed) were added to the features of each trial. The validity of these labels was confirmed using the reported arousal and valence SAM score given by the musicians.

C. Data Segmentation

The data segmentation differed according to the classification approach used.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

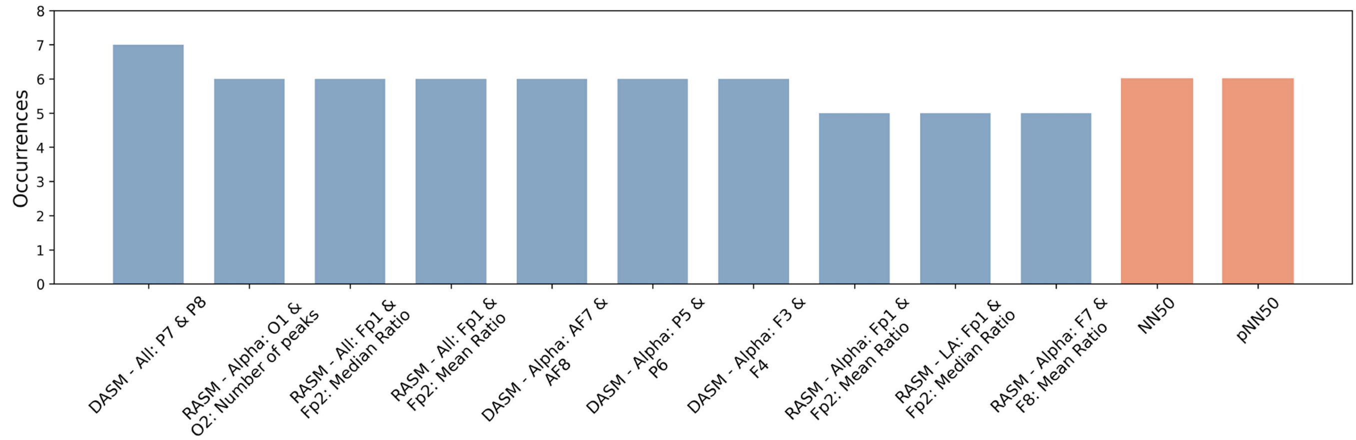

Fig. 4. Total occurrences of selected EEG (light blue) and ECG (light red) features for all subjects.

1) Subject-Dependent Classification: Each subject's dataset was split using the hold-out method. The training and testing subsets represented 66.66% and 33.33% of the data, respectively. These figures were selected as each trial was originally segmented into thirds, and thus, every number of observations will be divisible by three. Meaning we can also create a training dataset that s equally representative of each class. Subsequently, feature selection was applied.

2) Subject-Independent Classification: Training and testing datasets were created for each subject. The training dataset was created by collating all subject data excluding the data of the subject in question. The entirety of the subject's data was then used as a testing dataset. This process was repeated for each subject and the results were then averaged.

D. Feature Selection

The set of EEG, ECG, and acoustic features was combined to make a set of 1957 candidate features (238 EEG features, 5 ECG features, and 1714 acoustic features). We attempt to identify a subset of these features for use in predicting the emotions intentionally communicated by musicians. The Relief feature selection algorithm, taken from the scikit-learn library [57], was employed to select the features that were most informative for the given class. A total of 100 predictors with the highest importance weights were selected to be used in each classifier, except for the classification of the sole ECG signals. This was due to the fact that for ECG, only five features were extracted, and as such, no feature selection was applied. The relief algorithm was run for each subject.

Fig. 4 illustrates the total occurrences of the selected EEG and ECG features for all subjects using the Relief feature selection algorithm. The total selected EEG feature occurrences show that frontal and parietal electrodes were important for classification, particularly Fp1 and Fp2. The mode frequency band found in these results is the "Alpha" frequency band, which covers 8–12 Hz. When EEG and ECG were combined, two ECG features (NN50 and pNN50) were the only non-EEG features to consistently rank as important features (both of which had six occurrences as important features). Concerning acoustic features, the most prominent ones were related to chroma, MFCCs, beats per minute, and loudness.

E. Classification

We investigated the performances of two different classifiers: k-nearest neighbors (KNN) and support vector machine (SVM). The hyperparameters for both models were the same for all subjects. For the KNN classifier, optimization of hyperparameters was carried out to find the best KNN model and hyperparameter settings to use for classification. The KNN classifier hyperparameters used were as follows: a fine KNN model of 1 neighbor and a Euclidean distance metric of equal weight. For the SVM classifier, a linear SVM with an automatic kernel scale was used. To protect against overfitting, various means were employed: modulation of training and test sizes, regularization and normalization, and randomly generated data were also tested (which were performed at the chance level).

V. RESULTS

A. Subject-Dependent Classification

Table II shows the accuracy and F1-score for each type of signal and combination thereof, for both KNN and SVM classifiers.

An ANOVA was performed on two different linear mixed effect models, one for each metric utilized, with the aim to assess whether a classifier performed significantly better as well as the presence of significantly different conditions within each classifier. Specifically, each model had the metric (mean accuracy and mean F1-score), condition (EEG, ECG, acoustic, and the combination thereof), and classifier (KNN, SVM) as fixed factors, and a subject as a random factor. For each model, the assumption for the normality of the residuals was verified.

Regarding the analysis of the mean accuracy, significant main effects were found for the factors condition ($F(6,130) = 39.02$, $p < 0.001$) and the classifier ($F(1,130) = 6.9$, $p < 0.01$), as well

TABLE II
ACCURACY AND F1-SCORE OF THE SUBJECT-DEPENDENT CLASSIFICATION FOR EACH TYPE OF SIGNAL AND COMBINATION THEREOF, FOR BOTH KNN AND SVM CLASSIFIERS; IN BOLD THE HIGHEST VALUES FOR EACH METRIC

	Accuracy		**F1-score**	
	KNN	**SVM**	**KNN**	**SVM**
EEG	94.09%	94.55%	0.93	0.94
ECG	13.94%	13.64%	0.12	0.11
ACO	33.86%	31.21%	0.27	0.27
EEGECG	**98.11%**	63.79%	**0.97**	0.56
EEGACO	74.24%	80.15%	0.68	0.74
ECGACO	31.79%	35.76%	0.26	0.26
EEGECGACO	73.18%	63.79%	0.66	0.56

TABLE III
SIGNIFICANT PAIRWISE COMPARISONS FOR ACCURACY AND F1-SCORE IN THE SUBJECT-DEPENDENT CLASSIFICATION; LEGEND: *** = $p < 0.001$, ** = $p < 0.01$, * = $p < 0.05$, - = NOT SIGNIFICANT

	Accuracy		**F1-score**	
	KNN	**SVM**	**KNN**	**SVM**
EEG - ACO	***	***	***	***
EEG - ECG	***	***	***	***
EEG - ECGACO	***	***	***	***
EEGECG - ACO	***	–	***	–
EEGECG - ECGACO	***	***	**	***
EEGECG - ECG	***	–	***	–
EEGECG - EEG	–	–	–	***
EEGACO - ACO	**	***	**	***
EEGACO - ECG	***	***	***	***
EEGACO - ECGACO	***	***	**	***
EEGACOECG - ACO	**	*	**	–
EEGACOECG - EEG	–	–	–	***
EEGACOECG - ECG	***	***	***	***
EEGACOECG - ECGACO	**	***	**	–
EEGACOECG - EEGECG	–	–	*	–

as their interaction (F(6,130) = 6.69, $p < 0.001$). Regarding the analysis of the mean F1-score, significant main effects were found for the factors condition (F(6,130) = 38.52, $p < 0.001$) and the classifier (F(1,130) = 8.5, $p < 0.01$), as well as their interaction (F(6,130) = 8.08, $p < 0.001$).

Posthoc tests were performed on the fitted model using pairwise comparisons adjusted with the Tukey correction. Concerning the difference between classifiers by condition, for both metrics, the only significant difference was found for EEGECG, which was significantly higher for KNN than SVM (both $p < 0.001$). This result, coupled with the above result regarding the significant main effect of the classifier, allows us to conclude that, globally, KNN attained better discrimination performance than SVM.

Regarding the difference between conditions for each classifier, the analysis revealed several significant pairs, which are reported in Table III.

TABLE IV
ACCURACY AND F1-SCORE OF THE SUBJECT-INDEPENDENT CLASSIFICATION FOR EACH TYPE OF SIGNAL AND COMBINATION THEREOF, FOR BOTH KNN AND SVM CLASSIFIERS; IN BOLD THE HIGHEST VALUES FOR EACH METRIC

	Accuracy		**F1-score**	
	KNN	**SVM**	**KNN**	**SVM**
EEG	36.99%	**86.86%**	0.31	**0.89**
ECG	29.47%	14.77%	0.19	0.08
ACO	22.65%	19.74%	0.19	0.17
EEGECG	21.97%	51.07%	0.17	0.43
EEGACO	18.19%	23.37%	0.15	0.20
ECGACO	58.68%	56.79%	0.57	0.55
EEGECGACO	49.86%	51.07%	0.47	0.43

B. Subject-Independent Classification

Subject-independent classification performed considerably worse than its subject-dependent counterpart with both KNN and SVM classifiers (see Table IV).

VI. DISCUSSION

The conducted experiments and analysis of results allowed us to answer the four research questions described in Section I. Starting from RQ1, our findings indicate that only for subject-dependent classifications it possible to reliably recognize felt and musically communicated emotions from playing musicians using EEG and ECG.

Subject-independent classification performed poorly in comparison with subject-dependent classification (RQ4). This is a result that was expected since training and testing on the same subject typically is more likely to lead to better performances than involving data from other subjects. Our investigation of subject-independent classification showed that there were two main reasons for its poor results. The first is that features selected for both subject-independent and subject-dependent classification were somewhat unique, whereby the most frequently occurring selected feature occurred eight times in 11 of the subjects, while the most frequently occurring selected features occurred two times in 11 of the subjects. Meaning 33.19% of the total 550 selected features were only selected twice among the 11 subjects. This clearly indicates that the selected features have some individualistic nature. Second, the subject-independent classifiers created showed a bias toward a certain emotion: sadness. This is evident when looking at two outliers from the results, whereby a subject scored 0% and another 100%. In both cases, the classifier favored predicting sadness as the estimated emotion. The 100% result was due to the fact that the subject had played an imbalance toward the sad emotion (two sad songs and one of the rest). An acknowledged limitation of the preprocessing and feature selection is the inconsistent removal of electrode pairs across subjects. This may have led to an imbalanced feature set for the subject-independent classification.

It is worth noting that for the subject-independent case, the SVM classifier for EEG signals performed considerably better than the KNN classifier with a mean accuracy of 86.86%, which is very promising for future subject-independent studies. Further investigations would need to take place in order to find better subject-independent features and deal with class imbalances. Our current results suggest that in order to create an effective affective brain–computer interface (aBCI) it is necessary to train each model solely on the data of an individual user. Nevertheless, the results from the SVM classifier for EEG signals are a step in the right direction for the creation of an aBCI.

Concerning RQ2, our hypothesis that using multimodal signals (EEG, ECG, and acoustic) would have led to better classification performances compared to the use of the signals individually or in pairs was only partly satisfied. With regards to the subject-dependent classification and the KNN classifier (which was proven to perform significantly better than SVM), it can be noted from Tables II and III, that the classification that performed best was EEG plus ECG (M = 98.11%), followed by EEG alone (M = 94.09%), with no statistically significant difference between the two. Our findings indicate that emotions communicated by playing musicians can be inferred from the coupling of EEG and ECG signals with high accuracy. However, ECG signals alone achieved low levels of accuracy (M = 13.94%). Importantly, these classification results for EEG and EEG plus ECG not only well matched the emotion of the musical piece participants were supposed to express while playing; they also matched the emotions participants actually felt, as indicated by the SAM rankings. Moreover, our EEG results are in line with those of the study reported in [35], which showed that brain activity patterns differ for different emotional states communicated via music playing.

On the other hand, our results suggest that acoustic features are not likely, by themselves, to be optimal predictors of the emotions felt or musically communicated by playing musicians (M = 33.86%). This result is fully in accordance with other studies using state-of-the-art methods for communicated emotion recognition from individual musical instruments [6].

The combination of ECG and acoustic features led to an increase in performance compared to the ECG features alone (M = 31.79%). Conversely, the combination of acoustic features with EEG features led to a degradation of recognition performances compared to EEG features alone (M = 74.24%). This result is not in accordance with the findings reported in [11], which showed that the music-induced emotional response in a listener could be better predicted by using both EEG and acoustic features compared to the use of such signals alone. However, it is worth noting that the present study investigated a much different case, i.e., that of playing musicians. The use of all three kinds of signals (M = 73.18%) increased the accuracy performances compared to the signals alone, with the exception of EEG. It was also better than the use of pairs of signals only for ECG plus acoustic signals. The same outcomes expressed above are also seen for the F1-score.

In this study, we investigated the recognition of communicated and felt emotions of playing musicians, which is rather different from that of other studies investigating the use of music listening to induce participants' emotional states and collect their physiological signals (see, e.g., [11]). The most noticeable difference is that in the former case, movement is present along with intentional cognitive activities dedicated to communication, while in the second case, a passive activity is involved.

Notably, all participants musically communicated a certain emotion while actually feeling that emotion, i.e., there was no incoherence between what was communicated and what was felt (RQ3). Therefore, our collected dataset did not allow us to investigate the dichotomy that can potentially exist between a musician's actual emotional state and their intended emotional communication through a musical instrument. To our best knowledge, such a dichotomy has not been investigated yet using acoustic, EEG, ECG, or other physiological signals, and further research is needed to address this case.

Our findings indicate that frontal and parietal electrodes are important for emotion classification. As a matter of fact, all EEG features in the mean top ten used were from these positions (see Fig. 4). This is in accordance with other studies on emotion classification [9], [16], [25], in particular on music-induced emotions [11] as well as in emotions communicated during active playing [35]. Moreover, NN50 and pNN50 appear to be the most important ECG features for emotion classification when combined with EEG signals.

Our results suggest that physiological features can lead to more reliable detection of emotional states as well as emotional musical communication of playing musicians (when these two match) than the use of acoustic features derived from the music played. In part, this is likely to be caused by the inaccuracies of current MER algorithms and acoustic feature extraction methods [4]. On the other hand, our results may be interpreted as an indication that emotional communication and emotional responses during music playing are the result of processes that are more internal to the player than the acoustic properties of the music played.

It is important to reflect on the implications that the results of our study could offer in terms of both opportunities and challenges. Concerning the opportunities, the creation of effective aBCIs specific to the music playing scenario could spur the emergence of a whole new set of real-time or offline services for musicians, such as those that can be envisioned by relying on the recent Internet of Musical Things paradigm [15]. These include applications in which the classified emotional states of musicians during a performance are repurposed, wirelessly and in real-time, into visualizations, sounds, or haptic stimulations, thus enhancing the audience members' experience. Another example is represented by cloud-based applications that allow a musician to monitor his/her emotional states during musical activities such as composition, performance, teaching, and learning. Or again, data related to musicians' affective states could be collected and utilized in extended reality environments during emerging types of musical activities such as those envisioned in the Musical Metaverse [58].

On the other hand, ethical, privacy, and security issues need to be deeply considered for the development of such interfaces and services given the involvement of a highly sensitive type of data such as the physiological one. It is crucial that aBCIs for musicians adopt privacy-by-design approaches and privacy-enhancing technologies [59], as well as effective security methods [60]. This will contribute to ensuring that the machine learning solutions underlying such interfaces are both technically and socially robust. Furthermore, a set of questions need to be addressed, such as follows. Under which conditions or situations do musicians want to have their emotions recognized while playing in order to take advantage of the new musical services? Which impact such systems can have on musicians' practices? To address such challenges, there is a need for joint interdisciplinary research at the confluence of engineering, human-computer interaction, sociology, and music studies.

It is worth noting that the reported study presents some limitations. First, a relatively small number of musicians, music stimuli, and a restricted number of different musical instruments were involved. Nevertheless, the involved number of participants is in line with that of similar studies (see, e.g., [35]). Second, the participants' gender was not balanced, with more males than females. Nevertheless, an in-depth analysis showed that no major differences were present in the data of participants considered by gender. Third, not all possible physiological signals were utilized. The use of sensors measuring parameters related to oxygen consumption or galvanic skin responses could reveal other results.

In our study, participants did not improvise but were asked to prepare the musical excerpts in advance and practice to express well the intended emotions at the moment of recording. In principle, the number of times participants had practiced the pieces could have affected the emotions felt while recording, but it is not possible to conclude with certainty that this had a positive or a negative influence. Nevertheless, this procedure is consistent with what musicians normally do to perform a nonimprovised piece. Thus, our results are relevant to the design of concrete applications operating in real-world scenarios.

VII. Conclusion

This article investigated the automatic recognition of felt and musically communicated emotions using both physiological and acoustic signals, which were recorded from musicians instructed to perform music in order to communicate a given set of emotions. Specifically, we recorded the acoustic signals from the musical instrument along with signals from an EEG and ECG. Results showed that the best classification performances using a subject-dependent classification and a KNN classifier were achieved by using features derived from both the EEG and ECG. Such a classification was significantly more accurate than using ECG features alone but was not significantly more accurate than using EEG features alone. The use of acoustic features alone or in combination with EEG and/or ECG features did not lead to better performances than those achieved with EEG plus ECG or EEG alone.

Our results suggest that emotion detection of playing musicians, both felt and musically communicated when these two match, can be performed in a more reliable way using physiological features than involving acoustic features. This may be due in part to the fact that existing MER algorithms are still not optimal. On the other hand, it is possible that emotional responses during music playing are the result of processes that are more internal to the player than the acoustic properties of the music played. This is in line with the well-known fact that musicians are capable of musically communicating emotions that are different from those actually felt while playing. Our study highlighted the need to more deeply investigate such a dichotomy using physiological and acoustic signals.

The classification performances were accurate only using a subject-dependent approach, which suggests that to develop a successful aBCI it is necessary to train each model solely on the data of an individual user. The reported machine-learning results are a step toward the development of aBCIs capable of automatically inferring the emotions of a playing musician in real-time. The applications of such interfaces could be manifold, including neurofeedback training to improve creativity, detection of stress or flow states to support music learning or performance, or the control of peripherals wirelessly connected to musical instruments for the Internet of Musical Things applications. Our future work will focus on the development of such interfaces and applications.

References

[1] K. R. Scherer, "Which emotions can be induced by music? What are the underlying mechanisms? And how can we measure them?," *J. New Music Res.*, vol. 33, no. 3, pp. 239–251, 2004.

[2] P. N. Juslin and J. A. Sloboda, *Music and Emotion: Theory and Research*, P. N. Jusling and J. A. Sloboda, Eds. Oxford, U.K.: Oxford Univ. Press, 2001.

[3] P. Juslin, "Cue utilization in communication of emotion in music performance: Relating performance to perception," *J. Exp. Psychol.: Hum. Percep. Perform.*, vol. 26, no. 6, 2000, Art. no. 1797.

[4] J. S. Gómez-Cañón et al., "Music emotion recognition: Toward new, robust standards in personalized and context-sensitive applications," *IEEE Signal Process. Mag.*, vol. 38, no. 6, pp. 106–114, Nov. 2021.

[5] P. G. Hunter, E. G. Schellenberg, and U. Schimmack, "Feelings and perceptions of happiness and sadness induced by music: Similarities, differences, and mixed emotions," *Psychol. Aesthetics Creativity Arts*, vol. 4, no. 1, 2010, Art. no. 47.

[6] L. Turchet and J. Pauwels, "Music emotion recognition: Intention of composers-performers versus perception of musicians, non-musicians, and listening machines," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 30, pp. 305–316, 2022.

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

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

[9] I. Daly et al., "Neural correlates of emotional responses to music: An EEG study," *Neurosci. Lett.*, vol. 573, pp. 52–57, 2014.

[10] X. Hu, F. Li, and T.-D. J. Ng, "On the relationships between music-induced emotion and physiological signals," in *Proc. Int. Soc. Music Inf. Retrieval Conf.*, 2018, pp. 362–369.

[11] I. Daly et al., "Music-induced emotions can be predicted from a combination of brain activity and acoustic features," *Brain Cogn.*, vol. 101, pp. 1–11, 2015.

[12] A. Gabrielsson and P. Juslin, "Emotional expression in music performance: Between the performer's intention and the listener's experience," *Psychol. Music*, vol. 24, no. 1, pp. 68–91, 1996.

[13] P. Juslin, "Emotional communication in music performance: A functionalist perspective and some data," *Music Percep.*, vol. 14, no. 4, pp. 383–418, 1997.

[14] A. G. V. Zijl and J. Sloboda, "Performers' experienced emotions in the construction of expressive musical performance: An exploratory investigation," *Psychol. Music*, vol. 39, no. 2, pp. 196–219, 2011.

[15] L. Turchet, C. Fischione, G. Essl, D. Keller, and M. Barthet, "Internet of musical things: Vision and challenges," *IEEE Access*, vol. 6, pp. 61994–62017, 2018.

[16] J. Zhang and P. Chen, "Selection of optimal EEG electrodes for human emotion recognition," *IFAC-PapersOnLine*, vol. 53, no. 2, pp. 10229–10235, 2020.

[17] R. B. Firat, "Opening the "black box": Functions of the frontal lobes and their implications for sociology," *Front. Sociol.*, vol. 4, 2019, Art. no. 3.

[18] M. P. Van Den Heuvel and H. E. H. Pol, "Exploring the brain network: A review on resting-state fMRI functional connectivity," *Eur. neuropsychopharmacol.*, vol. 20, no. 8, pp. 519–534, 2010.

[19] X. Li et al., "EEG based emotion recognition: A tutorial and review," *ACM Comput. Surv.*, vol. 55, no. 4, pp. 1–57, 2022.

[20] R.-N. Duan, J.-Y. Zhu, and B.-L. Lu, "Differential entropy feature for eeg-based emotion classification," in *Proc. 6th Int. IEEE/EMBS Conf. Neural Eng.*, 2013, pp. 81–84.

[21] Y. Li et al., "EEG-based emotion recognition under convolutional neural network with differential entropy feature maps," in *Proc. IEEE Int. Conf. Comput. Intell. Virtual Environments Meas. Syst. Appl.*, 2019, pp. 1–5.

[22] B. Reuderink, C. Mühl, and M. Poel, "Valence, arousal and dominance in the EEG during game play," *Int. J. Auton. Adaptive Commun. Syst.*, vol. 6, no. 1, pp. 45–62, 2013.

[23] M. Li and B.-L. Lu, "Emotion classification based on gamma-band EEG," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2009, pp. 1223–1226.

[24] M. R. Mowla, R. I. Cano, K. J. Dhuyvetter, and D. E. Thompson, "Affective brain-computer interfaces: Choosing a meaningful performance measuring metric," *Comput. Biol. Med.*, vol. 126, 2020, Art. no. 104001.

[25] 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.

[26] W. D. Killgore and D. A. Yurgelun-Todd, "The right-hemisphere and valence hypotheses: Could they both be right (and sometimes left)?," *Social Cogn. Affect. Neurosci.*, vol. 2, no. 3, pp. 240–250, 2007.

[27] G. Gainotti, "Emotions and the right hemisphere: Can new data clarify old models?," *Neuroscientist*, vol. 25, no. 3, pp. 258–270, 2019.

[28] S. J. Reznik and J. J. Allen, "Frontal asymmetry as a mediator and moderator of emotion: An updated review," *Psychophysiol.*, vol. 55, no. 1, 2018, Art. no. e12965.

[29] E. S. Pane, A. D. Wibawa, and M. H. Purnomo, "Improving the accuracy of EEG emotion recognition by combining valence lateralization and ensemble learning with tuning parameters," *Cogn. Process.*, vol. 20, no. 4, pp. 405–417, 2019.

[30] Y.-P. Lin et al., "EEG-based emotion recognition in music listening," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 7, pp. 1798–1806, Jul. 2010.

[31] L. A. Dikaya and I. A. Skirtach, "Neurophysiological correlates of musical creativity: The example of improvisation," *Psychol. Russia: State Art*, vol. 8, no. 3, pp. 84–97, 2015.

[32] M. Sasaki, J. Iversen, and D. E. Callan, "Music improvisation is characterized by increase EEG spectral power in prefrontal and perceptual motor cortical sources and can be reliably classified from non-improvisatory performance," *Front. Hum. Neurosci.*, vol. 13, 2019, Art. no. 435.

[33] D. S. Rosen, Y. Oh, B. Erickson, F. Z. Zhang, Y. E. Kim, and J. Kounios, "Dual-process contributions to creativity in jazz improvisations: An SPM-EEG study," *NeuroImage*, vol. 213, 2020, Art. no. 116632.

[34] I. Ariza, L. J. Tardón, A. M. Barbancho, I. De-Torres, and I. Barbancho, "Bi-LSTM neural network for eeg-based error detection in musicians' performance," *Biomed. Signal Process. Control*, vol. 78, 2022, Art. no. 103885.

[35] J. E. Pousson et al., "Spectral characteristics of EEG during active emotional musical performance," *Sensors*, vol. 21, no. 22, 2021, Art. no. 7466.

[36] Y.-L. Hsu, J.-S. Wang, W.-C. Chiang, and C.-H. Hung, "Automatic ECG-based emotion recognition in music listening," *IEEE Trans. Affect. Comput.*, vol. 11, no. 1, pp. 85–99, First Quarter 2020.

[37] A. Gabrielsson and E. Lindström, *The Influence of Musical Structure on Emotional Expression*. Oxford, U.K.: Oxford Univ. Press, 2001, pp. 223–248.

[38] C. Laurier, O. Meyers, J. Serrà, M. Blech, P. Herrera, and X. Serra, "Indexing music by mood: Design and integration of an automatic content-based annotator," *Multimedia Tools Appl.*, vol. 48, no. 1, pp. 161–184, 2010.

[39] Y. Kim et al., "Music emotion recognition: A state of the art review," in *Proc. Int. Soc. Music Inf. Retrieval Conf.*, 2010, pp. 937–952.

[40] X. Yang, Y. Dong, and J. Li, "Review of data features-based music emotion recognition methods," *Multimedia Syst.*, vol. 24, no. 4, pp. 365–389, 2018.

[41] R. Panda, R. M. Malheiro, and R. P. Paiva, "Audio features for music emotion recognition: A survey," *IEEE Trans. Affect. Comput.*, vol. 14, no. 1, pp. 68–88, First Quarter 2023.

[42] C. Laurier, P. Herrera, M. Mandel, and D. Ellis, "Audio music mood classification using support vector machine," *MIREX Task Audio Mood Classification*, pp. 2–4, 2007.

[43] Y. Yang, Y. Lin, Y. Su, and H. Chen, "A regression approach to music emotion recognition," *IEEE Trans. Audio, Speech Lang. Process.*, vol. 16, no. 2, pp. 448–457, Feb. 2008.

[44] A. Aljanaki, Y. Yang, and M. Soleymani, "Developing a benchmark for emotional analysis of music," *PLoS One*, vol. 12, no. 3, 2017, Art. no. e0173392.

[45] R. Panda, R. Malheiro, and R. Paiva, "Novel audio features for music emotion recognition," *IEEE Trans. Affect. Comput.*, vol. 11, no. 4, pp. 614–626, Fourth Quarter 2020.

[46] M. Soleymani, M. Caro, E. Schmidt, C. Sha, and Y. Yang, "1000 songs for emotional analysis of music," in *Proc. ACM Int. Workshop Crowdsourcing Multimedia*, 2013, pp. 1–6.

[47] S. Chowdhury, A. Vall, V. Haunschmid, and G. Widmer, "Towards explainable music emotion recognition: The route via mid-level features," in *Proc. Int. Soc. Music Inf. Retrieval Conf.*, 2019, pp. 237–243.

[48] A. Gabrielsson and P. N. Juslin, "Emotional expression in music," in *Handbook of Affective Sciences*, R. J. Davidson, H. H. Goldsmith, and K. R.E. Scherer, Eds. Oxford, U.K.: Oxford Univ. Press, 2003, pp. 503–534.

[49] P. Alonso-Jiménez, D. Bogdanov, J. Pons, and X. Serra, "Tensorflow audio models in essentia," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2020, pp. 266–270.

[50] L. Turchet and A. Rodà, "Emotion rendering in auditory simulations of imagined walking styles," *IEEE Trans. Affect. Comput.*, vol. 8, no. 2, pp. 241–253, Second Quarter 2017.

[51] M. M. Bradley and P. J. Lang, "Measuring emotion: The self-assessment manikin and the semantic differential," *J. Behav. Ther. Exp. Psychiatry*, vol. 25, no. 1, pp. 49–59, 1994.

[52] W. Yang et al., "Affective auditory stimulus database: An expanded version of the international affective digitized sounds (IADS-E)," *Behav. Res. Methods*, vol. 50, no. 4, pp. 1415–1429, 2018.

[53] P. Rodway, L. Wright, and S. Hardie, "The valence-specific laterality effect in free viewing conditions: The influence of sex, handedness, and response bias," *Brain Cogn.*, vol. 53, no. 3, pp. 452–463, 2003.

[54] C. Mikutta, G. Maissen, A. Altorfer, W. Strik, and T. König, "Professional musicians listen differently to music," *Neuroscience*, vol. 268, pp. 102–111, 2014.

[55] A. Delorme and S. Makeig, "EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *J. Neurosci. Methods*, vol. 134, no. 1, pp. 9–21, 2004.

[56] D. Bogdanov et al., "Essentia: An audio analysis library for music information retrieval," in *Proc. Int. Soc. Music Inf. Retrieval Conf.*, 2013, pp. 493–498.

[57] N. Pilnenskiy and I. Smetannikov, "Feature selection algorithms as one of the python data analytical tools," *Future Internet*, vol. 12, no. 3, 2020, Art. no. 54.

[58] L. Turchet, "Musical metaverse: Vision, opportunities, and challenges," *Pers. Ubiquitous Comput.*, vol. 27, pp. 1–17, 2023.

[59] K. Xia et al., "Privacy-preserving brain–computer interfaces: A systematic review," *IEEE Trans. Computat. Social Syst.*, vol. 10, no. 5, pp. 2312–2324, Oct. 2023.

[60] S. L. Bernal, A. H. Celdrán, G. M. Pérez, M. T. Barros, and S. Balasubramaniam, "Security in brain-computer interfaces: State-of-the-art, opportunities, and future challenges," *ACM Comput. Surv.*, vol. 54, no. 1, pp. 1–35, 2021.