Emotional Impact of Source Localization in Music Using Machine Learning and EEG: a proof-of-concept study

Eleonora De Filippi [1], Timothy Schmele [1], ARIJIT NANDI [1], Adan Garriga Torres [1], and Alexandre Pereda-Banos [1]

[1]Affiliation not available

October 30, 2023

Abstract

Little is currently known about how varied source locations affect a listener's emotional reaction to music. Here, using spectral features extracted from electrophysiology (EEG) data, we tested through machine learning whether four music source positions (front, back, left, and right) could be accurately distinguished according to the type of valence in a subject-wise manner. The findings demonstrate that distinct EEG correlates can reliably classify the four source locations and that the effect is stronger when music with a negative emotional valence is played outside of the listener's visual field. This proof-of-concept study may pave the way for advanced spatial audio analysis approaches in music information retrieval by considering the listener's emotional impact depending on the source direction of incidence.

Emotional Impact of Source Localization in Music Using Machine Learning and EEG: a proof-of-concept study

Eleonora De Filippi*, Timothy Schmele*, Arijit Nandi, Adan Garriga Torres
Alexandre Pereda-Baños

Abstract—Little is currently known about how varied source locations affect a listener's emotional reaction to music. Here, using spectral features extracted from electrophysiology (EEG) data, we tested through machine learning whether four music source positions (front, back, left, and right) could be accurately distinguished according to the type of valence in a subject-wise manner. The findings demonstrate that distinct EEG correlates can reliably classify the four source locations and that the effect is stronger when music with a negative emotional valence is played outside of the listener's visual field. This proof-of-concept study may pave the way for advanced spatial audio analysis approaches in music information retrieval by considering the listener's emotional impact depending on the source direction of incidence.

Index Terms—Spatial Music, Emotion Recognition, Affective Computing, EEG, Machine Learning, SVM, Source Localization in Music.

1 Introduction

Darwin once said that "music mirrors or captures the relationship between affective state and sound [...]" [1]. Music's ability to modulate cognitive and emotional processes has been widely documented over the years [2], [3], [4], [5], [6], making it a relevant tool to investigate the brain correlates of emotional processes [7]. However, little is known about the impact of different sound-source locations on the emotional response to music, as only a few studies have addressed this issue [8], [9], [10], [11], [12]. In particular, the study conducted by Asutay et. al. [11], provided evidence that the effects of spatial source location on attentional processes are mediated by the emotional information conveyed by the sound [11]. The authors also demonstrated that a sound source behind the participant led to a more robust affective response in the listeners [11]. In another work, Tajadura-Jiménez et. al. concluded that sound sources outside the visual field produce emotional states of increased arousal [9], though these effects were more pronounced for natural sounds. Similarly, the work of Ekman and Kajastila showed that sounds are judged by the listener as scarier when they come from the back as compared to the front [8], although context has been found to be an important factor in eliciting the desired effect [12]. In a more detailed study using everyday sound events, Drossos et. al [10] showed that lateral positions do increase the listeners' affective state significantly, although this is dependent on the content of the audio sample used.

Moreover, there is a strong connection between space and music, as music can, in turn, evoke sensations of space and movement as a sense of *intrinsic space*, i.e. a metaphorical space, created by musical features in melody, harmony or rhythm, as opposed to the *literal, physical space* a sound source may occupy [13]. The most common effect is that of associating the perceived pitch with a sense of spatial height [14], [15], although alternative spatial representations for the same also exist latently [16]. Furthermore, a correlation between the absolute pitch of a musical piece and emotional affect has been shown [17]. In a study conducted by Eitan et. al. [18], in which participants were asked to associate music with imagined, spatial motions of a human character, it was shown that most musical parameters significantly affect the imaginary motion, indicating a strong correlation between music and space perception.

Here[1], we investigated whether four different music source spatial locations (i.e., front, back, left, and right) are reflected in a distinct pattern of electrophysiological activity that can be captured by a machine-learning approach. Moreover, we explored the interaction between distinct music source locations and the emotional salience of the musical excerpts played. The dimensional model supports the idea that emotions can be modeled as combinations of a few fundamental and basic dimensions. Valence and arousal, sometimes known as the "circumplex model," are two fundamental qualities that researchers unanimously concur are necessary to understand emotions [19]. The valence level

- *E. De Filippi, A. Nandi and A. Pereda-Baños are with the department of Big Data at Eurecat, Centre Tecnologic, 08005 Barcelona, Spain.*
- *T. Schmele and Adan Garriga Torres are with the department of Multimedia at Eurecat, Center Tecnologic, 08005, Barcelona, Spain. Timothy Schmele is as well affliated with the Institute for Musicology and Music Informatics at the University of Music, 76131 Karlsruhe, Germany.*
- ** = these authors equally contributed to this work*
- *E-mail: eleonora.defilippi@eurecat.org (E.D.F) timothy.schmele@eurecat.org (T.S) arijit.nandi@eurecat.org (A.N)*

1. *A preliminary and earlier result of this work has been presented as a Late-breaking/Demo (LBD) session to the Music-Information Retrieval community at ISMIR 2022 conference to get feedback from fellow researchers in the field of affective computing and music information retrieval. Link:* https://ismir2022.ismir.net/program/lbd/

varies from unpleasant (negative) to pleasant (positive), while the arousal level, specifically, ranges from not aroused (low arousal) to thrilled (high arousal). We used music excerpts as an audio stimulus. While the more common stimuli, such as noise or single frequencies can offer more control, as they reduce the number of variables, they also have several downsides to the purpose of this study, however. For example, improved auditory localization along the median plane requires broadband signals [20], which means that pure frequencies would make front-back confusions much more prevalent. Moreover, as is the case with noise signals, in particular, these types of acoustic stimuli are not commonly perceived or referred to as *music* [21]. In order to get the participants to listen to the audio stimulus *musically*, meaning that the stimulus is generally recognized as music, it was therefore decided to use musical examples as stimuli.

We recorded electrophysiological (EEG) data while participants were listening to musical excerpts characterized by either positive or negative valence, both with middle values of arousal, and occurring from different spatial source locations. To take into account individual differences, we performed subject-based classification between each pair of spatial locations, according to the type of valence. We hypothesized that when the music source was located outside the listener's visual field (i.e., back, right, left) it would lead to a different electrophysiological pattern and impact on the affective state as compared to frontal source localization.

2 Materials and Methods

2.1 Stimuli

The music excerpts used in this study were taken from the Database for Emotional Evaluation of Music (DEAM) [22]. DEAM provides over 1800 royalty-free, annotated samples of music, of which 58 are full-length songs. However, for this study, we were interested in the 1744 excerpts of 45s in length. The styles of these excerpts range widely across a variety of Western popular music genres, including rock, pop, electronic, country, and jazz, but also some more experimental examples that appear to be borderline spoken word or audio akin to field recordings. Since the experiment was carried out in Spain with mainly Spanish-speaking participants, we avoided music samples that appeared to rely too heavily on language.

The emotional characterization of the music in DEAM is done using two dimensions: valence (positive, neutral, or negative) and arousal (low, medium, or high). The data provided also came in two versions: a dynamic score on a scale from -10 to 10, sampled at 2Hz and begins only after the second 15, thus annotating only the last 30s of the music samples respectively, as well as static annotation data, which represents an overall score of the sample and was collected using a nine-point Self-Assessment Mannequin (SAM) scale [23].

In order to choose which samples to use for this study, we categorized the musical excerpts into 3 groups of high, mid, and low values along each emotional dimension (i.e., valence and arousal), based on the static evaluation metric. To evaluate the effect of valence in this study, those samples that fell into the mid-arousal category were first separated. Of those, samples that were additionally in the positive valence category (166 in total) or negative valence category (72 in total) were chosen and placed into two separate categories respectively. The samples had to be shortened to 30s in order to keep the overall length of the experiment within reasonable bounds. Thus, an average of the dynamic annotations was calculated for each sample selected. After standardizing the average dynamic and static annotations, the error between the two measures could be obtained. This error served as an ordering mechanism, along which the final samples could be selected, starting with the sample with the least error between the static and average dynamic annotation. Considering that each sample is now 30s long, 36 samples from each category were selected to achieve an approximate length of 50min for the entire the experiment. The final selection of samples by index number as corresponding to DEAM [22] can be found in Table 1.

TABLE 1
Final selection of stimulus indices in each positive and negative valence class. The numbers correspond to the sample index as given by DEAM [22].

Positive valence:	37, 42, 56, 65, 107, 131, 276, 317, 415, 431, 466, 486, 677, 693, 777, 814, 846, 976, 1071, 1079, 1122, 1156, 1204, 1298, 1340, 1346, 1523, 1538, 1544, 1614, 1647, 1742, 1749, 1853, 1865, 1954
Negative valence:	74, 146, 167, 176, 178, 187, 194, 198, 205, 218, 219, 223, 238, 239, 299, 316, 359, 361, 478, 480, 482, 501, 506, 560, 608, 621, 656, 715, 736, 794, 854, 855, 865, 957, 1146, 1823

As mentioned above, the first 15s of each sample were cut away to obtain only the last 30s, i.e. the region for which the dynamic annotation was available. To avoid clicking or startling the participant at the beginning of a sample, a linear 1s fade-in was applied to all samples. An objective perceptual loudness assessment was done over all samples using the *replay-gain* method [24]. A Python implementation was adapted from the original implementation in MATLAB[2], using pink noise at -20dBFS as a reference.[3] The mean loudness deviation was -6.49dB, with a standard deviation of 4.24dB. The maximum was 8dB and the minimum -18.35dB. Thus, the perceived loudness of the different samples ranged considerably. This was also commented on by the participants during the preliminary trials. However, equalizing the samples by their replay-gain values to have a more similar perceptive loudness against each other would make the data collected incomparable to the original annotations in DEAM. Thus, it was decided to not apply the replay-gain compensation values and keeping each music excerpt at its original relative loudness level.

2.2 Experimental design

For this study, we opted for a block-design experiment, in which the final sequence of audio samples was arranged into blocks of 3 randomly selected musical excerpts without repetition. Random selection was done anew for each participant. First, from each category, i.e. positive or negative valence, the samples were shuffled and grouped into blocks

2. See http://replaygain.hydrogenaudio.org/mfiles/, last accessed using on September 13th, 2022.
3. This code can be accessed via https://github.com/multimedia-eurecat/Neuromuse/tree/master/replaygain

of three musical excerpts. Then, the blocks from each category were combined in random order into a single sequence of $36 + 36 = 72$ samples. The whole experimental process is illustrated in Fig. 1.

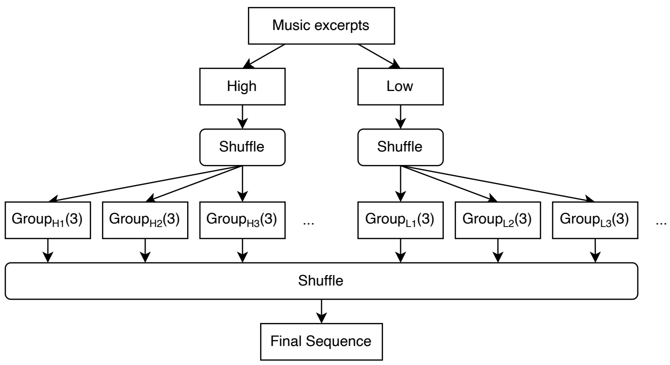

Fig. 1. Grouping and shuffling of the music samples.

Fig. 2 shows the final sequence of sections in the entire experiment. The participant is first shown an introduction, where the experiment and the SAM questionnaire are explained. Contained in this introduction is a short test if the participant has understood what the SAM represents. Then, a baseline rest period of $120\mathrm{s}$ is recorded, where the participant should remain entirely still, while in total silence. After that, the main experiment begins. In each group, 3 music samples of the same category are played. Before each sample, a rest period of $5\mathrm{s}$, represented by a cross in the middle of the screen is first presented as a sort of mental reset for the participant. At the end of each musical sample, the participant has to fill out the SAM questionnaire rating their emotional response to that particular musical stimulus. At the end of each block, another additional questionnaire was shown, where the participant answered how exhausted and attentive they felt.

The experiment ended after all blocks and their respective musical samples had been played to the participant. The total time necessary to play all 72 excerpts of music was about 36 minutes. Together with the rest periods, the minimum length of the experiment was around 44 minutes. Depending on the speed that the participants were able to read through the introduction and answer all questionnaires, the whole experimental session would last between 50 to 60 minutes.

2.3 Participants

We recruited a total of 20 healthy participants (10 males and 10 females) with a mean age of 28.66 years (SD = 5.53), no history of psychiatric or neurological disorders, and with normal hearing. Furthermore, subjects had no prior experience or formal music training. Nearly all participants were right-handed, with only one participant being left-handed. After preprocessing of the EEG signal, 3 participants were removed from the analysis due to excessive artifacts. Only 17 participants were included in the analysis, comprising 9 females and 8 males with a median age of 28 (20-38).

2.4 Experimental procedure

Upon arrival, the participant was seated in the studio control room and signed the consent form. After setting up the EEG cap, we guided the participant into the studio environment. The EEG data was recorded using a 19-electrode Neuroelectrics® Instrument Controller (NIC2)[4] at a sampling rate of 250 Hz. Next, subjects were informed about the experimental protocol, its approximate duration and the meaning of the SAM scales [23] used to rate their emotional response to the music stimuli presented throughout each block. During the whole experiment, the participant was comfortably seated in the center of the room (see Fig. 3). Prior to the musical stimuli presentation, we started the experiment by recording baseline EEG activity by showing participants a cross in the middle of the screen for 2 minutes (see Fig. 5). Lastly, we instructed the participants not to move, particularly during playback of the stimulus, to reduce muscular artifacts in the EEG data. The experiment was conducted according to the Helsinki Declaration and all subjects signed the consent form.

2.5 Experimental Setup

2.5.1 *Acoustic Environment*

All measurements were done in the 3D audio post-production studio of Eurecat in Barcelona, as shown in Fig. 3. The room has a size of $7.06\mathrm{m}$ by $5.13\mathrm{m}$ by $3.16\mathrm{m}$ and is acoustically isolated from the outside world. Inside, it is acoustically treated, with acoustic diffusion panels on the walls and absorption panels on the ceiling. The reverberation time is relatively short, with an RT_{60} of $0.398\mathrm{s}$ at $125\mathrm{Hz}$ to $0.253\mathrm{s}$ at $8\mathrm{kHz}$. The average RT_{60} between $500\mathrm{Hz}$ and $1000\mathrm{Hz}$ is around $0.293\mathrm{s}$.

2.5.2 *Spatial Conditions*

The audio stimuli were played from four positions: front, back, left, and right (see Fig. 4). A loudspeaker was placed in each position. No phantom sources were used, meaning that audio was always played through only one loudspeaker at any time. In order to keep the first reflections comparable between each position, we opted to keep each loudspeaker as close to the wall as possible. However, this also meant that the front and back loudspeakers would be positioned at a larger distance from the listener.

To correct for the differences in distance between the loudspeakers, each loudspeaker was calibrated to $75\mathrm{dBSPL}$ using pink noise at $-20\mathrm{dBFS}$ at the center listening position. The calibration was carried out using an NTI AL1 Acoustilyzer in combination with an NTI MiniSPL measurement microphone. It was determined during the pilot study that this level was an acceptable loudness for all participants and a good level to compromise between loud and quiet stimuli (see 2.1).

2.5.3 *Hardware*

The loudspeakers used were of the type Genelec 8040, fed by a Focusrite Scarlet 18i20 soundcard. The audio playback was done with the *sounddevice* module for Python, running on a windows laptop computer.[5] The participants were seated in

4. https://www.neuroelectrics.com/
5. See https://github.com/multimedia-eurecat/Neuromuse/blob/master/emotiondirectionexperiment/neuromuse_server.py

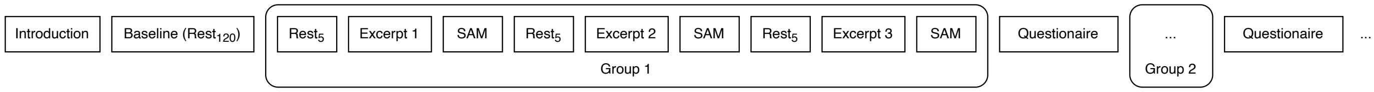

Fig. 2. Final sequence of music samples and full experiment.

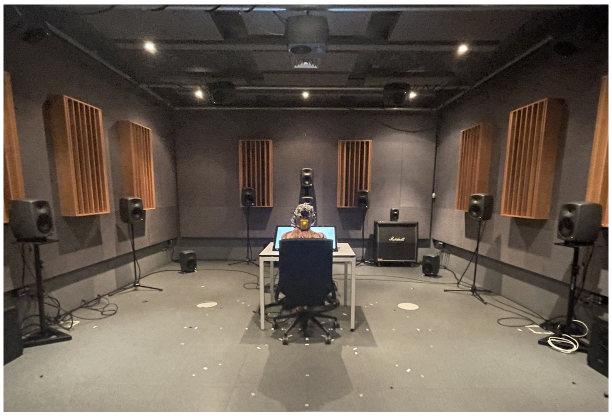

Fig. 3. The studio environment as seen from the back speaker using a wide-angle lens. Apart from the back speaker not seen in this picture, the other speakers especially placed for the experiment were the ones on the far left and right side of the participant and the one up ahead, covered by the head of the participant. All four speakers are placed at ear level. The touch screen is placed as flat as possible on the table in front, making sure it does not obstruct the view from the participant to the front speaker to avoid sound occlusion. All other speakers seen in this picture did not partake in the experiment. However, they had to remain for logistic purposes.

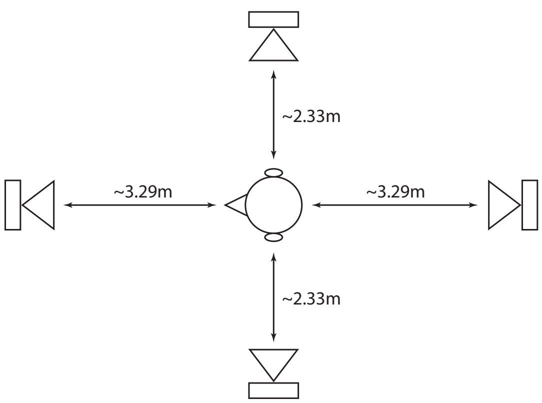

Fig. 4. Experiment layout showing the four playback positions and approximate distances of each loudspeaker to the listener in the center.

the center of the room, with a table in front of them. Laying nearly flat, the participants had a Raven MTi2 touch screen in front of them, on which they would have instructions shown to them, as well as answer questions during the experiment.

2.5.4 User Interface

The interface was designed in a web browser using HTML and CSS, with the functional elements programmed in Javascript using Jquery.[6] All interactive elements were designed with touch interaction in mind, meaning big buttons or clickable areas with only multiple-choice questions and no text input of any kind. Some examples of the interface are shown in Fig. 5. Accidental double taps are accounted for and filtered out. Buttons to continue are always hidden and deactivated until all questions in a questionnaire are answered to indicate the necessity of this task to the participant and avoid accidental progression without answers in the trial. Whenever a button to continue is shown, the participant is also able to change their mind before sending off the result to be recorded, meaning they can alter their answer before pushing the 'continue' button. Otherwise, the web page would change the view and progress immediately upon answering a question.

The Python script and the web front-end communicated using a simple socket connection over *localhost*. All activity on the touch screen was recorded in the form of signals sent over this socket connection to know where the participant clicked when and if their opinion on a certain question has changed before confirming. A USB and video extension was laid through the floor into the control room, where the experimenters could monitor the participant, their activity on the touchscreen, and all incoming sensor data.

6. See https://github.com/multimedia-eurecat/Neuromuse/blob/master/emotiondirectionexperiment/emotion.html

2.6 EEG preprocessing

EEG data were analyzed in an offline manner using the EEGLAB toolbox [25] on Matlab R2019b (The Mathworks, Inc.). The preprocessing steps included downsampling of the signal to 130 Hz and the application of a bandpass Butterworth filter ranging from 0.01 up to 40 Hz. To correct eye blinks and muscular artifacts, we used the Independent Component Analysis (ICA) algorithm. For each subject, we manually removed all components capturing artifacts. Afterward, we epoched the EEG data and created eight distinct datasets for each subject according to the experimental condition (i.e, spatial position and type of valence). Finally, we applied a spatial filter to reduce the volume conduction effect, using the surface Laplacian transform inspired by the spherical spline method described by [26], [27], [28].

2.7 Feature extraction

2.7.1 Time-frequency analysis

To preserve information about the temporal dynamics, we transformed the EEG data into the time-frequency domain using Complex Morlet Wavelet convolution (CMW). This method comprises a complex-valued sine wave tapered by a Gaussian window and is stated as follows:

$$CMW = e^{-t^2/2s^2} e^{i2\pi ft} \tag{1}$$

Where $e^{-t^2/2s^2}$ represents the real-valued Gaussian and $e^{i2\pi ft}$ is the result of Euler's formula combined with a sine wave [29].

We chose CMW instead of alternative approaches like the Short-time Fourier Transform or the Hilbert Transform,

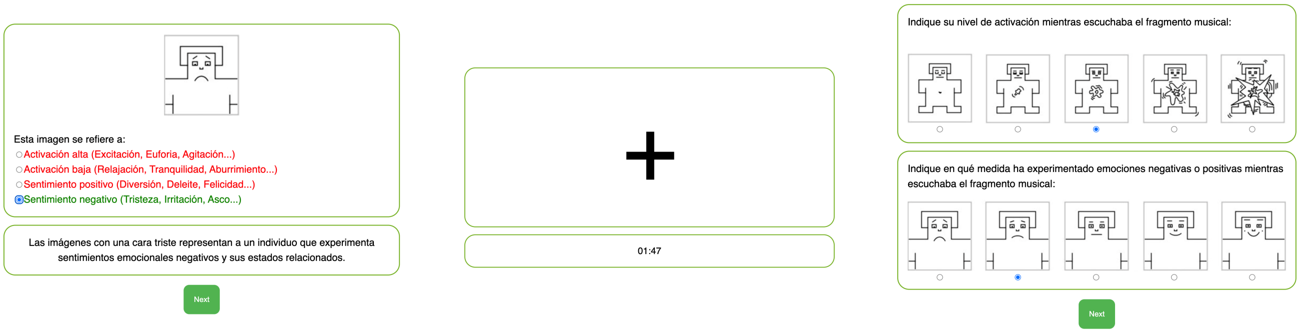

Fig. 5. Three examples of the touch interface in the web browser. On the left, an example of the preliminary test to reinforce the SAM model before beginning the experiment is shown. The text color indicating the correct answer only appears immediately *after* answering the question. The center image shows the view when the participant is asked to remain calm and focus on the cross during the baseline data acquisition and rest period before each stimulus. To avoid that the participant gets nervous during the 2 minute wait, the remaining time is also shown to indicate that UI remains responsive. The third image shows the SAM multiple choice questionnaire that is presented after every music sample to evaluate the arousal and valence state of the participant via self-reporting.

is because CMW is a Gaussian-shaped wavelet in the frequency domain. It is important to precisely set the width of the Gaussian, denoted here as s, while performing convolution with a Complex Morlet Wavelet since it affects how the time-frequency analysis trades off its ability to resolve temporal and spatial issues (see [29]).

The parameter s is expressed as $s = c\ /2\pi f$, where c denotes the number of cycles of the wavelet, which is dependent on the frequency f of the same. A narrower Gaussian with fewer cycles in the time domain leads to a high temporal resolution but reduced spectral precision, and vice-versa with a wider Gaussian. Because of this, we used a variable number of cycles, ranging from 3 to 10, rising as a function of frequency to achieve a fair trade-off between temporal and spectral resolutions.

Since we were interested in all frequency bands, we selected a range of frequencies going from 1 Hz up to 40 Hz.

Following the use of CMW convolution, we retrieved the power from the coefficients and then used a decibel-baseline normalization, utilizing all neutral trials as a baseline. We used a sliding-window strategy to reduce the time-frequency data for every trial in order to increase the sample size. There were a total of 39 windows every trial, each lasting 1 second and overlapping by half a second.

Then, we calculated the average change in power compared to the neutral baseline for seven frequency bands (delta $1-4$ Hz, theta $4-8$ Hz, low alpha $8-10$ Hz, high alpha $10-12$ Hz, low beta $13-18$ Hz, high beta $18-30$ Hz, and gamma $31-40$ Hz). Within each window and for all the 19 channels and the seven frequency bands, the features extracted were the mean power, the standard deviation of the mean, and the frontal alpha asymmetry (FAA). The FAA coefficients were calculated for the channel pairs Fp1-Fp2 and F3-F4 in both low-alpha ($8-10$ Hz) and high-alpha ($10-12$ Hz) bands. The resulting feature array consisted of 351 samples for each class with a total of 270 features.

2.8 Classification and feature selection

For data classification, we utilized MATLAB R2022ba Statistics and the Machine Learning Toolbox. As a base classifier, the linear Support Vector Machine (SVM) supervised learning approach was chosen, which uses a hyperplane as a decision boundary to optimize the margin of separation between two classes. Herewith, SVMs give a metric that permits scaling the certainty with which a window sample is allocated to one of the two classes: the sample's distance from the separation hyperplane. To evaluate the classifier's performance robustly, we used 6-fold cross-validation to train and test the classifier, allocating all windows in one trial to the same fold. Having said that, we also ran the 6-fold cross-validation fifty times and averaged the results across different classification runs.

It is well known that feature extraction and selection strategies assist to reduce computing complexity and develop models with greater generalization capabilities, in addition to enhancing predictive power [30], [31]. That is, we used the Bioinformatics toolbox of MATLAB R2022a to do feature selection due to the large dimensionality of our dataset. The goal was to improve the classifier's learning performance and find the most common discriminative characteristics shared by all participants. We rated the characteristics based on their importance between the classes, using the t-test as an independent criterion for binary classification. For each feature, the built-in function in MATLAB calculates the absolute value of the two-sample t-test with pooled variance estimate. Essentially, the technique calculates how probable it is for a particular characteristic that the difference in mean and variance across classes happened by chance. Finally, we identified the top 20 characteristics for each topic and combined them to determine which features were shared by all participants.

2.8.1 *Statistical comparisons*

We used the Wilcoxon rank-sum method to investigate whether the SVM performances were significantly above chance, thus we statistically compared accuracy distributions of real-labeled data with surrogate data (i.e., randomly shuffled labels). Furthermore, data from the self-assessment SAM questionnaire were analyzed using a general linear model, the multivariate analysis of variance (MANOVA), on IBM SPSS Statistics [32].

3 Results

We investigated through machine learning whether the four music position sources could be accurately differentiated according to the type of valence in a subject-wise manner using spectral features extracted from EEG data cut into 1-second windows. The results of all cross-validation runs for each participant are presented in Figure 6. The corresponding accuracy averaged across subjects for each binary classification run is summarized in Table 2. For both positive and negative valence, we showed that the highest average accuracy was reached when classifying the frontal localization versus each of the three sources located outside the visual field. In particular, this effect was stronger when classifying pairs of source locations using events characterized by negative valence.

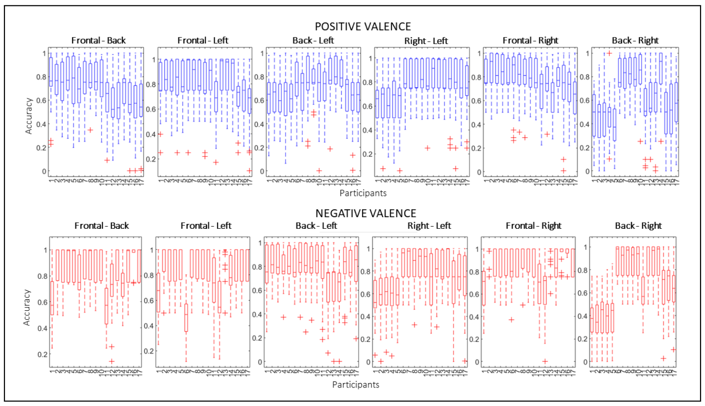

Fig. 6. Within-subject classification results of each binary classification between music position sources according to the type of valence. We performed 6-fold cross-validation 50 times, such that the boxplots depict the results of 50 classification runs for each participant.

TABLE 2
Classification results averaged across participants for each pair of binary classification runs presented according to the valence type. An asterisk indicates that the average accuracy is significantly above the chance level ($p<0.05$).

Type of valence	Pair of music source locations	Mean accuracy across subjects
Positive valence	Frontal - Back	70%*
	Frontal - Left	80%*
	Back - Left	69%*
	Right - Left	76%*
	Frontal - Right	77%*
	Back - Right	61%*
Negative valence	Frontal - Back	85%*
	Frontal - Left	82%*
	Back - Left	78%*
	Right - Left	74%*
	Frontal - Right	87%*
	Back - Right	67%*

3.1 Highest-ranked features

To understand which channels and frequency bands were the most discriminative between the four location sources depending on the type of valence, we applied a feature selection algorithm and merged together the top twenty features for each subject. As represented in Figure 7, the results showed that the electrophysiological correlates of the difference between frontal location and the three sources located outside the visual field (i.e., back, right, left) rely on different activities of channels mainly located in frontal and central areas, especially in the highest frequencies. In particular, we found that when using musical excerpts with negative valence, the difference between frontal location and each of the three sources out of the field of view is based on activity in beta (low and high) and gamma bands in channels Fp1, F3, F4, Fz, Cz, and T8, and in FAA measures for pairs of channels Fp1-Fp2 and F3-F4. On the other side, when comparing the source locations using positive valence, we showed that also brain activity in the alpha band, together with beta and gamma was important for differentiating frontal position from each of the other three sources. Moreover, in the case of positive valence, channels from posterior sites, especially Pz, P7, and P8, as well as central, frontal sites and FAA, were relevant for the classification of source locations, indicating a more widespread involvement of different brain areas.

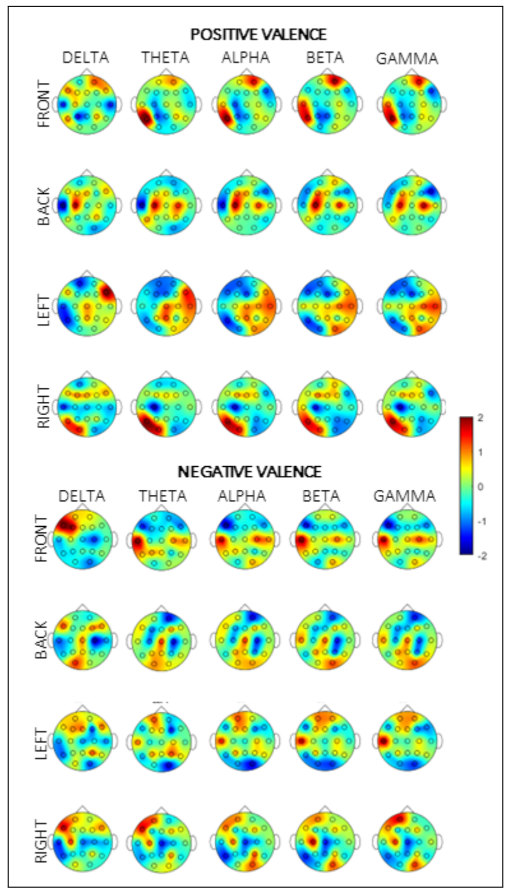

Fig. 7. Topoplots indicating brain activity for each of the main frequency bands according to the source locations (i.e., front, back, left, and right) and the type of valence (i.e., negative and positive). Colors in brain plots indicate the power in that specific channel and frequency band, with red showing the highest power and blue the lowest.

3.2 Self-assessment SAM questionnaire

We analyzed behavioral data using the multivariate analysis of variance (MANOVA) to assess differences in the self-reported ratings of valence and arousal of the SAM questionnaire according to source location and type of valence.

TABLE 3
Average reported levels of arousal and valence by means of the SAM questionnaire on a Likert scale from 1 to 5.

Type of valence	Source location	Average SAM rating for arousal	Average SAM rating for valence
Positive valence	Frontal	3,41	3,25
	Back	3,57	3,21
	Left	3,22	3,17
	Right	3,45	3,36
Negative valence	Frontal	2,63	2,03
	Back	2,90	2,18
	Left	2,88	2,28
	Right	3,45	2,15

Averaged reported levels of perceived arousal and valence are presented in Table 3 and shown in Figure 8 for each participant.

Results showed that there was a significant difference in arousal and valence ratings based on the type of event (source location and type of valence of musical excerpts), $F(14, 12428) = 28.133$, $p = 0.000$, $Wilk'slambda = 0.740$, partial eta squared = 0.14. The source locations depending on the type of valence had a significant effect both on reported levels of perceived arousal, $F(7, 1215) = 18.477$, $p = 0.000$, partial eta squared = 0.096, and valence, $F(7, 1215) = 47.886$, $p = 0.000$, partial eta squared = 0.216.

Post-hoc analysis revealed that there were significant differences between trials with positive valence and negative valence within each source location ($p = 0.000$, Bonferroni corrected). In particular, musical excerpts characterized by positive valence elicited higher reported levels of arousal ($p = 0.000$) and valence ($p = 0.000$) for each of the four source locations. Differences between sources were not significant ($p > 0.05$) when comparing the same type of valence (i.e., either positive or negative), except the levels of reported arousal between the back and right when music with positive valence was played ($p = 0.03$, Bonferroni corrected).

4 Discussion

In this work, we analyzed the impact of different source locations of music depending on the type of valence on the listener's affective brain processing by employing machine learning tools.

We demonstrated that frontal location can be accurately distinguished from each of the three sources (back, right, and left) located outside the listener's visual field. In particular, our results suggested that the emotional connotation of music (i.e., positive and negative valence) mediated the impact of the different source locations on the brain's electrophysiological signal, as reflected by music characterized with negative valence yielding higher classification performances in differentiating between the spatial sources as compared to musical excerpts characterized by positive valence.

Furthermore, by applying a feature selection procedure we showed that playing music from different source locations led to different electrophysiological brain responses in the highest frequencies (alpha, beta, and gamma) and in channels belonging to the frontal, central, and also parietal areas in the case of positive valence. The importance of beta and gamma bands that we found here is consistent with earlier research showing the significance of these bands for differentiating between various emotional states [33], [34], [35]. Moreover, a previous study has found that the alpha band in parietal channels was associated with the processing of auditory stimuli, while the gamma band activity was related to music awareness [36]. Interestingly, we found the FAA between pairs of channels Fp1-Fp2 and F3-F4 to be an important measure for distinguishing between different locations, both for positive and negative valence conditions. Alpha activity in the frontal site has been largely used as an index of emotional processing, reflecting motivation and dominance of perceived emotion. Indeed, in the literature, positive emotional stimuli have been related to a relative increase in left hemisphere activity, whereas negative emotional stimuli have been associated with a larger right hemisphere activity [37], [38]. For example, a previous study has found that musical excerpts characterized by positive valence induced lower frontal alpha power in the left hemisphere [39]. In addition to valence and arousal, frontal asymmetry was also linked to other factors, such as self-reported dominance [40].

Music has generally been used in research as a tool to elicit emotional responses in participants and study emotional processes in the brain [7], [41], [42]. Together with this, recent applications of Brain-Computer Interfaces have used music as a way to convey information and/or feedback in a real-time manner to the subjects based on their own brain activity [43], [44], [45], [46]. However, the difficulty of participants in engaging and sustaining genuine emotional states in an experimental context, particularly when trying to elicit complex emotions, has generally been a significant hurdle for neuroimaging and BCI studies based on affective processes. In this regard, the results of this study may pave the way for more effective use of music as a stimulus in experimental settings, by playing it from a source located outside the visual field, especially when trying to elicit emotions falling within the negative valence.

On the other side, analysis of sounds and their spatial orientation in relation to the listener is relevant in the context of spatial music. Here, spatial music refers to musical composition practices that specifically target spatial aspects of sound as a compositional parameter, such as the sound position or specific aspects of room acoustics [47], [48], [49]. Indeed, emotion elicited through spatial music listening is not an aspect that is not often considered in the field. The discussion around space in music tends to be often of philosophical [47], or conceptual nature [49], [50] and often centers around aspects in electronics or hardware [48], [50], technology [51], [52], [53] or taxonomy [54]. The most common considerations in spatial composition techniques center around the analytical location of a sound source, following its trajectory or simulating, or alluding to acoustics that differ from those present in the current concert hall [53]. In a survey conducted in [55], composers are more often than not concerned with those spatial aspects in music that can be parameterized on a technical level and lesser with the emotional space can have on the listener, e.g. the "dramatic role" of space [55].

The results in this study indicate that an analysis of spatial music would need to take into account the correlation between the extracted emotional impact of the more

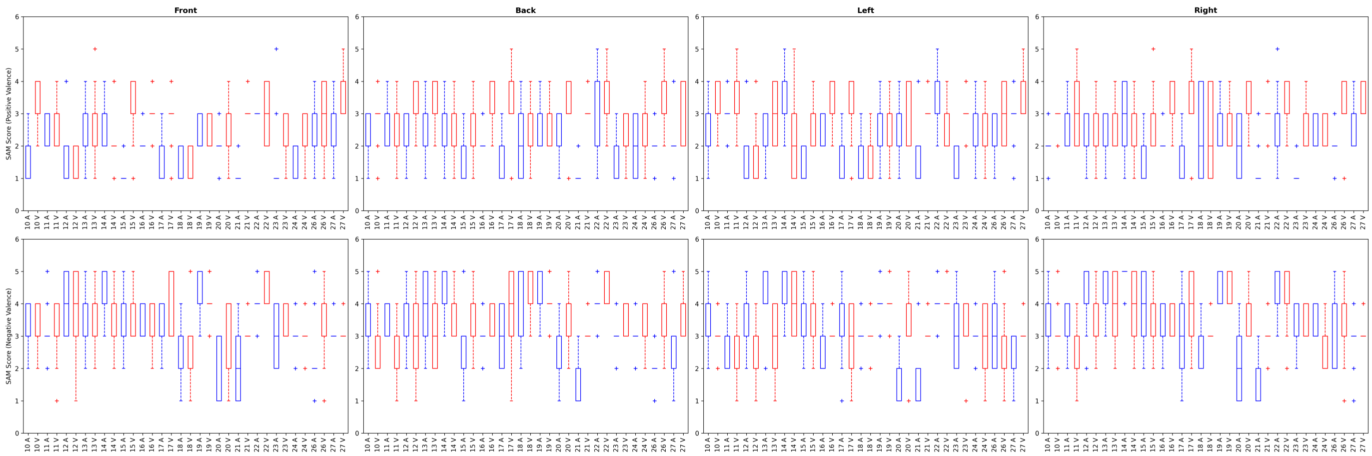

Fig. 8. Results of the SAM questionnaire for each position and valence class (positive or negative). For each participant, both the valence (blue) and arousal (red) results are shown in each case.

traditional musical parameters, such as melody and rhythm, with spatial features, when it comes to the emotional impact music can have on the listener. This can either be done by analyzing the score and correlating musical phrases with the spatial position of the corresponding instruments at any given time, or by analyzing the multichannel audio signal in the case of electroacoustic works. For example, using Ambisonic signals [56], the energy vector can be computed to estimate the audio source location [57]. Otherwise, other formats such as Parametric Object-Based Audio Coding in MPEG [58] can provide this information as stored in its metadata. This can then be combined with other emotional factors in music to achieve a comprehensive prediction of the listener's emotional response.

Nevertheless, this study does present some limitations. Most importantly, the relatively small sample size may limit the generalization of our results. Also, despite having found a different brain pattern of activity in the EEG signal related to affective processing of music depending on the various source locations, this effect was not reflected in the analysis of subjective ratings of both arousal and valence. In fact, we did not find any significant differences between positions when analyzing trials with the same type of valence. This may have been due to overthinking, or conscious evaluation on behalf of the subjects, rather than an immediate and intuitive reaction. When being asked how one would evaluate a piece of music, one would have to execute said task by recollecting what was just heard. This evaluation will thus be skewed by the importance a subject might place on different aspects in the music. Therefore, if a subject has little to no experience in associating spatial position with musical significance, then the recollection of the heard excerpt will most likely be focused on aspects like melody and rhythm, filtering out the spatial direction from which the excerpt was heard from. This means that mentally, i.e. in the inner ear, the music may have been heard *aspatially*.

5 Conclusion

The present study has shown that machine learning methods are able to discern a listener's affective brain processing between different spatial positions of sound sources as a function of positive or negative affect. Annotated musical excerpts were classified into two groups of both median arousal and low or high valence values respectively. These samples were presented to the listener from the front, the lateral left or right positions, or the back, in random order. Our results showed that frontal location, compared to each of the other three sources located outside the visual field, is associated with different brain electrophysiological patterns related to emotional processing. In fact, we found a significant involvement of alpha, beta, and gamma frequency bands in frontal and central sites, together with FAA measures, in distinguishing between such source locations. These findings were not reflected in the subjective rating analysis, hinting that the subjects may have excluded the spatial aspect of the music when consciously evaluating the heard excerpts. While more analysis is necessary, these first results prove promising. Further analysis is necessary to understand how the source location is able to influence the emotional impact of music, particularly focusing on arousal. Lastly, future work will also have to include the median plane to get a more comprehensive view of the effects of spatial source locations on the listener's affective state.

References

[1] I. Cross, "Music and cognitive evolution," in *Oxford Handbook of Evolutionary Psychology* (L. Barrett and R. Dunbar, eds.), ch. 45, pp. 649–668, Oxford: Oxford University Press, 2007.

[2] J. A. Sloboda, "Music structure and emotional response: Some empirical findings," *Psychology of music*, vol. 19, no. 2, pp. 110–120, 1991.

[3] C. L. Krumhansl, "An exploratory study of musical emotions and psychophysiology.," *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, vol. 51, no. 4, p. 336, 1997.

[4] Y. Hou and S. Chen, "Distinguishing different emotions evoked by music via electroencephalographic signals," *Computational intelligence and neuroscience*, vol. 2019, 2019.

[5] V. Putkinen, S. Nazari-Farsani, K. Seppälä, T. Karjalainen, L. Sun, H. K. Karlsson, M. Hudson, T. T. Heikkilä, J. Hirvonen, and L. Nummenmaa, "Decoding music-evoked emotions in the auditory and motor cortex," *Cerebral Cortex*, vol. 31, no. 5, pp. 2549–2560, 2021.

[6] P. Vuust, O. A. Heggli, K. J. Friston, and M. L. Kringelbach, "Music in the brain," *Nature Reviews Neuroscience*, vol. 23, no. 5, pp. 287–305, 2022.

[7] S. Koelsch, "Brain correlates of music-evoked emotions," *Nature Reviews Neuroscience*, vol. 15, no. 3, pp. 170–180, 2014.

[8] I. Ekman and R. Kajastila, "Localization cues affect emotional judgments–results from a user study on scary sound," in *Audio Engineering Society Conference: 35th International Conference: Audio for Games*, Audio Engineering Society, 2009.

[9] A. Tajadura-Jiménez, P. Larsson, A. Väljamäe, D. Västfjäll, and M. Kleiner, "When room size matters: acoustic influences on emotional responses to sounds.," *Emotion*, vol. 10, no. 3, p. 416, 2010.

[10] K. Drossos, A. Floros, A. Giannakoulopoulos, and N. Kanellopoulos, "Investigating the impact of sound angular position on the listener affective state," *IEEE Transactions on Affective Computing*, vol. 6, no. 1, pp. 27–42, 2015.

[11] E. Asutay and D. Västfjäll, "Attentional and emotional prioritization of the sounds occurring outside the visual field.," *Emotion*, vol. 15, no. 3, p. 281, 2015.

[12] S. Hughes and G. Kearney, "Fear and localisation: Emotional fine-tuning utlising multiple source directions," in *Audio Engineering Society Conference: 56th International Conference: Audio for Games*, Audio Engineering Society, 2015.

[13] E. Di Bona, "Listening to the space of music," *Rivista di estetica*, no. 66, pp. 93–105, 2017.

[14] E. Rusconi, B. Kwan, B. L. Giordano, C. Umilta, and B. Butterworth, "Spatial representation of pitch height: the smarc effect," *Cognition*, vol. 99, no. 2, pp. 113–129, 2006.

[15] S. C. Hedger, H. C. Nusbaum, O. Lescop, P. Wallisch, and B. Hoeckner, "Music can elicit a visual motion aftereffect," *Attention, Perception, & Psychophysics*, vol. 75, no. 5, pp. 1039–1047, 2013.

[16] Z. Eitan and R. Timmers, "Beethoven's last piano sonata and those who follow crocodiles: Cross-domain mappings of auditory pitch in a musical context," *Cognition*, vol. 114, no. 3, pp. 405–422, 2010.

[17] L. Jaquet, B. Danuser, and P. Gomez, "Music and felt emotions: How systematic pitch level variations affect the experience of pleasantness and arousal," *Psychology of Music*, vol. 42, no. 1, pp. 51–70, 2014.

[18] Z. Eitan and R. Y. Granot, "How music moves:: Musical parameters and listeners images of motion," *Music perception*, vol. 23, no. 3, pp. 221–248, 2006.

[19] P. J. Lang, M. K. Greenwald, M. M. Bradley, and A. O. Hamm, "Looking at pictures: Affective, facial, visceral, and behavioral reactions," *Psychophysiology*, vol. 30, no. 3, pp. 261–273, 1993.

[20] J. Blauert, *Spatial hearing: the psychophysics of human sound localization*. Cambridge, MA, USA: MIT press, 1997. Chapter 2.3.1: Directional hearing the median plane, pp. 97–116.

[21] C. Nussbaum, "Musical perception," in *The Oxford handbook of philosophy of perception* (M. Matthen, ed.), pp. 495–514, Oxford, UK: Oxford University Press, 2015.

[22] A. Aljanaki, Y.-H. Yang, and M. Soleymani, "Developing a benchmark for emotional analysis of music," *PLOS ONE*, vol. 12, March 2017.

[23] M. M. Bradley and P. J. Lang, "Measuring emotion: The self-assessment manikin and the semantic differential," *Journal of Behavior Therapy and Experimental Psychiatry*, vol. 25, no. 1, pp. 49–59, 1994.

[24] D. J. Robinson, *Perceptual model for assessment of coded audio*. PhD thesis, University of Essex, UK, 2002.

[25] A. Delorme and S. Makeig, "Eeglab: an open source toolbox for analysis of single-trial eeg dynamics including independent component analysis," *Journal of neuroscience methods*, vol. 134, no. 1, pp. 9–21, 2004.

[26] F. Perrin, O. Bertrand, and J. Pernier, "Scalp current density mapping: value and estimation from potential data," *IEEE Transactions on Biomedical Engineering*, no. 4, pp. 283–288, 1987.

[27] F. Perrin, J. Pernier, O. Bertnard, M.-H. Giard, and J. Echallier, "Mapping of scalp potentials by surface spline interpolation," *Electroencephalography and clinical neurophysiology*, vol. 66, no. 1, pp. 75–81, 1987.

[28] F. Perrin, J. Pernier, O. Bertrand, and J. Echallier, "Spherical splines for scalp potential and current density mapping," *Electroencephalography and clinical neurophysiology*, vol. 72, no. 2, pp. 184–187, 1989.

[29] M. X. Cohen, "A better way to define and describe morlet wavelets for time-frequency analysis," *NeuroImage*, vol. 199, pp. 81–86, 2019.

[30] H. Liu, E. R. Dougherty, J. G. Dy, K. Torkkola, E. Tuv, H. Peng, C. Ding, F. Long, M. Berens, L. Parsons, *et al.*, "Evolving feature selection," *IEEE Intelligent systems*, vol. 20, no. 6, pp. 64–76, 2005.

[31] G. Anthony and H. Ruther, "Comparison of feature selection techniques for svm classification," in *10th International Symposium on Physical Measurements and Signatures in Remote Sensing*, pp. 1–6, 2007.

[32] IBM Corp., "IBM SPSS Statistics for Windows 25.0," 2017. Armonk, NY: IBM Corp.

[33] E. S. D. Glauser and K. R. Scherer, "Neuronal processes involved in subjective feeling emergence: Oscillatory activity during an emotional monitoring task," *Brain Topography*, vol. 20, no. 4, pp. 224–231, 2008.

[34] M. Li and B.-L. Lu, "Emotion classification based on gamma-band eeg," in *2009 Annual International Conference of the IEEE Engineering in medicine and biology society*, pp. 1223–1226, IEEE, 2009.

[35] P. Li, H. Liu, Y. Si, C. Li, F. Li, X. Zhu, X. Huang, Y. Zeng, D. Yao, Y. Zhang, *et al.*, "Eeg based emotion recognition by combining functional connectivity network and local activations," *IEEE Transactions on Biomedical Engineering*, vol. 66, no. 10, pp. 2869–2881, 2019.

[36] M. J. Mollakazemi, D. Biswal, and A. Patwardhan, "Target frequency band of cognition and tempo of music: Cardiac synchronous eeg," in *2018 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, pp. 696–700, IEEE, 2018.

[37] R. J. Davidson and J. Henriques, "Regional brain function in sadness and depression," *The neuropsychology of emotion*, pp. 269–297, 2000.

[38] B. D. Poole and P. A. Gable, "Affective motivational direction drives asymmetric frontal hemisphere activation," *Experimental brain research*, vol. 232, no. 7, pp. 2121–2130, 2014.

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

[40] B. Reuderink, C. Mühl, and M. Poel, "Valence, arousal and dominance in the eeg during game play," *International journal of autonomous and adaptive communications systems*, vol. 6, no. 1, pp. 45–62, 2013.

[41] S. Brown, M. J. Martinez, and L. M. Parsons, "Passive music listening spontaneously engages limbic and paralimbic systems," *Neuroreport*, vol. 15, no. 13, pp. 2033–2037, 2004.

[42] W. Trost, T. Ethofer, M. Zentner, and P. Vuilleumier, "Mapping aesthetic musical emotions in the brain," *Cerebral Cortex*, vol. 22, no. 12, pp. 2769–2783, 2012.

[43] E. R. Miranda and J. Castet, *Guide to brain-computer music interfacing*. Springer, 2014.

[44] I. Daly, D. Williams, A. Kirke, J. Weaver, A. Malik, F. Hwang, E. Miranda, and S. J. Nasuto, "Affective brain–computer music interfacing," *Journal of Neural Engineering*, vol. 13, no. 4, p. 046022, 2016.

[45] J. Heo, H. J. Baek, S. Hong, M. H. Chang, J. S. Lee, and K. S. Park, "Music and natural sounds in an auditory steady-state response based brain–computer interface to increase user acceptance," *Computers in Biology and Medicine*, vol. 84, pp. 45–52, 2017.

[46] S. K. Ehrlich, K. R. Agres, C. Guan, and G. Cheng, "A closed-loop, music-based brain-computer interface for emotion mediation," *PloS one*, vol. 14, no. 3, p. e0213516, 2019.

[47] M. A. Harley, *Space and spatialization in contemporary music: History and analysis, ideas and implementations*. PhD thesis, McGill Univeristy, 1994.

[48] E. Bates, *The Composition and Performance of Spatial Music*. PhD thesis, Trinity College Dublin, 2009.

[49] P. Miller, *Stockhausen and the Serial shaping of Space*. University of Rochester, Eastman School of Music, 2009.

[50] M. Brech and H. von Coler, "Aspects of space in luigi nono's prometeo and the use of the halaphon," in *Kompositionen für hörbaren Raum/Compositions for Audible Space*, pp. 193–204, transcript Verlag, 2015.

[51] N. Barrett, "Spatio-musical composition strategies," *Organised sound*, vol. 7, no. 3, pp. 313–323, 2002.

[52] R. Normandeau, "Timbre spatialisation: The medium is the space," *Organised Sound*, vol. 14, no. 3, pp. 277–285, 2009.

[53] M. A. Baalman, "Spatial composition techniques and sound spatialisation technologies," *Organised Sound*, vol. 15, no. 3, pp. 209–218, 2010.

[54] D. Smalley, "Space-form and the acousmatic image," *Organised sound*, vol. 12, no. 1, pp. 35–58, 2007.

[55] F. Otondo, "Contemporary trends in the use of space in electroacoustic music," *Organised Sound*, vol. 13, no. 1, pp. 77–81, 2008.

[56] J. Daniel, *Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia*. PhD thesis, University of Paris VI, France, 2000.

[57] C. Dimoulas, G. Kalliris, K. Avdelidis, and G. Papanikolaou, "Improved localization of sound sources using multi-band processing of ambisonic components," in *Audio Engineering Society Convention 126*, Audio Engineering Society, 2009.

[58] J. Breebaart, J. Engdegård, C. Falch, O. Hellmuth, J. Hilpert, A. Hoelzer, J. Koppens, W. Oomen, B. Resch, E. Schuijers, *et al.*, "Spatial audio object coding (saoc)-the upcoming mpeg standard on parametric object based audio coding," in *Audio Engineering Society Convention 124*, Audio Engineering Society, 2008.

Eleonora De Filippi is an advanced researcher at Eurecat, where she uses artificial intelligence and neuroimaging tools to study brain processes. Her background is in psychology and cognitive neuroscience and she is specialized in the study of brain mechanisms involved in new therapeutic approaches. Ms. De Filippi has recently obtained a Ph.D. in computational neuroscience and also worked for two years as a Neurofeedback therapist. She is the author of many articles on the topic of psychological well-being and of several scientific papers published in international journals. During her PhD, Ms. De Filippi was invited to the D'Or Institute in Brazil, where she led two projects related to the application of technology for clinical purposes. In 2022, she gave a talk at an international congress in Paris on the topic of emotion classification techniques for braincomputer interface application.

Timothy Schmele is a doctoral candidate at the University of Music in Karlsruhe, investigating the compositional implications of spatialized electronic music. He also works at Eurecat - Centro Tecnológico de Cataluña as a Senior resercher in the Multimedia Technologies Unit (MTU). His interest in electronic music are organic sounds native to the electronic realm, though he also composes with musical instruments that could be described as more traditional.

Finding a way to adequately describe these rich electronic sounds in an absolute way is one of the bigger goals he pursues in founding this unusual ensemble.

Arijit Nandi is pursuing his Ph.D. in Artificial Intelligence from Department of Computer Science at Universitat Politècnica de Catalunya, Barcelona, Spain. He also works at Eurecat - Centro Tecnológico de Cataluña as an AI researcher in the Big Data and Data Science Unit. His research interests are on Data Stream Mining, Affective Computing, Artificial Neural Network, E-Learning technologies etc. He can be reached at *jit.ari172@gmail.com*. Please visit also https://dblp.org/pid/241/6152

Adán Garriga Torres is the director of Eurecat's Audiovisual Technologies Unit. PhD in Physics from the University of Barcelona in 2004, he has taught at the University of Barcelona (2003-2004), the Pompeu Fabra University (2004-2008) and the University of the Balearic Islands (2008-2015). At Pompeu Fabra University he was deputy director of Computer Science Studies (2005-2008). He is the author of more than fifteen indexed scientific publications in the fields of mathematical physics and acoustics. He is currently responsible for the development of the Sfëar brand for the development of 3D audio technologies for music and virtual reality.

Alexandre Pereda Baños received the PhD degree in cognitive neuroscience from Trinity College Dublin, in 2008. He leads the Perception and Cognition research line within the Digital Humanities Department. He was a post-doctoral fellow in the Multisensory Research Group, Universitat Pompeu Fabra, in Barcelona. He was Torres a Quevedo fellow from 2012 to 2015. In 2010, he took charge of the PiC research group, with the aim of applying the methodologies of experimental psychology and cognitive neuroscience to the research of the interaction between users and new technologies, field commonly referred to as User Experience (UX). Since then, he has been involved in quite successful European research projects, mainly carrying the user evaluation tasks, while also disseminating the use of these methodologies in scientific meetings aimed at a variety of stakeholders in the ICTs field. He currently chairs the Working Group on User Experience metrics of the COST action 3DConTourNet. Finally, he has published his research in international journals and conferences, is a regular TCP on conferences on the field of HCI and collaborates with the University of Barcelona in the formation of practice students.