Investigating the Impact of Sound Angular Position on the Listener Affective State

Konstantinos Drossos, Andreas Floros, Andreas Giannakoulopoulos, and Nikolaos Kanellopoulos

Abstract—Emotion recognition from sound signals represents an emerging field of recent research. Although many existing works focus on emotion recognition from music, there seems to be a relative scarcity of research on emotion recognition from general sounds. One of the key characteristics of sound events is the sound source spatial position, i.e. the location of the source relatively to the acoustic receiver. Existing studies that aim to investigate the relation of the latter source placement and the elicited emotions are limited to distance, front and back spatial localization and/or specific emotional categories. In this paper we analytically investigate the effect of the source angular position on the listener's emotional state, modeled in the well-established valence/arousal affective space. Towards this aim, we have developed an annotated sound events dataset using binaural processed versions of the available International Affective Digitized Sound (IADS) sound events library. All subjective affective annotations were obtained using the Self Assessment Manikin (SAM) approach. Preliminary results obtained by processing these annotation scores are likely to indicate a systematic change in the listener affective state as the sound source angular position changes. This trend is more obvious when the sound source is located outside of the visible field of the listener.

Index Terms—Sound events, affective state, binaural processing, emotions, affective acoustic ecology

1 INTRODUCTION

Sound is a dominant component of human perception. It can emerge from various activities and sources, e.g. nature, human activities, machine operation and others. It is also a fundamental component for a wide number of applications in the area of entertainment (i.e. music, sound in video games, etc.) and communications (i.e. as speech/voice, alarms and sonification output in human-machine interfaces) [1]. Hearing does not require neither visible contact with the source nor a free path between the source and the receiver. When we hear sounds, we process them, appraise them and maybe decide to take some actions. In some or all of the above steps emotions are raised [2].

Music may be regarded as sound with well organized structure and form [3]. It is stated that it was originally developed as a technique to mimic and enhance the human voice and the conveyed emotions [4]. This fact, combined with the well established notion of emotion conveyance from music [3], [5], has led into a deep proliferation of music information retrieval (MIR) and music emotion recognition (MER) disciplines: MIR focuses on using emotion as a feature for content-based categorization and retrieval, with a typical example being the enhancement/proposed substitution of legacy music classification based on "Band-Artist-Year-Genre" scheme by clustering the musical data according to the elicited emotion and/or mood[1] [6]. MER on the other hand investigates the relation between music and the elicited listener's emotions [7]. Both raw and symbolic music data are considered, while various emotional models are employed [8], [9], [10].

However, music is only a portion of what we hear. There are numerous non-linguistic and non-musical sounds (termed as *Sound Events*) that construct the ambient audio environment [1]. Despite this significant role of sound events, only recently a limited number of studies have focused on emotion recognition from them [11]. Sound events are not simply signal representations. They inherently incorporate additional information related to many attributes of the sound source and its surroundings, such as it's spatial position or movement relative to the listener, the nature of the sound generation mechanism (e.g. impact noise), the volume and texture of the surroundings and others [12]. Typically, they emanate from human activities (e.g. walking, hand clapping, coughing), natural phenomena (e.g. sound of rain, wind, thunders), animals (e.g. dog barking, birds singing), machines (e.g. car noise, gunshots, soda fizz) or interaction between objects (e.g. impact, scraping) [1], [11]. They are also artificially synthesized within the scope of modern interactive audiovisual applications, including artificial ambient audio environments [1], soundscapes [13], virtual or augmented reality acoustic environments [14], video games and auditory-based Human-Computer Interaction, in a continuous attempt to enhance the immersion and the experience of the user [15].

- *K. Drossos and A. Floros are with the Department of Audiovisual Arts, Audiovisual Signal Processing Lab, Ionian University, Corfu 49100, Greece. E-mail: {kdrosos, floros}@ionio.gr.*
- *A. Giannakoulopoulos is with the Department of Audiovisual Arts, Interactive Arts Lab, Ionian University, Corfu, Greece. E-mail: agiannak@ionio.gr.*
- *N. Kanellopoulos is with the Department of Audiovisual Arts, Ionian University, Corfu, Greece. E-mail: kane@ionio.gr.*

Manuscript received 29 May 2014; revised 13 Dec. 2104; accepted 9 Jan. 2015. Date of publication 14 Jan. 2015; date of current version 3 Mar. 2015. Recommended for acceptance by S. Narayanan. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below. Digital Object Identifier no. 10.1109/TAFFC.2015.2392768

1. e.g. http://www.allmusic.com

The information conveyed by sound events represents essential data for the interaction between the listener and his environment. Such input and the resulting interaction is likely to engender emotions to the listener [16]. This particular fact has recently led to the concept of Affective Acoustic Ecology, in a generalized attempt to take into account the relation between the surrounding sound events and the listener's affective reactions within the legacy notion of Acoustic Ecology [1]. The key component of this definition is the sound event representation as a data structure that includes multiple information types: the sound waveform, the event time-duration, the semantic content and the sound source spatial position relative to the listener.

Under the above perspective, in this paper we aim to analytically investigate the potential impact of sound events source spatial position parameter on the human listener affective state. We particularly focus on the angle variation between the listener and the sound source under constant distance conditions, bridging the gap with existing works that only investigate the potential relation between elicited emotions and distance [17]. On the other hand, two studies already exist that explore the connection of the sound source angular position and the elicited emotion, considering only two generic spatial ranges at the front and at the back of the listener. The first one exclusively focuses on the emotion of "Fear" [18], while the second [19] reaches the conclusion that the placement of sound sources beyond the field of vision of the listener introduces increased arousal conditions. On the contrary, we hereby consider a full affective model and multiple angular positions on the horizontal plane. Towards this aim, we employ sound spatial rendering using binaural technology [20].

It is well-known that binaural processing results into a two-channel waveform that inherently encapsulates all the spatial information required for accurately modeling a sound event under the Affective Acoustic Ecology scope. Binaural signals may at some extent include multiple information related to a sound event nature, such as the reverberant field of the reproduction space [21]. Also, it has been reported that in reverberant rooms the source-receiver distance can be estimated with the utilization of binaural technology by employing signal statistics regarding the reverberant field [22]. As a consequence, they are utilized in a wide range of application fields, like hearing aids [23], dereverberation [24] or even stereo recordings enhancement [25]. Thus, within the scope of this work, they can be considered as a well-defined approach for modeling generalized sound events, with constant distance from the listener, under authentic and realistic terms.

The assessment performed in this work included a series of subjective evaluation experiments using an emotionally pre-annotated sound database with 167 sounds having multiple semantic content [26]. Binaural processing of the original sound data was realized using the KEMAR HRTF library [27]. For each original sound signal, five binaural versions were created, corresponding to angles equal to 0, 45, 90, 135 and 180 degrees. Finally, the participants' affective state was defined in terms of a 2-dimensional (2D) affective model, while the annotations were performed using the Self Assessment Manikin (SAM) method, which is suitable for affective state depiction [28].

The rest of the paper is organized as follows: Section 2 presents a brief overview of the emotion recognition process from sound events, including a concise description of the affective model employed and the corresponding emotion annotation methods followed. In Section 3 the exact experimental process is described, followed by the obtained results summarized in Section 4. The obtained results are analytically discussed in Section 5, while Section 6 concludes the work and depicts some issues and recommendations for future work.

2 Emotion Recognition from Sound

Emotion recognition from audio data can be regarded as an Artificial Intelligence/Pattern Recognition task [29] that requires the employment of a ground truth (or training), emotionally annotated dataset. This annotation involves emotion labels/classes appropriately chosen to closely match the employed affective model. The same data set is used for extracting appropriately selected signal technical features. Both emotional annotations and the extracted feature values are feeding a machine learning algorithm and a classification model is finally acquired. The annotations and the technical features of the testing dataset are finally used for the evaluation of the developed model, providing an estimation of the classification accuracy. The rest of this Section provides a brief report on well-established affective models and the corresponding annotation methods followed, alongside with an abridged overview of the emotion recognition from SEs.

2.1 Emotions Modeling

Modeling of emotions seems to be a widely debated and interdisciplinary aspect. Although many affective models exist [29], audio emotion recognition mainly employs a) discrete and b) continuous models [8], [11]. In the former category, emotions are represented as distinct terms expressed with verbal descriptions. Different words are assigned to different emotions. The most common models in this group are the basic emotions model and the list of adjectives [30]. The first one introduces a set of primary emotions, i.e. "Happiness", "Sadness", "Fear" and "Anger", where their combinations can lead to any other emotion [31]. Although this model was throughly questioned by [32], it appears that it is preferred for studying physiological responses to particular emotions, e.g. in neurogical research where there is a direct connection of particular brain's regions activity with specific emotions which lie in the basic emotions set [1]. For example, there are works reporting the connection of the activity in limbic/paralimbic system with emotions belonging in the basic emotions set [33], [34].

The list of adjectives modeling approach employs a set of distinct emotional conditions. For each of these, a set of synonym adjectives is provided. It was originally introduced by [35] and its difference from the basic emotions model is that it considers a set of affective or emotional states, each one described with a set of synonym adjectives. There are also studies that propose extensions of the list of adjectives model with the amount of the distinct emotional conditions reaching up to 13 [36].

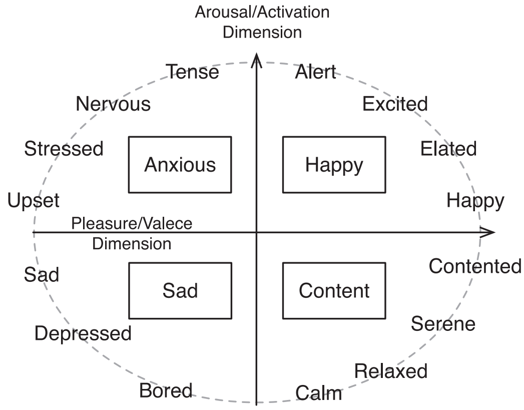


Fig. 1. Clustering and qualitatively mapping of VA space to verbal description of emotions, based on the circumplex model [41] and as proposed by [40].

On the contrary, continuous affective models consider emotions as a resultant condition from a set of basic affective states, e.g. arousal and valence. These are numerically mapped on orthogonal axes with continuous values. A specific emotion is expressed in terms of the unit vectors in the space defined by the above axes. Although the usual amount of axes is 2, leading to a 2D representation [37], there are studies reporting the usage of more dimensions with the preferred amount set to 3 [38]. Typical affective states used in a 3D space are valence, arousal and dominance [37], [38]. However, the 2D model can be considered advantageous over the 3D one, since it reduces complexity [39]. Correlation of the resulting values with the verbal descriptions of emotions is quantitively performed through clustering or qualitatively through depiction of points in the dimensional representation. The words assigned to different emotions are either used as labels for areas usually situated in different quadrants in the valence/arousal (VA) plane [40], for the former case, or as indicating points in the circumscribed circle of the 2D space [41], for the latter one. An illustration of both cases is in Fig. 1, after [40], [41].

Although there is a qualitative correlation of the values in the space of a continuous model with the verbal descriptions used in a discrete model, no quantitative one is reported. In the continuous models there are no numerical sets of values for the affective states which define areas (on the continuous model's space) that could be mapped to various individual emotions of a discrete model. An exception to this is the separation of a 2D space into quadrants, as shown in Fig. 1, using the median values of the two dimensions.

There are continuous 2D models depicting more than one discrete emotions in one quadrant [5], [8], [42]. But, there is no published report for actual values that can be used in order to quantitively cluster arousal and valence values and distinct individual emotions. Also, it has been previously shown that a categorization of emotion in the 2D plane is ambiguous [8]. This obscurity can be resolved by the continuous approach, assuming that each value in the 2D plane is a distinct emotional state and without any verbal description for the separate plane's values [8].

2.2 Emotion Annotation Methods

Affective annotations are performed through labeling an audio dataset using written verbal descriptors for emotions or values of affective states. In either case they form the classes required by the emotion recognition process for the classification task. The actual labels are assigned by human subjects. The exact choice of an annotation method is likely to be well connected with the chosen affective model. For example, when a discrete model is utilized, the annotation is performed with the use of verbal descriptions of emotions, i.e. "Happiness", "Joy", "Enlightment" etc [4]. But the usage of words seems to introduce an inconsistency between different approaches ascribed to the possible different words used for same or similar emotions and the particular perception of the words actually employed by the annotators [4].

In [28] a method for self assessment and annotation of emotion is introduced which exploits the capabilities of the dimensional models. This method uses figures of a manikin (i.e. the Self Assessment Manikin—SAM) in order to portray the discrete values of the affective states used in a dimensional model, i.e. valence, arousal and dominance. In its original form, five manikin figures are mapped to each valence and arousal dimensions along with the intermediate values between the available choices, leading to nine accessible values for each affective dimension.

2.3 Sound Events and Emotions

In general, research in the field of emotion recognition from SEs is rather sparse and loosely connected [11]. Attempting a brief literature overview, we can start from an early work [43] which aimed to define everyday listening through an alternative ecological approach. This work argued that human listening is actually the experience of hearing events and not sound waveforms. Additionally, it introduced a framework that maps various sound sources' attributes (including their spatial location relative to the listener) to the actual sound hearing experience. For example, it is stated that the resonating cavities of a source affect the spectral pattern of the sound, while the material of the source has impact on the frequency and complex effects on the amplitude of the perceived sound event.

The above work does not particularly focus on emotions conveyed by sound events, although it provides an alternative scope for defining sound as a listening experience that definitely involves emotions. Only lately research has particularly focused on the affective impact of sound as a generalized, perceived event, thus extending the relative existing investigations that deal exclusively with musical content and speech signals. Towards investigating this extended field, in [26] the International Affective Digitized Sound (IADS) emotional SEs database was presented, aiming to provide a ground-truth reference for future experimental studies. This database contains emotionally annotated general sounds which were marked up by a relative large number of people: the amount of annotations per sound was performed by an average of 100 people. The IADS dataset contains 167 sound events with a variety of semantic content ranging from every-day human activities, e.g. coughing, to animal sounds, e.g. dog barking, and usual or private activities, e.g. vomiting or sexual interjection. However, no spatial positioning information of the recorded sound sources is included. The annotation of the dataset was performed with the utilization of the 3D (arousal, valence and dominance) continuous emotional model and the SAM self-assessment method mentioned earlier.

Recognizing the potential (and probably significant) impact that the sound source position can result into the listener's affective domain, a recent research has focused on fear as the target emotion under investigation [18]. A set of 24 loudspeakers was used for sound spatialization at two positions/areas: in front and at the back of the listener. Only a limited number of humans participated in the subjective evaluation process. The results demonstrated that the participants perceived as most scary the sounds originating from outside of their field of vision, i.e. when the sound source was located at their back. Emotional responses with the variation of the source-listener distance and the lateral reflections imposed by the reproduction enclosure were studied in [44]. This work was also limited to fear. The results obtained portray an impact of the source location (and movement) to the listener's raised emotion. In particular, the listeners exhibited stronger emotional reactions and increased fear as the sound source was approaching. In addition, the room size was also found to affect the emotional responses in an analogous fashion, i.e. the larger the room, the stronger were the emotional reactions of the listeners.

Recently in [1], the authors proposed an extension of the legacy Acoustic Ecology concept in order to include the affective aspects of sound. This extension was termed Affective Acoustic Ecology. It's key-concept is the sound event, defined as a complex structure that encapsulates a number of legacy sound representation attributes, such as waveform, duration and sound source spatial positioning in the listener's direct acoustic environment. Based on this Affective Acoustic Ecology framework, an attempt for a concurrent recognition of both arousal and valence dimensions from the IADS dataset was also performed. The classification results obtained were not encouraging, being in the range of 50 percent for the arousal and close to 40 percent for the valence. More robust and encouraging results were obtained in [3], which focused exclusively on the arousal recognition from the rhythmic characteristics of generalized sound events. Again the IADS dataset was employed and the classification accuracy reached up to 89 percent. A set of 26 rhythm-related acoustic features was considered for classification purposes. The sound corpus consisted of seven different versions of the original IADS dataset. Each version was composed by the IADS waveforms windowed with different window time-lengths (ranging from 0.8 to 2.0 seconds).

In [11] emotion recognition from sound events was attempted, using a new dataset. It was performed in conjunction with the evaluation of features that can be utilized in emotion recognition for music and speech. The utilized dataset consisted of sounds from the FindSounds database [45]. These sounds were emotionally annotated by four subjects and the obtained mark ups were averaged using the evaluation weighted estimator [46]. The results showed that it is difficult to obtain a set of acoustic features that can be commonly used for emotion recognition for all sound events, music and speech cases. On the contrary, an inverse trend of features' variation in these three different application domains was reported, which shows that specific features have opposite effects in the above sound content categories.

From the above literature overview it becomes clear that there is a lack of research towards the investigation of the potential impact that the sound source positioning (or movement) may raise to the affective state of the human listener. Towards the systematic exploration of this area of research, in this paper we investigate how the human listener emotional state is affected by various sound sources placed at different locations around him. Specific questions are addressed, such as: are there any variations in the affective state of the listener as the source spatial position is changing? And how are these alterations related with the angle between the listener's reference axis and the sound source position? Through a series of subjective listening tests, we try to outline an analytical relation between the source's location modification and the listener's emotional state variation.

Since there is no available emotionally annotated and spatially-varied sound events database, a secondary, collateral contribution of this work is to offer such a ground-truth reference set. We have formed this database based on the original IADS dataset and have produced the sonic spatial representation using binaural processing. Due to the well-known, limited efficiency of binaural technology for representing moving sound sources [47], our investigations here do not consider moving sound sources. Instead, we aim to assess the affective state relative differentiation that is potentially imposed when a sound event occurs at different angles around the human listener. Additionally, in all test cases, the sound source distance is kept constant and equal to 1 meter. Hence, all sound event spatial positions are located on a circle with its center being the listener's head. Finally, without any significant impact to the generality we assume that a) sound propagation is performed in the open field, hence no close room-related acoustic phenomena occur and b) all sound sources are facing towards the listener; thus no directional properties of sound on the source side are applied.

3 Experimental Setup

We followed an experimental sequence that consists of the following parts: a) the creation of the binaural sound corpus, b) the subjective listening tests that derived the affective annotations for all sounds included in the binaural sound corpus, and finally c) appropriate pre-processing of the original subjective ratings, in order to derive alternative meaningful representations that can be further analyzed to obtain meaningful results. These parts are analytically described in the next subsections.

3.1 Binaural Sound Corpus Creation

The utilized sound corpus was created using the IADS sound events library [26]. IADS was chosen instead of another available dataset [30] due to the larger amount of available ratings per sound. It consists of 167 monophonic sounds (denoted here as $s(i)$, with $i \in [1, 167]$), with various semantic

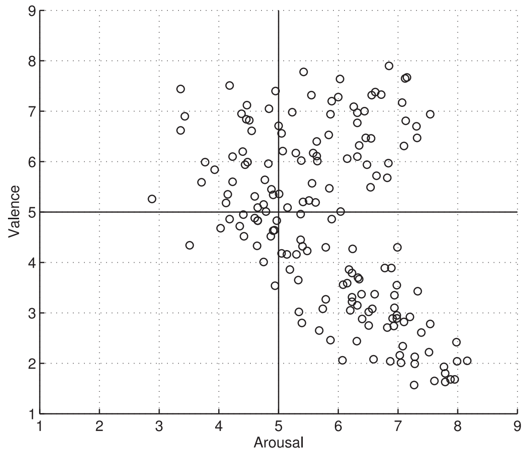


Fig. 2. IADS mean annotation values.

content. The dataset also includes the mean and standard deviation values for emotion annotations from 100 participants for each sound waveform. The averages of these annotation ratings per $s(i)$ in the VA space are graphically presented in Fig. 2. Clearly, these mean values are assembled in three quadrants of the VA space. As it is stated in [26], this is due to the fact that it is unlikely for a human to hear a sound that he/she does not like and not feel aroused.

All $s(i)$ waveforms were binaurally processed using the KEMAR HRTF library [27]. The set of the spatial angles considered is $\theta(k) = \{0^o,\ 45^o,\ 90^o,\ 135^o,\ 180^o\}$, where $k \in [1, 5]$ is the corresponding spatial index. Taking into account the symmetry of HRTFs on the horizontal plane, we do not consider angles larger than 180 degree. It must be noted here that although this assumption is valid from a technical point of view, psychological studies have already reported a clear left-ear advantage for identifying emotional qualities [48]. Under this perspective, our approach cannot be generalized to cover the overall horizontal plane. Hence, binaural processing produced five binaural versions $s_b(i, k)$ of the ith original IADS waveform. The derived sound corpus was finally organized in terms of five angular subsets $S_a(k)$, each containing all 167 $s_b(i, k)$ binaural waveforms, i.e. $S_a(k) = \{s_b(1, k),\ s_b(2, k),\ \ldots,\ s_b(167, k)\}$. Finally, all binaural signals were normalized to an average of -4.44 relative to full scale (dBFS) in order to avoid further sound clipping artifacts which could be introduced by the amplification adjustments performed by the human subjects. Each $S_a(k)$ subset was stored in the database organized in tuples $t_a(k)$ formed as:

$$t_a(k) = (s_b(i, k), c(i, k)), \tag{1}$$

where c is an integer counter defined for controlling the selection of the $s_b(i, k)$ sounds to be reproduced during the experimental process (the details of this control mechanism are provided next in Section 3.2).

3.2 Corpus Annotation

The sequence of the annotation tests was performed remotely through a web-based platform,[2] thus introducing a number of advantages:

TABLE 1
Participants Country Origin Distribution

Country	(%)	Country	(%)
Greece	18.92%	France	5.85%
The Netherlands	3.28%	Finland	5.85%
Australia	3.28%	New Zealand	3.28%
United Kingdom	12.27%	Denmark	4.56%
Canada	4.56%	Germany	3.28%
Unknown (.com and .org domains)		34.87%	

- simultaneous access of different participants reduces the time required for successfully organizing the experiments
- each participant is able to perform the experiment at his convenience, since no time and location restrictions apply. This fact also diminishes the undesired effect of the user fatigue, as well as any negative feelings raised by the unfamiliar environment of a typical laboratory space
- access to distant participants from all over the world is feasible. This is an important feature that can lessen the substantial effect of cultural distinctiveness on the final subjective ratings. To ensure such participation, the conduction of the experiment was advertised through international and widely-known lists of electronic mails, such as the auditory-list.[3] Hence, although the instructions were written only in two languages, the usage of English language can be considered sufficient since the majority of mails in the employed electronic lists are written in English. Table 1 summarizes the distribution of the geographical origin of the participants, as it was obtained by extracting the internationalized domain names from the submitted e-mail addresses used for login purposes.

On the other hand, this remote experimental execution also imposes some risks. Headphone equalization that is typically required for artifact-free binaural reproduction is not feasible. More importantly, it is not verified whether the subjects were using headphones, although they were strongly requested to do so. The quality of the user playback consumer equipment may also influence the sound reproduction quality, mainly in terms of induced noise and harmonic distortion. A series of additional potential risks originates from the non-laboratory experimental conditions: For example, multiple subjects participation and external user distractions imposed by other people or noisy conditions may render the participants inattentive and acoustically disturbed; thus, these conditions may consequently affect the validity of the recorded annotation mark-ups. We have to note though that such practical but important issues are subject to the trust that should be ascribed to human volunteers participating in a subjective test, thus overcoming possible arguments related to the participants' answering truthfulness. For example, even in a fully-controlled laboratory environment, the outcome of a subjective evaluation can be equivalently affected by the degree of responsibility,

2. http://www.audemo.eu

3. http://www.auditory.org

faithfulness and truthfulness of the participants, or even the user fatigue that may be induced during in-laboratory subjective assessments [49].

For avoiding multiple participations of the same subjects we employed a login by e-mail mechanism combined with a user registration prerequisite. Under this mechanism, each participant had to create an individual account, then log into the platform with his own credentials and perform the experiment. Although this technique can allow the creation of multiple accounts (and thus introduce noise in the obtained annotations) we have tried to reduce this possibility by rejecting the results obtained by users with the same (or very similar) registered profiles.

A total number of 215 valid participants responded providing a set of 3,905 annotations. The actual listening tests were organized in two stages. During the first one, a short introductory text is presented summarizing the usage of the platform and strongly prompting the participant to use headphones. Next, a 1 kHz tone is reproduced and the subject is asked to adjust the reproduction gain for his headphones to a comfortable, non-annoying level. Following the level adjustment process, an informative video[4] is shown, demonstrating the details of the listening tests and explaining the SAM scores. It is also clearly stated that there is no correct or wrong answer.

For each human subject, a playlist is automatically generated containing 15 $s_b(i,k)$ soundtracks. Three sounds from each angular subset are chosen randomly according to the condition:

$$I(k) = \min_{c(i,k)}(t_a(k)), \tag{2}$$

where $I(k)$ is a tuple of the form of $t_a(k)$. Upon the selection of an $I(k)$ the corresponding $s_b(i,k)$ is selected for reproduction and the $c(i,k)$ element is increased by one, identifying the amount of participation for the corresponding $s_b(i,k)$ in a playlist. This counting is chosen in order to decrease the probabilities for a $s_b(i,k)$ to be selected repeatedly from the random collection process.

During the second part of the experiment the subjects are able to sequentially listen to the sound tracks of the corresponding playlist. Then, the subjective SAM rating options are shown and the subjects are asked to select their own arousal and valence ratings. The selected scores are stored in a database along with the corresponding playlists.

3.3 Subjective Ratings Post-Processing

Due to the remote execution of the experiment, it is observed that some of the participants did not complete it. Thus, their ratings have to be excluded from the final results derivation. This leads to a total of 50 fully annotated sound events $s(i')$ (or a total of 250 $s_b(i',k)$ annotated sounds). For reasons of simplicity, we use a separate index $i' \in [1,50]$ for these sounds, mapped to the i-index of the original IADS $s(i)$ sounds using the equation:

$$i = i' + f(i), \tag{3}$$

4. https://www.youtube.com/watch?v=COhmf206VIg

TABLE 2
The Mean and σ Values for the Amount of Ratings that Each Sound Received Along with the Amounts of Total and per Angle Obtained Annotations

	Angles				
Amount of ratings	**0°**	**45°**	**90°**	**135°**	**180°**
Mean	9.54	10.60	9.84	10.36	9.92
σ	3.11	3.72	3.84	3.02	3.90
				Average mean	10.05
				Average σ	3.52

where $f(i)$ is an integer representing the total number of IADS sounds that are excluded up to the ith sound. In the final dataset a total of $2,946$ subjective annotations are included.

Following the definition of the angular subsets $S_a(k)$, we can hereby define new angular subsets $S'_a(k)$ containing all the binaural sounds $s_b(i',k)$ ($i' \in [1,50]$) that correspond to the same sound source angle $\theta(k)$. These notations are useful for analytically modeling the metrics employed for presenting the results in the following section.

4 Results

In this Section, a draft presentation of the overall results structure and values obtained is included. Their analysis and subjective interpretation is provided next in Section 5.

Table 2 shows the mean and standard deviation (σ) values for the amount of ratings that each $s_b(i',k)$ sound received from the subjective annotators separately for each angle $\theta(k)$ and in total. Furthermore, the obtained subjective rates for the binaural signals $s_b(i',1)$ are compared against the rates of the corresponding $s(i'+f(i))$ sound events of the original IADS dataset. In particular, if we assume that $\overline{A_a}[x]$ and $\overline{A_v}[x]$ denotes the mean annotations' value obtained for a signal x and for the arousal and valence dimensions respectively, then the comparison above is performed in terms of the differences:

$$D_a(i') = \overline{A_a}[s(i'+f(i))] - \overline{A_a}[s_b(i',1),] \tag{4}$$

$$D_v(i') = \overline{A_v}[s(i'+f(i))] - \overline{A_v}[s_b(i',1)]. \tag{5}$$

The $D_a(i')$ and $D_v(i')$ values are presented in Table 3, together with their mean and standard deviation (σ) values. In addition, the mean and standard deviation values of the corresponding absolute differences values are shown.

Figs. 3a, 3b, 3c, 3d, and 3e illustrate the scatter plots of the mean arousal and valence annotation values for all angular subsets $S'_a(k)$. Hence, each point in these graphs is defined by the cartesian coordinates $(a(i',k), v(i',k))$ equal to corresponding mean arousal and valence values, respectively.

Another interesting metric is the variation of the $\overline{A_v}[s_b(i',k)]$ and $\overline{A_a}[s_b(i',k)]$ mean values with the spatial angle $\theta(k)$. The above pairs of values (termed here as VA vectors) for consecutive/neighboring pairs of spatial angles $\theta(k)$ can be expressed as:

$$\overrightarrow{va}(i',k) = \left(\overline{A_a^d}[s_b(i\,k)]\right)\overrightarrow{a} + \left(\overline{A_v^d}[s_b(i\,k)]\right)\overrightarrow{v}, \tag{6}$$

TABLE 3
Difference of the Mean Values for Arousal and Valence Ratings of the Final Binaural and the Original IADS Dataset

i'	Arousal Difference ($D_a(i')$)	Valence Difference ($D_v(i')$)	i'	Arousal Difference ($D_a(i')$)	Valence Difference ($D_v(i')$)
1	-2.09	-0.75	26	0.58	0.23
2	2.25	1.67	27	0.14	1.55
3	-0.54	1.07	28	-0.72	1.21
4	-1.04	0.84	29	0.52	1.31
5	0.70	-0.97	30	-0.07	1.02
6	0.77	0.35	31	-1.99	1.75
7	2.53	-0.81	32	0.22	1.46
8	-1.24	0.55	33	0.22	-0.37
9	1.65	-1.63	34	0.26	0.46
10	1.18	-0.17	35	1.37	-0.59
11	0.05	0.14	36	1.99	-0.68
12	-0.11	0.95	37	1.11	-0.54
13	0.19	0.13	38	-2.83	1.79
14	0.88	0.01	39	-0.33	0.31
15	0.43	-0.54	40	0.72	2.13
16	1.48	0.72	41	-1.42	0.61
17	-0.83	-0.42	42	-0.20	0.71
18	-0.06	0.04	43	-1.45	1.86
19	0.30	0.08	44	0.00	1.99
20	-0.58	0.79	45	2.09	0.08
21	0.70	-1.08	46	-0.49	-0.30
22	1.58	-0.58	47	0.58	1.88
23	0.52	-0.25	48	-0.05	1.03
24	0.72	-1.79	49	-1.14	1.76
25	0.10	0.24	50	0.25	-0.01
Mean arousal:		0.18	Mean valence:		0.38
σ of arousal:		1.12	σ of valence:		0.98
Absolute mean of arousal:		0.87	Absolute mean of valence:		0.84
Absolute σ of arousal:		0.73	Absolute σ of valence:		0.62

where $\overrightarrow{a}$ and $\overrightarrow{v}$ are the unit vectors for the arousal and valence dimension respectively, $k \in [2, 5]$ and $\overline{A_a^d}[s_b(i\,k)]$ and $\overline{A_v^d}[s_b(i\,k)]$ are defined as:

$$\overline{A_a^d}[s_b(i\,k)] = \overline{A_a}[s_b(i', k)] - \overline{A_a}[s_b(i', k-1)], \tag{7}$$

$$\overline{A_v^d}[s_b(i\,k)] = \overline{A_v}[s_b(i', k)] - \overline{A_v}[s_b(i', k-1)]. \tag{8}$$

The angles of the above vectors (calculated relative to the arousal unit vector $\overrightarrow{a}$ counterclockwise) as well as their magnitudes are calculated using the equations (6) and:

$$\phi(i', k) = \arctan\left(\frac{\overline{A_a^d}[s_b(i\,k)]}{\overline{A_v^d}[s_b(i\,k)]}\right). \tag{9}$$

The resulting values of $\overrightarrow{va}(i', k)$ vector angles (in degrees) and magnitudes are shown in Table 4.

5 Discussion

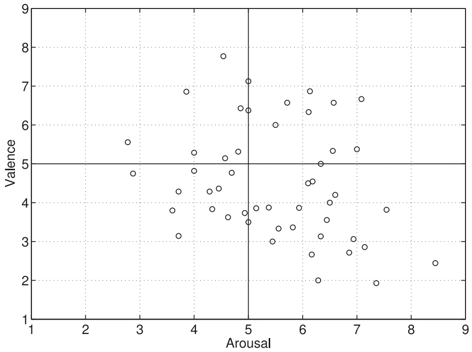
A detailed discussion on the results presented in the previous Section may be organized in terms of the following three assessment attempts:

- The investigation of the accuracy of the subjective results performed by evaluating the differences between the annotations provided with the IADS dataset and the ones obtained here. Since the IADS sounds are monophonic, the best approach for performing the above comparison is to consider the binaural sounds for $\theta(1) = 0^{\text{o}}$. This also allows to investigate whether the conveyance of any spatial information can alter the measured ratings for the above case.
- The analytic assessment of the effect of the variation of the sound source angular position on the affective state of the listener in terms of the annotated VA values.
- The particular interpretation of the above VA values for sound events that have music content. Due to widespread and accustomed adaptation of music listening in every day life, the effect of the musical source's spatial position to the listener's affective state is likely to reveal noteworthy results.

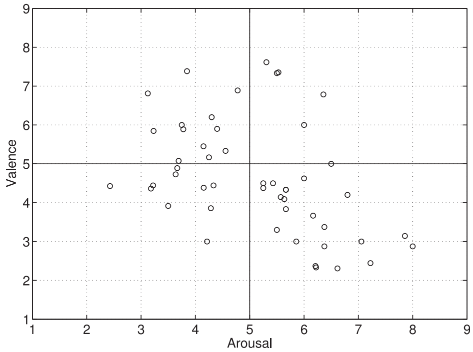
These assessment attempts are analytically described next.

5.1 VA Values Comparison to the IADS Dataset

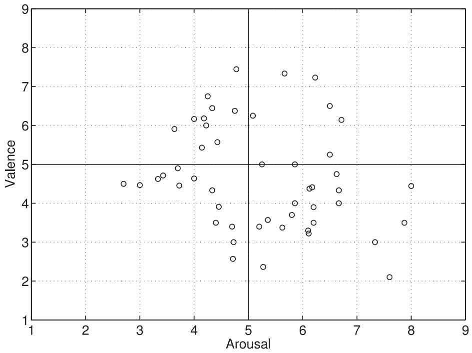
Focusing on the $D_a(i')$ values presented in Table 3, the mean and absolute-mean difference of the subjective ratings obtained in this work for $k = 1$ and the ones provided with the IADS dataset is below unity, with values equal to 0.18 and 0.87 respectively. The same fact is observed for the standard deviation of the above differences, which is marginally equal or less than 1. Since the SAM subjective ratings were defined as integers in both cases, difference values less than one do not imply actual differentiation in the annotation scores. Only values in the range of $[1\;\;2)$ indicate a difference of one intermediate SAM rating step, while greater ratings in the $[2\;\;3)$ interval denote that the ratings' distance corresponds to one actual SAM figure. A number of sound events in the set $s_b(i', 1)$ though demonstrate $D_a(i')$ and $D_v(i')$ values that exceed the unity margin. But, given the small amount of ratings acquired for each sound event in $S_a'(1)$ (presented in Table 2), one may assume that the obtained mean values of annotations are not significantly different from the IADS ones. Hence, before proceeding to any further analysis, a series of hypothesis tests is carried out for each sound event that exhibits difference above the unity margin. The list of the p values derived from the hypothesis tests for these sounds is presented in Table 5. For these tests, the null hypothesis is defined as: the procured mean values of arousal and valence are not different from the IADS dataset, with the alternative stating that there is such a difference. Since the full annotations' distribution of the IADS dataset is not available but, instead, only the corresponding mean and σ values are provided, Z-Tests are applied. The results of these tests show that 10 out of 18 (55.55 percent) sound events have significantly different values for each affective dimension, with significance level (p value) set to 0.01. If p is raised to 0.05 then there are 13 out of 18 (72.22 percent) sound events for each category that portray a significant difference at their ratings and if further increment of p is used, i.e. 0.1, then there are 14 out of 18 (77.77 percent) for arousal and 15 out of 18 (83.33 percent) for valence. Thus, the obtained ratings of sound events that



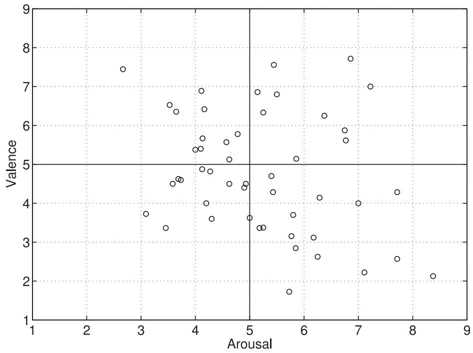
(a) VA mean values for $S'_a(1)$ subset ($\theta(k)$=0°)



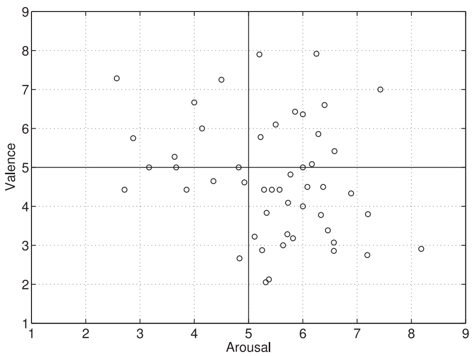
(b) VA mean values for $S'_a(2)$ subset ($\theta(k)$=45°)



(c) VA mean values for $S'_a(3)$ subset ($\theta(k)$=90°)



(d) VA mean values for $S'_a(4)$ subset ($\theta(k)$=135°)



(e) VA mean values for $S'_a(5)$ subset ($\theta(k)$=180°)

Fig. 3. VA mean values for $S'_a(k)$ and $k \in [1,\, 5]$.

have $D_a(i')$ and $D_v(i')$ values above the unity margin can be considered different from the IADS dataset.

From these, only four (for i'=1, 2, 7 and 38) are annotated with the next (or previous) SAM arousal figure, whereas the remaining are annotated with the immediate next or previous rate. In addition, the semantic content of sound events in Table 5 cannot be systematically grouped, since it ranges from animals and erotic actions to crashes and musical content. Thus, no exact relation of the semantic content can be indicated for explaining this divergence of annotation values. In overall, it can be concluded that, no significant variations are observed on the arousal affective dimension annotation between the two tests under comparison.

Focusing on valence, the mean of $D_v(i')$ and $|D_v(i')|$ values are 0.38 and 0.84 respectively. Eighteen sound events exhibit a value of $|D_v(i')| \geq 1$, while only one imposes a difference value above 2. These are listed in Table 5 along with their semantic content description. Clearly, only one of them is annotated at a distance of one SAM figure and an intermediate state (compared to the original IADS sound) and 17 at a distance of one intermediate state. Moreover, the semantic content of these 17 sound events appears non-correlated; hence, no indication of a potential impact of the semantic content to the valence rating can be concluded.

Summarizing the above outcomes, the arousal and valence ratings demonstrated a relative coherence between the sound events located at $\theta(1) = 0^{\text{o}}$ and the original IADS dataset. The results of the performed hypothesis tests indicate that the obtained values significantly vary from the IADS ones and thus they can be considered as not emerging from potential statistical deviations imposed by the obtained amount of annotations per sound event. Since the only

TABLE 4
VA Vectors Angles and Magnitudes versus Applied Angular Transitions

	Angular Transitions							
	$0^o \rightarrow 45^o$		$45^o \rightarrow 90^o$		$90^o \rightarrow 135^o$		$135^o \rightarrow 180^o$	
i' **index**	$\phi(i',2)$	$\lvert\overrightarrow{va}(i',2)\rvert$	$\phi(i',3)$	$\lvert\overrightarrow{va}(i',3)\rvert$	$\phi(i',4)$	$\lvert\overrightarrow{va}(i',4)\rvert$	$\phi(i',5)$	$\lvert\overrightarrow{va}(i',5)\rvert$
1	−150.89	1.80	−171.35	1.11	69.78	0.84	47.54	0.88
2	85.14	2.10	−79.94	1.27	82.54	0.77	164.05	1.27
3	144.58	1.31	−7.99	0.74	−137.54	1.33	91.83	1.57
4	88.75	1.35	−112.02	1.19	39.54	2.30	−112.31	1.20
5	126.05	1.18	−18.71	0.60	14.29	1.56	176.19	2.15
6	−132.75	1.15	173.93	0.53	0.00	1.92	141.99	1.25
7	136.90	1.04	−126.03	0.45	49.14	0.87	−17.44	1.86
8	174.68	2.60	−36.43	1.09	142.79	0.59	141.51	1.22
9	−78.24	1.44	82.52	0.55	−104.93	0.34	5.06	2.16
10	179.06	1.36	26.57	0.99	135.00	0.31	24.15	0.88
11	−148.39	1.27	45.00	0.71	34.70	0.88	−111.97	1.71
12	61.46	0.51	−60.95	0.74	−82.41	0.27	58.90	1.31
13	−115.69	1.40	93.84	1.02	95.71	0.71	−15.54	1.96
14	−114.02	1.94	56.31	0.90	−146.63	0.72	1.96	2.21
15	143.51	1.44	−71.57	1.36	56.31	0.97	−132.27	0.21
16	105.48	0.80	78.50	1.00	−106.26	0.42	−14.62	0.28
17	−150.36	1.06	133.78	1.48	26.57	0.67	−87.99	0.52
18	48.37	1.51	−136.70	1.79	106.77	1.48	−42.88	1.45
19	121.15	1.52	−103.64	0.96	−54.46	0.78	141.67	0.52
20	175.36	0.26	118.07	0.27	0.00	0.14	−66.64	0.62
21	−60.95	1.14	150.86	1.44	−93.22	0.55	−5.87	0.81
22	−23.20	1.90	75.47	1.16	−125.42	1.42	106.70	0.46
23	159.01	1.23	29.36	1.29	−105.95	0.81	81.76	0.53
24	−145.95	0.64	22.75	0.60	−55.30	0.28	−82.57	0.37
25	−144.46	0.66	−57.69	0.56	136.23	1.00	−6.30	1.56
26	175.96	1.11	−35.33	1.49	122.35	1.33	93.18	0.38
27	55.01	0.55	180.00	0.64	34.01	1.10	−141.34	0.99
28	104.04	0.52	90.00	0.62	−120.26	0.69	−19.98	1.17
29	14.38	0.64	169.77	1.02	−35.80	1.07	−15.22	2.34
30	168.02	0.34	91.70	1.02	−64.71	1.14	172.57	0.97
31	152.88	0.84	107.65	0.58	99.46	1.52	−71.39	2.59
32	−135.00	0.94	−60.26	0.38	121.26	2.08	−38.66	2.05
33	−120.74	0.47	134.88	1.78	171.47	0.36	−75.96	1.55
34	80.54	0.65	−71.57	1.19	−25.02	0.61	129.92	1.95
35	−28.07	1.89	142.68	1.41	−77.47	0.69	148.57	0.72
36	162.65	2.00	−31.86	2.71	164.05	1.32	55.01	1.57
37	6.48	1.85	169.58	1.92	−22.09	1.73	172.98	2.48
38	171.87	0.94	2.73	2.34	174.39	2.61	−61.75	0.69
39	−176.69	1.08	−27.31	1.96	152.41	2.73	−20.03	1.48
40	1.17	0.56	144.46	0.22	134.47	1.10	18.12	0.69
41	−134.44	1.03	129.62	1.90	−33.69	1.56	−60.26	1.07
42	160.35	0.83	48.61	0.77	175.96	1.25	25.23	1.37
43	151.78	2.58	−5.57	1.72	−70.02	1.46	103.84	0.81
44	171.23	3.36	28.49	1.25	−170.54	0.17	−93.27	0.42
45	−1.71	0.76	−71.57	0.29	−131.19	0.97	118.24	0.80
46	−127.12	1.16	−101.51	0.79	55.01	1.74	148.57	0.75
47	59.71	2.12	3.67	0.93	180.00	2.11	−30.26	1.54
48	76.07	1.28	−114.21	1.36	29.54	0.86	−90.00	0.70
49	102.76	0.78	−17.20	2.20	−170.54	0.65	−120.96	0.83
50	−20.62	2.06	135.00	1.68	−45.00	0.47	108.43	0.35

difference applied between the sound waveforms of the above datasets is binaural processing, this fact is likely to imply that the spatial information attached on the IADS dataset introduces a slight variation of listener's arousal and valence. On a first investigation, this variation seems to indicate that by considering sound source spatial positioning, we can adjust the affective ratings of the IADS sounds to a more realistic perspective: in real-world, all sound events do demonstrate such information; however the original IADS dataset does not include it. Thus, by incorporating the previously excluded—but important—sound spatial position information to the dataset possibly leads to more realistic, revised annotations.

Through a more detailed view though and taking into account the spread of the sound sources, two groups can be identified within the $S'_a(1)$ dataset. One with sound events emanated from point-like sources and the second containing sound events which demonstrate pure ambient

TABLE 5
Semantic Content and p Values from Z-Tests of the Sound Events that Exhibit $|D_a(i')| > 1$ and $|D_v(i')| > 1$

i'	Semantic cont.	$D_a(i')$	p	i'	Semantic cont.	$D_v(i')$	p
	$\|D_a(i')\| > 1$				$\|D_v(i')\| > 1$		
1	Cat	−2.09	0.00	2	Dog	1.67	0.02
2	Dog	2.25	0.00	3	Baby	1.07	0.18
4	Kids 1	−1.04	0.23	9	Tropical	−1.63	0.06
7	Rattlesnake	2.53	0.00	21	Fight 2	−1.08	0.26
8	Robin	−1.24	0.15	24	Creep	−1.79	0.00
9	Tropical	1.65	0.05	27	Horse race	1.55	0.01
10	Erotic fem 1	1.18	0.01	28	Paint	1.21	0.00
16	Couple sneeze	1.48	0.06	29	Sink	1.31	0.01
22	Fight 3	1.58	0.01	30	Rain 1	1.02	0.05
31	Helicopter	−1.99	0.00	31	Helicopter	1.75	0.00
35	Plane crash	1.37	0.03	32	Countdown	1.46	0.01
36	Engine failure	1.99	0.01	38	Thunderstorm	1.79	0.07
37	Bike wreck	1.11	0.15	40	Phone 1	2.13	0.00
38	Thunderstorm	−2.83	0.00	43	Alarm	1.86	0.03
41	Clock	−1.42	0.01	44	Slotmachine	1.99	0.00
43	Alarm	−1.45	0.02	47	Harp	1.88	0.00
45	Walking	2.09	0.00	48	Bach	1.03	0.07
49	Choir	−1.14	0.24	49	Choir	1.76	0.01
	Amount of SEs		18		Amount of SEs		18

characteristics. In [44] it is reported that the perceived spread of the sound source influences the elicited emotion, and in particular "Fear". Focusing on the utilized dataset, from Table 3 it can be seen that narrowing the perceived width of the source, i.e. applying binaural processing and reproducing the corresponding sound event through headphones and not from a set of loudspeakers, can result in a reduced mean arousal and increased valence. This arousal variation conforms well with the results obtained in [44], especially for the sound events that emanate from a localized point sound source (such as a dog or a rattle snake with i' index equal to 2 and 7 respectively). On the other hand, there are sound events, like rain ($i' = 30$), tropical forest ($i' = 9$) and thunderstorm ($i' = 38$) which are the only ones among the $S'_a(k)$ dataset that exhibit pure ambient sound characteristics. Thus, although the affective annotations made here can be considered to adjust the corresponding ratings obtained from the original IADS dataset, for the three ambient-like sound events described above, this adjustment cannot be considered as a realistic one.

But does the above subjective annotation adjustment result into variations in the affective state of the listener? This question is rather important, since it designates the significance of the obtained subjective annotation adjustments for $\theta(1) = 0^{\circ}$. This question can be addressed by mapping the above measured annotation differences to the raised emotions, as they are modeled in the VA space. Hence, a further quantitative examination can be performed that associates the derived arousal and valence differences shown in Table 5 with possible emotional variations implied by the valence/arousal affective model. Specifically, following the fact analyzed in Section 2.1 that existing MER and MIR works map only the quadrants of the VA space to different emotions, this investigation regards whether the measured arousal and valence values, for 0 degree, lay within the same VA space quadrant with the IADS ones.

Fig. 4 illustrates these values obtained in the cases of the sound events listed in Table 5. From this Figure it can be outlined that for 0 degree only four sound events (with i' = 1, 2, 28 and 38) demonstrate values of arousal and valence that are in different quadrants from the corresponding ones in the IADS dataset. Furthermore, there are seven sound events (for i' equal to 4, 7, 8, 9, 16, 22 and 36) whose arousal values are in different quadrants compared to $\overline{A_a}[s(i'+f(i))]$ and five more (with i' equal to 27, 29, 30, 31 and 40) whose valence values lay within a different quadrant compared to $\overline{A_v}[s(i' + f(i))]$. Thus, from a total of 30 unique sound events in Table 5, 16 are found to alter the listener's emotional state.

Abridging the above outcomes, the encompass of spatial information to the utilized sound events results into slightly altered values for arousal and valence. Although the new values could be considered as more realistic and representative of a real-world sound event representation, the emotional state of the listener is not significantly

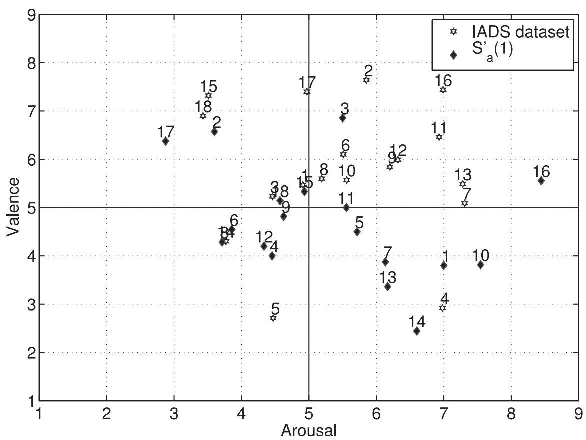


Fig. 4. Scatter plot for the VA values of the original IADS and $S'_a(1)$ datasets.

affected for the majority of the sound events in the employed dataset.

5.2 VA Variation with Source Location Alteration

A close examination of the VA vectors magnitudes and angles presented in Table 4 can reveal three different classes of sound events, depending on the arousal and valence variation trends observed as $\theta(k)$ increases:

1) Class 1 (C_1), which includes sound events rated with higher arousal and lower valence values
2) Class 2 (C_2), which includes sound events rated with lower arousal and higher valence values
3) Class 3 (C_3), which includes all sound events that do not fall into the above two cases

Table 6 lists the sound events included in C_1 and C_2 for all spatial transitions followed by their VA vector parameters. Obviously, there are sound events which cannot be assigned exclusively to either C_1 or C_2 classes as k increases, since they interchangeably move from one class to the other. In contrast, there are sound events that do belong to C_1 or C_2, exhibiting a systemic alteration of the received emotional annotations. Fig. 5 illustrates the number of sound events included in C_1 or C_2 and their union $C_1 + C_2$ (expressed as the percentage relative to the total amount of the IADS waveforms included in the binaural sound corpus) as the sound source is moving from the front to the back side of the listener. Clearly, there is a trend of population increment for classes C_1 and C_2 as $\theta(k)$ increases. In addition, there is likely no indication that the semantic content has any effect on the resulting values since in each of the classes the corresponding semantic contents are fairly different.

The statistics of the VA vectors' absolute angle and magnitude for classes C_1 and C_2 for all $\theta(k)$ cases are shown in Table 7. For all angular transitions, most sound events alter $\overline{A_a^d}[s_b(i',k)]$ and $\overline{A_v^d}[s_b(i',k)]$ almost inversely analogously, i.e. a decrement on one dimension (e.g. arousal or valence) is analogous to the increase to the other. Thus, as the source moves towards the boundaries, or even outside of the listener's field of view both arousal and valence component demonstrate a rather systematical change. This indication is strengthen when the source moves outside of the listener's field of view, i.e. to 135 and 180 degree, where the percentage of sound events allocated to classes 1 and 2 reaches up to 70 percent.

But why this assignment to classes is important? And why is it examined here? Considering the distribution of verbal description of emotions proposed by the circumplex model of affect, certain areas of the VA space can be mapped to specific emotions or group of emotions [40], [41]. Hence, the systematic differentiation of the $\overrightarrow{va}(i',k)$ vectors can reveal a tendency for the alteration of the annotated emotions for sound events in C_1 and C_2 for the aforementioned particular emotions/group of emotions. Moreover, following [26] and as it is shown in Fig. 2, it is highly unlike for a listener to experience low valence (i.e. highly dislike) without feeling activated (i.e. high arousal).

Taking into account the above analysis, the members of C_1 and C_2 demonstrate a clear $\overrightarrow{va}(i',k)$ placement and direction trend towards specific areas in the VA space. Sound events in C_1 exhibit the likelihood to have new $\overrightarrow{va}(i',k)$ values in the second quadrant or to elicit emotion that is located counterclockwise comparatively to the previous ones. On the other hand, the $\overrightarrow{va}(i',k)$ vectors for the C_2 class members tend to move towards the fourth quadrant, thus raising emotions located clockwise relatively to the previous ones.

More specifically, considering sound events in C_1 and for all k cases, an increment of the $\overline{A_a^d}[s_b(i',k)]$ value is combined with an equal decrement in $\overline{A_v^d}[s_b(i',k)]$. This is in conformance with the findings in [44] and can be analyzed further: As k increases, the maximum, minimum and mean absolute angle values also increase. This fact demonstrates that as the source is moving from the front of the listener towards lateral positions, the receiver's valence exhibits an increased effect of alteration up to the maxima of the angular transition of $45 \rightarrow 90$ degree. From that maxima and further (i.e. for greater k values), the alteration in the affective state of the listener is more focused on his arousal. Due to the lack of a global quantitively association for the values in the VA space with verbal descriptions of emotions, no analytical relation between the sound events in C_1 and the resulting emotions can be established. Nevertheless, the combination of vectors' magnitudes and absolute angle values can reveal the variance with angular transitions of the annotation using the SAM method. Based on the mean values of $|\phi(i',k)|$ and $|\overrightarrow{va}(i'k)|$ presented in Table 7, Table 8 shows the corresponding $\overline{A_a^d}[s_b(i',k)]$ and $\overline{A_v^d}[s_b(i',k)]$ VA vector components for C_1 as k increases. Clearly, there is a significant difference of the arousal's and valence's variations when the source is moving towards the back of the listener, with an increased impact on the listener's affective state when the source is exactly behind him.

Summarizing the analysis for C_1, sound events assigned to this class tend to alter the listener's arousal slightly more than his valence especially when the source is located at the lateral limits and beyond of his field of vision. In addition, the changes of the emotional state are oriented towards the high arousal and low valence quadrant of the VA space, resulting into emotions verbally described with words assigned to the second quadrant or with words located near to the valence axis (for third and fourth quadrants) or near to the high values portion of arousal axis (for the first quadrant case). Thus, sound events in C_1 are likely to activate the listener and make him feel less pleasant, especially when they are located exactly behind him.

Sound events in C_2 present a rather similar behavior with those in C_1 in respect to the resulting listener's emotional state but for the opposite quadrant. Table 7 shows that their mean effect to the elicited emotional state is oriented slightly towards arousal. In particular, they exhibit a mean absolute angle value close to 135 degree. The partial components $\overline{A_a^d}[s_b(i',k)]$ and $\overline{A_v^d}[s_b(i',k)]$ of the mean VA vectors, for all angular transitions are shown in Table 8: sound events in C_2 do not impose notable differences on arousal or valence. Instead, their mean impact on both VA space dimensions is rather the same and demonstrates a decrease as the source moves towards the locations that are outside of the listener's field of vision. In addition, sound events in C_2 receive affective annotations that are towards high valence and low arousal values (i.e. laying within the

TABLE 6
Sound Events Included in Class C_1 and C_2 for All k

i'	Semantic content	$\phi(i',k)$	$\lvert\overrightarrow{va}(i',k)\rvert$	i'	Semantic content	$\phi(i',k)$	$\lvert\overrightarrow{va}(i',k)\rvert$
	Class 1				**Class 2**		
			$k = 2$ (angular transition from 0° to 45°)				
9	Tropical	−78.24	1.44	3	Baby	144.58	1.31
21	Fight 2	−60.95	1.14	5	Bees	126.05	1.18
22	Fight 3	−23.20	1.90	7	Rattlesnake	136.90	1.04
35	Plane crash	−28.07	1.89	8	Robin	174.68	2.60
45	Walking	−1.71	0.76	10	Erotic fem 1	179.06	1.36
50	Electricity	−20.62	2.06	15	Male cough	143.51	1.44
				16	Couple sneeze	105.48	0.80
				19	Vomit	121.15	1.52
				20	Whistling	175.36	0.26
				23	Victim	159.01	1.23
				26	Writing	175.96	1.11
				28	Paint	104.04	0.52
				30	Rain 1	168.02	0.34
				31	Helicopter 1	152.88	0.84
				36	Engine failure	162.65	2.00
				38	Thunderstorm	171.87	0.94
				42	Cuckoo	160.35	0.83
				43	Alarm	151.78	2.58
				44	Slot machine 2	171.23	3.36
				49	Choir	102.76	0.78
			$k = 3$ (angular transition from 45° to 90°)				
2	Dog	−79.94	1.27	6	Chickens	173.93	0.53
3	Baby	−7.99	0.74	13	Male laugh	93.84	1.02
5	Bees	−18.71	0.60	17	Man wheeze	133.78	1.48
8	Robin	−36.43	1.09	20	Whistling	118.07	0.27
12	Boy laugh	−60.95	0.74	21	Fight 2	150.86	1.44
15	Male cough	−71.57	1.36	27	Horse race	180.00	0.64
25	Typewriter	−57.69	0.56	28	Paint	90.00	0.63
26	Writing	−35.33	1.49	29	Sink	169.77	1.02
32	Countdown	−60.26	0.38	30	Rain 1	91.70	1.02
34	Wind	−71.57	1.19	31	Helicopter 1	107.65	0.58
36	Engine failure	−31.86	2.71	33	Car horns	134.88	1.78
39	Explosion	−27.31	1.96	35	Plane crash	142.68	1.41
43	Alarm	−5.57	1.72	37	Bike wreck	169.58	1.92
45	Walking	−71.57	0.29	40	Phone 1	144.46	0.22
49	Choir	−17.20	2.20	41	Clock	129.62	1.90
				50	Electricity	135.00	1.68
			$k = 4$ (angular transition from 90° to 135°)				
6	Chickens	0.00	1.92	8	Robin	142.79	0.59
12	Boy laugh	−82.41	0.27	10	Erotic fem 1	135.00	0.31
19	Vomit	−54.46	0.78	13	Male laugh	95.71	0.71
20	Whistling	0.00	0.14	18	Male sneeze	106.77	1.48
24	Creep	−55.30	0.28	25	Typewriter	136.23	1.00
29	Sink	−35.80	1.07	26	Writing	122.35	1.33
30	Rain 1	−64.71	1.14	31	Helicopter 1	99.46	1.52
34	Wind	−25.02	0.61	32	Countdown	121.26	2.08
35	Plane crash	−77.47	0.69	33	Car horns	171.47	0.36
37	Bike wreck	−22.09	1.73	36	Engine failure	164.05	1.32
41	Clock	−33.69	1.56	38	Thunderstorm	174.39	2.61
43	Alarm	−70.02	1.46	39	Explosion	152.41	2.73
50	Electricity	−45.00	0.47	40	Phone 1	134.47	1.10
				42	Cuckoo	175.96	1.25
				47	Harp	180.00	2.11
			$k = 5$ (angular transition from 135° to 180°)				
7	Rattlesnake	−17.44	1.86	2	Dog	164.05	1.27
13	Male laugh	−15.54	1.96	3	Baby	91.83	1.57
16	Couple sneeze	−14.62	0.28	5	Bees	176.19	2.15
17	Man wheeze	−87.99	0.52	6	Chickens	141.99	1.25
18	Male sneeze	−42.88	1.45	8	Robin	141.51	1.22
20	Whistling	−66.64	0.62	19	Vomit	141.67	0.52
21	Fight 2	−5.87	0.81	22	Fight 3	106.70	0.46
24	Creep	−82.57	0.37	26	Writing	93.18	0.38
25	Type writer	−6.30	1.56	30	Rain 1	172.57	0.97
28	Paint	−19.98	1.17	34	Wind	129.92	1.95
29	Sink	−15.22	2.34	35	Plane crash	148.57	0.72
31	Helicopter 1	−71.39	2.59	37	Bike wreck	172.98	2.48
32	Countdown	−38.66	2.05	43	Alarm	103.84	0.81
33	Car horns	−75.96	1.55	45	Walking	118.24	0.80
38	Thunderstorm	−61.75	0.69	46	Cork pou	148.57	0.75
39	Explosion	−20.03	1.48	50	Electricity	108.43	0.35
41	Clock	−60.26	1.07				
47	Harp	−30.26	1.54				
48	Bach	−90.00	0.70				

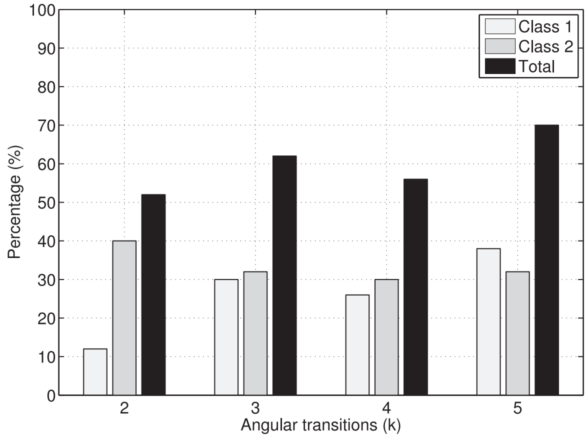


Fig. 5. Percentage of the population of classes C_1, C_2 and $C_1 + C_2$ as $\theta(k)$ increases.

fourth quadrant of the VA space) or values that are close to a) the valence axis when the emotional annotation falls into the first or second quadrant and b) the low values of the arousal axis for affective annotations in the third one. Again the lack of an exact mapping of verbal descriptions of emotions to VA space cannot allow an analytical investigation of the connection between emotions and angular transitions. What can be inferred though is that sound events in C_2 are increasing the listener's pleasure and decrease his activation.

Recapitulating the above discussion, clearly there are sound events that impose a systematic effect on the listener's emotional state as the source is moving towards the lateral limits of his field of vision. This change is either an increased activation combined with lowered pleasantness or an increased pleasantness integrated with a lowered activation. The amount of events that can be assigned to either C_1 of C_2 is increasing with the angular position k. In addition, when the source is located exactly at the rear of the listener, the amount of the events that cause increased arousal and lower valence is greater than the amount of the sound events that demonstrate increased valence and decreased arousal. The alteration of $\overrightarrow{va}(i', k)$ seems to be coherent in both classes, since the σ values for $|\phi(i', k)|$ have a maximum limited to 29.67 degree. To clarify the latter claim, the small σ values for $|\phi(i', k)|$ indicate that the changes of arousal and valence for both classes result in values focused to a small and concentrated area in the VA space, i.e. second and fourth quadrant for the first and second class respectively, and are not dispersed.

TABLE 7
Statistics of the $|\phi(i', k)|$ and $|\overrightarrow{va}(i', k)|$ Values for the SEs Included into C_1 and C_2

	Class 1		Class 2	
	$\|\phi(i',k)\|$	$\|\overrightarrow{va}(i',k)\|$	$\|\phi(i',k)\|$	$\|\overrightarrow{va}(i',k)\|$
$k = 2$ (angular transition $0° \rightarrow 45°$)				
Max	78.24	2.06	179.06	3.36
Min	1.71	0.76	102.76	0.26
Mean	35.46	1.53	149.37	1.30
σ	28.44	0.51	25.50	0.79
$k = 3$ (angular transition $45° \rightarrow 90°$)				
Max	79.94	2.71	180.00	1.92
Min	5.57	0.29	90.00	0.22
Mean	43.60	1.22	135.36	1.10
σ	25.35	0.70	29.44	0.58
$k = 4$ (angular transition $90° \rightarrow 135°$)				
Max	82.41	1.92	180.00	2.73
Min	0.00	0.14	95.71	0.31
Mean	43.54	0.93	140.82	1.37
σ	27.11	0.59	28.33	0.75
$k = 5$ (angular transition $135° \rightarrow 180°$)				
Max	90.00	2.59	176.19	2.48
Min	5.87	0.28	91.83	0.35
Mean	43.33	1.30	135.01	1.10
σ	29.67	0.68	28.63	0.65

TABLE 8
Mean Arousal and Valence Vector Components for C_1 and C_2 as k Increases

	Arousal	Valence	Arousal	Valence
	C_1		C_2	
$k = 2$	1.25	0.89	−1.12	0.66
$k = 3$	0.88	0.84	−0.78	0.77
$k = 4$	0.67	0.64	−1.06	0.87
$k = 5$	0.95	0.89	−0.78	0.78

5.3 Musical Sound Events Ratings

Table 9 summarizes the arousal and valence values for those sound events that have musical content, showing a significant change in valence, especially when the source is located in lateral positions. Thus, the listener feels more pleasant when the sound source moves from just in front of him towards his side. In particular, a common peak of valence is observed at 45 degree ($k = 2$). This fact can be partially related with the widely-employed encoding forms for immersive music reproduction, that aim to locate music sound sources within an expanded spatial field in the front of the listener. But when the source is located at 90 degree or more (i.e. for $k \geq 3$), a relatively unusual case for common music panning mixes, not all musical sound events demonstrate an increment in valence.

Moreover, for k values equal to 3 and 4, the sound events that continue to impose increased valence are a musical instrument playing (harp), or a musical piece excerpt (Bach). The one that does not exhibit the same trend in valence annotation is a singing voice (choir). This observation may also rely on existing listening conventions: in typical music reproduction setups, the singing voice is panned

TABLE 9
$a(i', k)$ and $v(i', k)$ Values for Music Sound Events

		k				
Sem. content	i'	1	2	3	4	5
		$a(i', k)$				
Harp	47	2.78	3.85	4.78	2.67	4.00
Bach	48	5.00	5.31	4.75	5.50	5.86
Choir	49	4.57	4.40	6.50	5.86	5.43
		$v(i', k)$				
Harp	47	5.56	7.39	7.45	7.45	6.67
Bach	48	6.38	7.62	6.37	6.80	6.10
Choir	49	5.14	5.90	5.25	5.14	4.43

in the vicinity of 0 degree. Thus, its virtual placement in extreme lateral positions may lower the pleasantness of the receiver. This trend of valence increment is not valid for all musical sound events when the source is located exactly at the back side of the listener. Finally, the arousal values seem not to indicate an observable, systematic behavior with either semantic content or different k values.

6 Conclusions

In the work at hand we investigate the potential relation of the spatial position (taking into account 5 different angles between the listener and the source which cover the right half of the horizontal plane) of a sound event to the elicited emotion of the listener. Does the location of the sound source alter the affective state of a listener? And if yes, what is the difference in the emotional state as the source is moving around the listener? Is there any relation with the listener's field of vision? Although the relation between sound (and particularly music) and emotions is the particular topic of interest for many existing studies, forming answers to these simple questions that focus on the affective impact of spatial position of generalized sound events is definitely limited. Thus, this work aims to contribute to the research foundations in the field of emotion recognition from general sounds.

This contribution is threefold: apart from the main exploration of the potential relation of emotions and spatial source positioning, we also developed an emotionally annotated and spatially-varied sound events database, which can be employed as the ground-truth dataset in future investigations, and provide a basis for emotionally enhanced sound design for multimedia and computer applications. The aforementioned dataset was formed through binaural processing of the IADS pre-annotated dataset that originally consists of 167 sound events. The final sound corpus included sound events located at five different spatial positions, effectively covering the entire horizontal plane. This binaural sound corpus is emotionally annotated through a series of online subjective evaluation experiments using a custom web platform specifically developed for the needs of the presented work. With the announcement of the experiments through international mailing lists, subjects originating from all over the world participated, thus suppressing the impact of the potential cultural differences on the subjective ratings. International participation was confirmed by the different domain origins of the email addresses submitted during the applied login process.

The analysis of the set of annotations obtained reveals various, multilevel outcomes. For example, one conclusion regards the sound events emotional annotation itself. In particular, short-range differences are obtained between the original IADS annotations and the ones derived in the case of placing the sound source in the front of the listener. Since spatial position represents an inherent characteristic of a sound event, which is naturally and automatically communicated to the listener, the presence of this information to the emotional annotation can be considered that adjusts the annotation results in order to accurately meet real-world acoustic conditions. There is one issue that regards sound events perceived as emanating from spread sources and that it is related to whether spatial positioning is an important physical parameter for such events. Further investigation is needed in order to fully resolve this question risen for these particular sound events, since other effects can also alter the annotated emotional states (e.g. lab versus no-lab conditions, etc.).

It was also observed that the majority of the binaural sound events impose a specific trend to the elicited emotional state of the receiver. This trend can be analyzed in two categories with respect to the movement of the sound source from the front to the rear side of the listener: events which increase the arousal and decrease the valence of the listener and vice versa. This reveals that the movement of the source has a rather systematic effect on the listener's arousal and valence and tends to result in emotions that correspond to either the second or fourth quadrants of the VA space or in emotions located near the VA axes. Nevertheless, due to the wide spread of the semantic content of the sound events in the original IADS dataset, a focused investigation on the possible relation between the semantic content, the affective state of the listener and the source's spatial position could not be examined. Towards this aim, future work may focus on the creation of emotionally annotated datasets from online sound databases that include sound events with similar semantic content (e.g. findsounds [45]).

The work at hand can be considered as a starting point that offers a useful set of evidences about the relation of sound events placement around the listener and the elicited emotions, under a limited set of spatialization parameters (i.e. angular variation limited at the half of the horizontal plane, constant distance and no directional sound source parameterization). Further work is clearly needed towards the complete exploration of the impact of source's spatial position on the listener's emotional state, taking into account all the above spatial conditions. For example, from a non-technical, but psychological point of view, the assumption made on the symmetry of the derived binaural signals on the two halves of the horizontal plane should by carefully reconsidered, due to the measured asymmetry in the perception of emotional stimuli through dichotic listening between the two brain hemispheres [48], [50]. The potential impact of the semantic content represents an additional approach. An extension of the current work may also consider sound source placement in three dimensions, taking into account the vertical plane. Another approach may incorporate more complicated affective models with additional dimensions (such as valence, arousal and dominance). Finally, the employment of a cloud source platform may provide significant advantages in the process of extending the developed, emotionally annotated binaural dataset.

Being a significant part of auditory interfaces, an affective-driven sonification/sound design process is expected to provide the means for delivering more realistic interfaces in a wide range of application fields were auditory representation and feedback is essential for conveying information. Under this perspective, the investigation of the relation of spatial position of everyday sound events to the elicited emotions may offer new principles and guidelines for sound designers that are involved in the process of developing authentic and immersive digital environments, such audio-only, video and serious games, virtual worlds and augmented reality applications.

Acknowledgments

Konstantinos Drossos is the corresponding author.

References

[1] K. Drossos, A. Floros, and N.-G. Kanellopoulos, "Affective acoustic ecology: Towards emotionally enhanced sound events," in *Proc. 7th Audio Mostly Conf.: Conf. Interact. Sound*, 2012, pp. 109–116.

[2] K. R. Scherer, "Appraisal theory," in *Handbook of Cognition and Emotion*, T. Dalgleish and M. J. Power, Eds. West Sussex, England: Wiley, 1999.

[3] K. Drossos, R. Kotsakis, G. Kalliris, and A. Floros, "Sound events and emotions: Investigating the relation of rhythmic characteristics and arousal," in *Proc. 4th Int. Conf. Inform., Intell., Syst. Appl.*, 2013, pp. 1–6.

[4] P. N. Juslin and P. Laukka, "Communication of emotions in vocal expression and music performance: Different channels, same code?" *Psychol. Bull.*, vol. 120, no. 5, pp. 770–814, Sep. 2003.

[5] J. Sanghoon, S. Rho, B.-j. Han, and E. Hwang, "A fuzzy inference-based music emotion recognition system," in *Proc. 5th Int. Conf. Vis. Inform. Eng.*, 2008, pp. 673–677.

[6] Z. Shiliang, T. Qi, J. Shuqiang, H. Qingming, and G. Wen, "Affective MTV analysis based on arousal and valence features," in *Proc. IEEE Int. Conf. Multimedia Expo*, 2008, pp. 1369–1372.

[7] K. F. MacDorman, S. Ough, and C.-C. Ho, "Automatic emotion prediction of song excerpts: Index construction, algorithm design, and empirical comparison," *J. New Music Res.*, vol. 36, no. 4, pp. 281–299, 2007.

[8] Y.-H. Yang, Y.-C. Lin, Y.-F. 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.

[9] A. P. Oliveira and A. Cardoso, "Modeling affective content of music: A knowledge base approach," in *Proc. 5th Sound Music Comput. Conf.*, Jul. 2008.

[10] W. L. Cheung and G. Lu, "Music emotion annotation by machine learning," in *Proc. IEEE 10th Workshop Multimedia Signal Process.*, 2008, pp. 580–585.

[11] F. Weninger, F. Eyben, B. W. Schuller, M. Mortillaro, and K. R. Scherer, "On the acoustics of emotion in audio: What speech, music and sound have in common," *Front. Psychol.*, vol. 4, pp. 1–12, May 2013.

[12] M. Marcell, M. Malatanos, C. Leahy, and C. Comeaux, "Identifying, rating, and remembering environmental sound events," *Behav. Res. Methods*, vol. 39, no. 3, pp. 561–569, 2007.

[13] K. Wassermann, K. Eng, P. F. M. J. Verschure, and J. Manzolli, "Live soundscape composition based on synthetic emotions," *IEEE MultiMedia*, vol. 10, no. 4, pp. 82–90, Oct. 2003.

[14] N. Moustakas, A. Floros, and N. Grigoriou, "Interactive audio realities: An augmented/mixed reality audio game prototype," in *Proc. 130th Audio Eng. Soc. Conv.*, May 2011, pp. 427–434.

[15] P. Shah, S. Grant, and W. Chapin, "Calibration and 3-d sound reproduction in the immersive audio environment," in *Proc. IEEE Int. Conf. Multimedia Expo*, 2011, pp. 1–6.

[16] T. Garner and M. Grimshaw, "A climate of fear: considerations for designing a virtual acoustic ecology of fear," in *Proc. 6th Audio Mostly Conf.: Conf. Interact. Sound*, 2011, pp. 31–38.

[17] A. Tajadura-Jiménez, A. Väljamäe, E. Asutay, and D. Västfjäll, "Embodied auditory perception: The emotional impact of approaching and receding sound sources," *Emotion*, vol. 10, no. 2, pp. 216–229, Apr. 2010.

[18] I. Ekman and R. Kajastila, "Localization cues affect emotional judgments—Results from a user study on scary sound," in *Proc. Audio Eng. Soc. Conf.: 35th Int. Conf.: Audio Games*, Feb. 2009.

[19] 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, pp. 416–422, 2010.

[20] J. Blauert, *Spatial Hearing-Revised Edition: The Psychophysics of Human Sound Localization*, revised ed., J. S. Allen, Ed. Cambridge, Massachusetts, USA: MIT Press, 1983.

[21] A. Tsilfidis, E. Georganti, and J. Mourjopoulos, "Binaural extension and performance of single-channel spectral subtraction dereverberation algorithms," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2011, pp. 1737–1740.

[22] E. Georganti, T. May, S. van de Par, and J. Mourjopoulos, "Extracting sound-source-distance information from binaural signals," in *The Technology of Binaural Listening*, series Modern Acoustics and Signal Processing, J. Blauert, Ed. Berlin, Germany: Springer, 2013, pp. 171–199.

[23] J. Desloge, W. Rabinowitz, and P. Zurek, "Microphone-array hearing aids with binaural output .I. fixed-processing systems," *IEEE Trans. Speech Audio Process.*, vol. 5, no. 6, pp. 529–542, Nov. 1997.

[24] M. Jeub and P. Vary, "Binaural dereverberation based on a dual-channel wiener filter with optimized noise field coherence," in *Proc. IEEE Int. Acoust. Speech Signal Process. Conf.*, 2010, pp. 4710–4713.

[25] K. Drossos, S. Mimilakis, A. Floros, and N. Kanellopoulos, "Stereo goes mobile: Spatial enhancement for short-distance loudspeaker setups," in *Proc. 8th Int. Conf. Intell. Inform. Hiding Multimedia Signal Process.*, 2012, pp. 432–435.

[26] M. M. Bradley and P. J. Lang, "The international affective digitized sounds (2nd edition; iads-2): Affective ratings of sounds and instruction manual," NIMH Center for the Study of Emotion and Attention, Gainesville, FL, USA, Tech. Rep. B–3, 2007.

[27] B. Gardner and K. Martin, "Hrtf measurements of a kemar dummy-head microphone," MIT Media Lab Perceptual Computing, Tech. Rep. 280, 1994.

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

[29] R. Reisenzein, E. Hudlicka, M. Dastani, J. Gratch, K. Hindriks, E. Lorini, and J.-J. Meyer, "Computational modeling of emotion: Toward improving the inter- and intradisciplinary exchange," *IEEE Trans. Affective Comput.*, vol. 4, no. 3, pp. 246–266, Jul.–Sep. 2013.

[30] B. Schuller, S. Hantke, F. Weninger, W. Han, Z. Zhang, and S. Narayanan, "Automatic recognition of emotion evoked by general sound events," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2012, pp. 341–344.

[31] P. Ekman, "An argument for basic emotions," *Cogn. Emot.*, vol. 6, nos. 3 and 4, pp. 169–200, 1992.

[32] A. Ortony and T. J. Turner, "What's basic about basic emotions?" *Psychol. Rev.*, vol. 97, no. 3, pp. 315–331, 1990.

[33] S. Koelsch, "Towards a neural basis of music-evoked emotions," *Trends Cogn. Sci.*, vol. 14, no. 3, pp. 131–137, Mar. 2010.

[34] R. Adolphs, "Neural systems for recognizing emotion," *Curr. Opin. Neurobiol.*, vol. 12, no. 2, pp. 169–177, Apr. 2002.

[35] K. Hevner, "Experimental studies of the elements of expression in music," *Am. J. Psychol.*, vol. 48, no. 2, pp. 246–268, Apr. 1936.

[36] T. Li and M. Ogihara, "Detecting emotion in music," in *Proc. Int. Symp. Music Inform. Retrieval*, 2003, pp. 239–240.

[37] C. A. Smith and P. C. Ellsworth, "Patterns of cognitive appraisal in emotion," *J. Pers. Soc. Psychol.*, vol. 48, no. 4, pp. 813–838, Apr. 1985.

[38] M. M. Bradley and P. J. Lang, "Affective reactions to acoustic stimuli," *Psychophysiology*, vol. 37, no. 2, pp. 204–215, 2000.

[39] C. Stickel, M. Ebner, S. Steinbach-Nordmann, G. Searle, and A. Holzinger, "Emotion detection: Application of the valence arousal space for rapid biological usability testing to enhance universal access," in *Proc. 5th Int. Conf. Universal Access in Human-Computer Interaction. Addressing Diversity*, 2009, pp. 615–624.

[40] J. Posner, J. A. Russel, and B. S. Peterson, "The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology," *Dev. Psychopathol.*, vol. 17, no. 9, pp. 715–734, 2005.

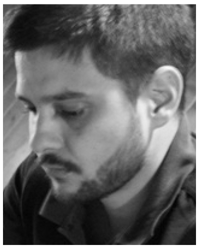
[41] J. A. Russel, "A circumplex model of affect," *J. Pers. Soc. Psychol.*, vol. 39, no. 6, pp. 1161–1178, 1980.

[42] J. Posner, J. A. Russell, and B. S. Peterson, "The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology," *Dev. Psychopathol.*, vol. 17, no. 3, pp. 715–734, Sep. 2005.

[43] W. W. Gaver, "What in the world do we hear? an ecological approach to auditory event perception," *Ecol. Psychol.*, vol. 5, no. 1, pp. 1–29, 1993.

[44] F. Hagman, "Emotional response to sound—Influence of spatial determinants," Master's thesis, Chalmers Univ. Technology, Göteborg, Sweden, 2010.

[45] Findsounds—Search the web for sounds [Online]. Available: www.findsounds.com, 2014.
[46] M. Grimm and K. Kroschel, "Evaluation of natural emotions using self assessment manikins," in *Proc. IEEE Workshop Autom. Speech Recognit. Understanding*, Nov. 2005, pp. 381–385.
[47] C. I. Cheng and G. H. Wakefield, "Moving sound source synthesis for binaural electroacoustic music using interpolated head-related transfer functions (HRTFS)," *Comput. Music J.*, vol. 25, no. 4, pp. 57–80, Dec. 2001.
[48] M. Bryden, R. G. Ley, and J. Sugarman, "A left-ear advantage for identifying the emotional quality of tonal sequences," *Neuropsychologia*, vol. 20, no. 1, pp. 83–87, 1982.
[49] R. Schatz, S. Egger, and K. Masuch, "The impact of test duration on user fatigue and reliability of subjective quality ratings," *J. Audio Eng. Soc*, vol. 60, no. 1/2, pp. 63–73, 2012.
[50] A. Carmon and I. Nachshon, "Ear asymmetry in perception of emotional non-verbal stimuli," *Acta Psychol.*, vol. 37, no. 6, pp. 351–357, 1973.



Konstantinos Drossos received the BEng degree equivalent on sound and musical instruments technology with distinction from the Technological Educational Institute of Ionian Islands. In 2007, he received the MSc degree from ISVR, University of Southampton. Currently, he is working toward the PhD degree at the Department of Audiovisual Arts of Ionian University. His main research interests include emotion recognition from sound events, affective acoustic ecology, audio processing, audio perception, and audio interfaces. He was an acoustic consultant and an adjunct lecturer at the Department of Sound and Musical Instruments Technology of the Technological Educational Institute of Ionian Islands. He is a student member of the Audio Engineering Society and of Hellenic Institute of Acoustics.

Andreas Floros received the engineering and PhD degrees in the area of digital audio technology from the Department of Electrical and Computer Engineering, University of Patras. In 2001, he joined the semiconductors industry, leading projects in the area of digital audio delivery over PANs and WLANs, quality-of-service, mesh networking, wireless VoIP technologies and lately with audio encoding and compression in embedded processors. During 2003-2005, he was a member of IEEE Tasks Groups 802.11e, .11k, and .11s. For a period of three years (2005-2008), he was an adjunct professor at the Department of Informatics, Ionian University, teaching also at the postgraduate (MSc) program "Arts and Technologies of Sound" organized by the dept. of Music Studies. Since 2009, he has been an assistant professor at the Department of Audiovisual Arts, Ionian University. His current research interests focus on analysis, processing and conversion of digital audio signals, digital audio coding and distribution techniques, digital audio technologies for multimedia and networking applications, audio systems for consumer and professional applications, wireless technologies for multimedia applications (with emphasis on quality of service for wireless LANs) and high-quality audio streaming over packet networks. He is a member of the ACM, the Audio Engineering Society (currently serving as the Secretary of the AES Greek Section), and the treasurer of the Hellenic Institute of Acoustics.

Andreas Giannakoulopoulos received the BAs (Ptychio) degree in economics and in communication and media studies from the University of Athens (UoA), the master's of arts degree in communication and media studies from UoA, and the master's of science degree in logic from the University of Amsterdam. He is a lecturer at the Department of Audiovisual Arts of Ionian University, where he teaches courses related to Internet communication, new media and the web technologies. His doctoral dissertation was in the field of web accessibility. The main field of his academic activities is computer mediated communication with particular interest on web development technologies. His research interests focus on information architecture, web-based media, content management platforms, and e-learning systems as means of effective communication via the web.

Nikolaos Kanellopoulos received the BSc, MSc, and PhD degrees in 1976, 1978, and 1982, respectively, from the University of Newcastle-upon-Tyne, United Kingdom. Currently, he is a professor of computer systems and informatics applications in the Department of Audiovisual Arts, Ionian University and an elected member of the Ionian University Council. In the past, he has served as the head of the Audiovisual Arts Department, as a faculty member of the Computer Science Engineering and Informatics Department, University of Patras, the Archives and Library Science Department, Ionian University, and as the president for the Greek National School of Dance. Since 1986, he has been contributed in the development of the Department of Computer Engineering & Informatics of Patras University and the Research Academic Computer Technology Institute of the Ministry of Education. He has been a member of the Governmental National Committee of Informatics and has served as a special scientific advisor, reviewer, and an evaluator on information technology matters to the Greek Ministers of Education, Research and Development, Justice, Defense, Presidency, Press & Mass Media. He is a member of the internal evaluation committee of Ionian University. His main research interest is in the application of digital technology in audio-visual art interactive systems.