Affective Acoustic Ecology: Towards Emotionally Enhanced Sound Events

Konstantinos Drossos
Dept. of Audiovisual Arts
Ionian University
Corfu, Greece
kdrosos@ionio.gr

Andreas Floros
Dept. of Audiovisual Arts
Ionian University
Corfu, Greece
floros@ionio.gr

Nikolaos - Grigorios Kanellopoulos
Dept. of Audiovisual Arts
Ionian University
Corfu, Greece
kane@ionio.gr

ABSTRACT
Sound events can carry multiple information, related to the sound source and to ambient environment. However, it is well-known that sound evokes emotions, a fact that is verified by works in the disciplines of Music Emotion Recognition and Music Information Retrieval that focused on the impact of music to emotions. In this work we introduce the concept of affective acoustic ecology that extends the above relation to the general concept of sound events. Towards this aim, we define sound event as a novel audio structure with multiple components. We further investigate the application of existing emotion models employed for music affective analysis to sonic, non-musical, content. The obtained results indicate that although such application is feasible, no significant trends and classification outcomes are observed that would allow the definition of an analytic relation between the technical characteristics of a sound event waveform and raised emotions.

Categories and Subject Descriptors
H.5.1 [**Information Interfaces & Presentation**]: Audio input/output

General Terms
Human factors, theory, measurement

Keywords
Affective acoustic ecology, emotion recognition, sound events

1. INTRODUCTION
Audio, in its multiple alternative forms, represents one of the fundamental means for human communication. Starting from speech, as the primary medium for information distribution [1], and followed by music, which was firstly used as a way to enhance orally expressed messages by mimicking and extending vocal expression's characteristics [2], it is also used as a transfer path for conveying emotional information.

But speech and music seem to be only a portion of the sounds that we hear. There are also non-musical and non-linguistic sounds [3] which carry information that constructs the ambient audio environment presented around a listener [4]. These sounds may originate from human activities (e.g. coughing, screaming or walking), nature (wind blow or a water stream), animals (barking) or interaction between objects (rattling noise). These sounds bear information about their source (like speed, placement and movement, size of the source) [3, 4], and provide to the receiver the necessary input for taking decisions or interacting with the environments. A typical example is the noise from a car approaching that can affect the decisions and actions made by a person that wants to cross a road. The establishment of this kind of relationship between the receiver and a sound event, as well as the encapsulated information that a sound event can carry may also trigger appropriate responses and engender emotion(s) to the receiver, such as fear [5]. Hence, sound events are also closely related to emotions.

Regarding emotions, there are a number of issues that multiple researchers from different disciplines are trying to investigate. For example, how emotions are created? What is the exact relationship between them? And how they can be accurately modeled? The above represent some typical fundamental questions of an ongoing research debate on emotions [2]. Affective models examine these aspects from a specific perspective and, in some cases, offer a graphical and quantitative representation of emotions. Typical examples of such models are the arousal valence plane [6] and the basic emotions model [7].

Neuroscience and cognitive sciences widely use these models in research of physiological responses and their relation to emotions [8]. Focusing on music, Music Information Retrieval (MIR) and Music Emotion Recognition (MER) fields excessively use them for exploiting and exploring the ability of music signals to evoke emotions on the listener [9]. More specifically, MIR is mainly targeted to a more efficient and content-oriented music categorisation (compared to the legacy "Band / Artist - Genre / Style - Album" one) using emotions [10]. On the other hand, MER aims at investigating the relationship between music and listeners' emotions [9]. Nevertheless, the above MER and MIR application fields are confined by the exclusive usage of music as sound stimuli, since they do not explore the link between emotions and sound events.

It is also widely known that the relationship of the leaving beings and their sonic environment is a fundamental issue of acoustic ecology [11, 12]. This relation is ubiquitous in both real and virtual (or synthetic) worlds. On one hand, there are everyday audio stimuli from every possible sound event source, while, on the other hand, audio environments aid the immersion to the created realms [13]. There are studies examining the relation between soundscapes, as a formed and complete audio entity comprised by sound events, and emotions [14] or quality [15]. Additionally, there are researches concerned with emotional reactions to specific sound events, like for example dental engine [16]. But, to the best of the authors' knowledge, there are no research efforts aiming to study a general relation of emotions and discrete surrounding sound events.

In this work, we define the concept of affective acoustic ecology as the relation of the surrounding sound events with the emotions that these can evoke to the listener. Towards this aim, we use well-established affective models and existing techniques borrowed from emotion recognition from organized forms of sound (i.e. music or speech) and we try to investigate their application to a more general, non-linguistic and non-musical form of sound [3]. In order to provide an initial indication about the potential relation of sound-events and emotions, we use the International Affective Digital Sounds (IADS) emotionally annotated sound events database, which provides a number of sound events engrossing various emotions [17]. The reason for this selection is twofold. Firstly, all IADS sound events correspond to a large variety of everyday situations. Secondly, the affective annotation applied follows a thoroughly used emotions' model in MIR and MER researches.

The rest of the paper is organized as follows: In Section 2 a brief overview of the existing emotions' models is presented. Section 3 is concerned with the music emotion recognition process while in Section 4 the sound event concept under the scope of affective acoustic ecology is defined. Section 5 demonstrates the results obtained by the combination of an affective model and the emotion recognition process applied to sound events. Finally, Section 6 concludes the work.

2. MODELLING OF EMOTIONS

One of the main components of emotion recognition, is the emotions' modeling approach employed, which is likely to be one of the major debates in the area of affective modeling and identification [2, 19]. In general, models of emotions can be divided in two main categories; discrete and continuous [18]. The most frequently used models in MIR and MER research fields are the so called basic emotions, the list of adjectives and the dimensional models [20, 21]. The former two are discrete models, while the latter represents a typical continuous approach.

Discrete emotions' models refer to the use of specific (or group of) words for describing an emotion. The basic emotions model for example employs anger, fear, happiness and sadness [22], based on the knowledge of Darwin era [7, 22]. The above emotions are considered as the basic ones, and all the rest can be derived by specific combinations of them. The main concept of this model was argued by [23]. Nevertheless, it seems to offer an attractive basis for emotion-related researches. The main reasons for this are likely to be the use of specific words and the actual emotions that are considered as basic. Specific word usage can offer an integrity between studies, by not referring to the same state with different words. On the other hand, basic emotions are frequently used in neurioscience due to their relation with specific brain areas, like amygdalae, or with physiological responses [21, 47, 25].

The list of adjectives proposed by [26] represents an alternative discrete emotions modeling approach. Emotions are not referred as single words, like in the case of basic emotions. Instead, a number of word groups is used, with most words in the same group to be synonyms [27]. These groups serve as a way to refer to and model emotions or emotional states. Each group is denoted as a class and consists of words capable for describing a similar emotional condition. The number of classes originally used was eight, but some researches extend it through additional classes, reaching a total of thirteen [28]. This model is frequently used in music psychology assessments. In Figure 1, a visual representation of the original eight classes is provided, as proposed by [26].

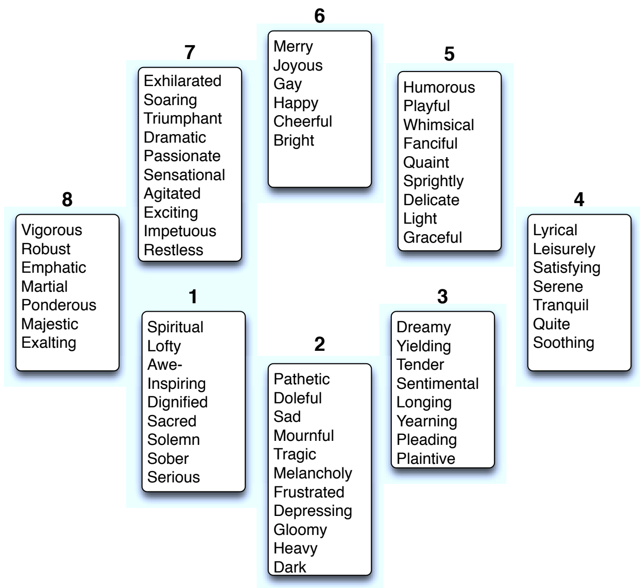

Figure 1: The list of adjectives with 8 classes according to [26, 27]

Focusing now on continuous affective models, they employ a combination of continuous values for representing the affective state of a human subject. As a result, this affective state is expressed as a point in an N-dimensional space, where N equals to the number of values used. These values are unit vectors serving as basic components of the affective states which lead to emotions. The amount of the chosen states equals to the amount of the dimensions used in the model. The most consistent across studies have found to be valence and arousal [29], leading to a two-dimensional (N=2) space for modeling the affective state, although more dimensions, like dominance [6] or activation and quality [30], were also proposed. The usage of arousal and valence dimensions is usually referred to as arousal-valence (AV) plane.

The derived affective state of the listener is mapped to a specific emotion through its position in the N-dimensional space. In the AV plane for example, high arousal and low valence values can indicate distress. On the contrary, in an arousal, valence and dominance space, a high arousal, low valence and low dominance situation can indicate fear. However, it should be noted that a specific match between values and emotions cannot be verified. What is likely to be used is a cluster of values which indicate the same emotion. Due the usage of affective states instead of emotions, different words can serve for verbal mapping, like misery and sadness for neutral arousal and low valence. This can lead to the incorporation of discrete models in the mapping procedure.

The center of the axes in a dimensional model can be mapped to the neutral affective state or to the extremes of each dimension. In the proposal of the circumplex model of affect [31] for example, which served as the basis for the dimensional approach, the center of axes was defined as the neutral state. Figure 2 illustrates the AV plane with emotions defined as specific points in the corresponding plane [19, 31, 32, 33]. These points can serve as center points during the AV values clustering process. On the other hand, Figure 3 provides a graphical representation of the AV plane with four emotions mapped to clusters of AV values [19, 33].

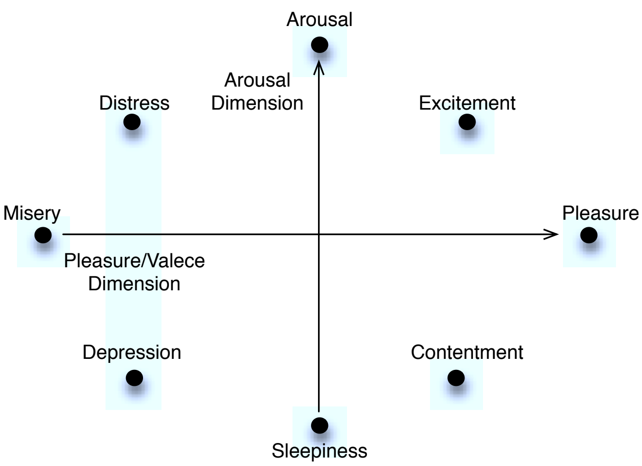

Figure 2: AV plane representation with indication to corresponding emotions

3. MUSIC EMOTION RECOGNITION

Since the advent of digital media, the available for distribution music data experienced a significant growth that rendered content classification a necessity. The legacy forms of classification based on the "Band/Artist-Genre-Style" arguments were found to be not sufficient provided the volume of the content [10, 34]. This fact is further strengthen by the increasing functionality that user-oriented classification can offer, focusing on the content and/or the perceived emotion of the music [35, 36].

Music emotion recognition can be considered a pattern recognition and/or classification problem. The audio content is

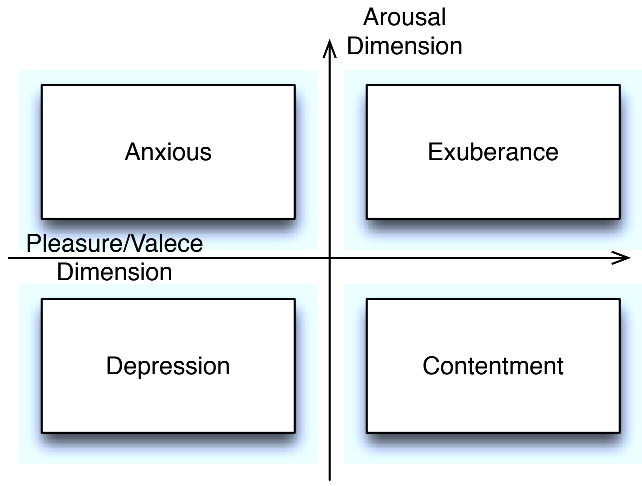

Figure 3: The AV plane with four emotions presented as clusters

firstly emotionally annotated by human subjects. Then, a relationship is derived between its technical characteristics / features and the outcome of the emotional annotation. These characteristics along with the annotation itself form the training data set. Emotion recognition from musical data can consider both raw (i.e. PCM encoded) audio data as well as symbolic musical data (presented using the Musical Instrument Digital Interface - MIDI). Nevertheless, in both cases emotions' models can be used in order to model the emotion that the listener experiences.

Extracted features from raw audio data typically concern timbral, temporal/rhythmic and energy/intensity characteristics of the signal. When symbolic data are used, structural music characteristics can be additionally extracted, like the musical key [37]. Timbral characteristics usually include spectral flatness, spectral flux, spectral centroid, spectral roll off and pitch. On the other hand, the zero-crossing rate, tempo, rhythm regularity and the rhythm strength parameters are frequently used as common temporal/rhythmic characteristics.

As mentioned previously, the annotation of the music material is carried out through listening tests with human subjects. A perceived emotion is indicated after the test, using either a discrete emotions' model (by presenting a list of emotions to choose from), or through a continuous one. A method presented in [38] offers an efficient way for self-annotation of affective states. It incorporates the usage of specific images depicting a manikin with specific characteristics. These are organized in three groups of five pictures each. Each group corresponds to the dimensions of arousal, valence and dominance. In each of them, only one characteristic of the manikins is varied, with this change being escalated in order to efficiently depict the corresponding variation in the affective state, i.e. from lower to higher. Thus, during the annotation process, the participant listens to the audio material and then chooses one manikin for each dimension. It has to be noted that the participant can choose also the state between two manikins which results in a total of nine available values for each dimension. The manikins for arousal and valence groups are shown in Figures 4 and 5 respectively.

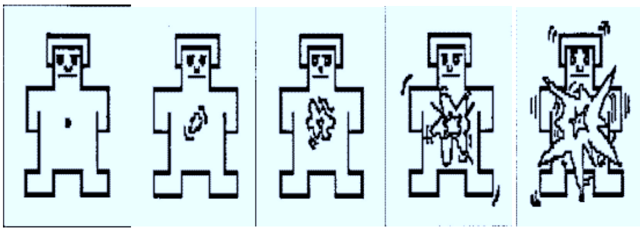

Figure 4: The images used for arousal estimation, from lower to higher, as proposed in [38]

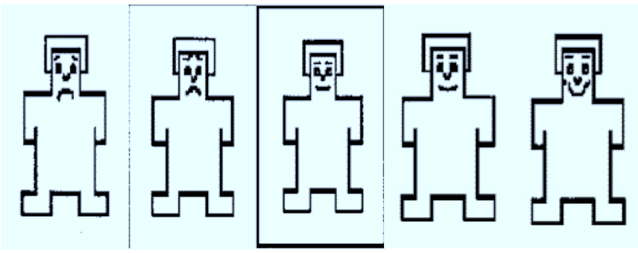

Figure 5: The images used for valence estimation, from lower to higher, as proposed in [38]

Subsequently, the above affective annotations along with the extracted features from the music signal serve as input to a classification algorithm. Many algorithms can be employed, such as the Gaussian Mixture Models (GMM) [19], Artificial Neural Networks (A-NN) [39] and Support Vector Machines (SVM) [28, 46]. Both discrete and continuous affective models can be used with the classification algorithms. This choice, on one hand, determines the available output of the recognized emotion and, on the other hand, can give rise to hierarchical structuring of the recognition process. Thus, if a discrete model is used, then the training of the algorithm is likely to produce a classification model which will be able to categorize the test input against the used emotions. On the contrary, if a continuous approach is adopted, then the classification algorithm will provide output values on the used dimensions in the emotions' model. The latter can be seen in [41], where the test data result into real values in the range of [-1 +1]. In addition, in [19, 33] a hierarchical approach was adopted using the AV plane. The resulting quadrants were divided in two groups, i.e. one with high arousal and one with low. Then, the intensity related features of the music material were correlated with the resulting arousal, allowing a hierarchical categorization of the material by firstly deciding on which of the two quadrants' group the resulting affective state would belong.

4. SOUND EVENTS

Although music is an important aspect and form of sound, in everyday life, an acoustic environment consists of non-linguistic and non-musical sounds. We can acoustically focus on these sounds under a more "musical" perspective, by considering for example their pitch and timbre. While in music listening such a consideration about timbre, pitch, musical scales or any other music related characteristic can enhance the listening experience, in a non-musical sonic content this may obscure information that the audio stimuli can bear. Alternatively, we can also hear these sounds with the "*everyday listening*" [4] approach, focusing on the attributes of the source, like its size, its movement/placement etc. [4].

The information of the non-musical listening allow the receiver to identify the sound sources and perceive their forthcoming or already established relation. Sonic content conveying this kind of data can be referred as sound events and such are widely used in a variety of applications or are apparent in various situations. They are evident in everyday life. Moving objects can be identified and also their route or speed can be estimated. The way that the sound is produced can be perceived, like in the case of someone scratching a wooden surface or like in the case that someone is hitting an object [4].

Sound events are widely used in a variety of applications. Multimedia and personal computers use sound events as a means to communicate with the user, indicating various states of the machine. To this scope, there are recent works which consider the sonification of information in order to incorporate such an auditory communication to human - computer interaction [42]. Video games use them as well, both as environmental sounds, in order to enhance user's immersion, and as scenario components [43], e.g. video games that are fighting oriented tend to have a fighting audio environment. Also, humans use, on purpose or accidentally, sound events communicating various information like emotional condition, e.g. a nervous walking [3, 4]. Moreover, in day-to-day communication one listens or produces sound events, like clapping, yelling, crying etc, which definitely provide a basis for emotion identification.

In [4] a framework regarding the nature of the sound events is proposed. According to it, different interacting material, e.g. vibrating objects or liquids, will result in different and distinguishable sound events which will have different values and characteristics for their technical features. Such features, like the high-frequency components' energy, the amplitude modulation or the bandwidth, allow the discrimination of the source's attributes that the sound event carry [4]. In addition to the aforementioned attributes of the source(s), sound events can also evoke or communicate emotions, as the IADS database clearly demonstrates. Its sonic content spans from simple and ordinary situations, like a dog barking, to extraordinary or unpleasant ones, like guns' sounds or vomiting [6].

Combing all the above, we can state that as a sound event can be considered as an audio structure that contains:

- a sound waveform
- manifestation of source's instantaneous spatial position, relative to the receiver
- duration
- indication of the sound's creation procedure
- evidence of vibrating objects' nature state (i.e. solid, liquid, gas)

- semantic content

The combination of these characteristics in a unified approach that extends acoustic ecology by considering elicited emotions, leads to the affective acoustic ecology concept proposed here.

In the present work we consider only the waveform of a sound event, applying methods and models used in research concerned also with the same characteristic but for musical content. To this scope, we employ and analyze the sound events provided by the IADS database [6]. The inspection of the utilized sonic material is based on technical features, classification algorithms and emotional models throughly used in MER and MIR researches.

5. EMOTION RECOGNITION FROM EMOTIONALLY ANNOTATED SOUND EVENTS

Prior to emotion recognition from the considered sound events, a pre-processing stage was carried out. All available audio files were resampled at a 16 kHz sampling frequency. From the annotation values, arousal and valence were used and these were clustered in four classes. The correspondence of each class with the annotated arousal and valence values is presented in Figure 6.

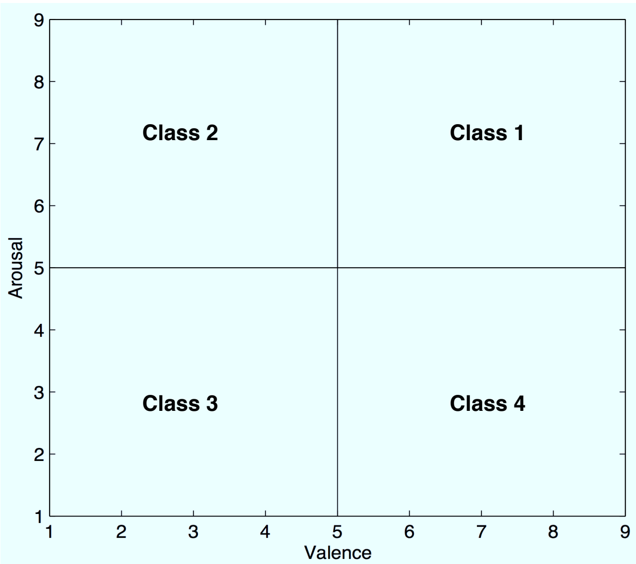

Figure 6: Mapping of AV values to the classes used in the emotion recognition process

The emotion recognition process was divided in two stages. First, each audio signal was processed and various technical features were extracted. Then, a classification algorithm was employed and trained using the emotional classes presented above and the extracted features from the signals. The accuracy that the classification algorithm could achieve was examined using the ten fold cross validation process from the available training data set. Ten fold cross validation means that for ten times, a 10% of the available training set, different for each of the ten times, was used as the testing set against the remaining 90% which was used as training dataset. Both training and testing was carried out using Weka [44]. Figure 7 graphically illustrates the above process.

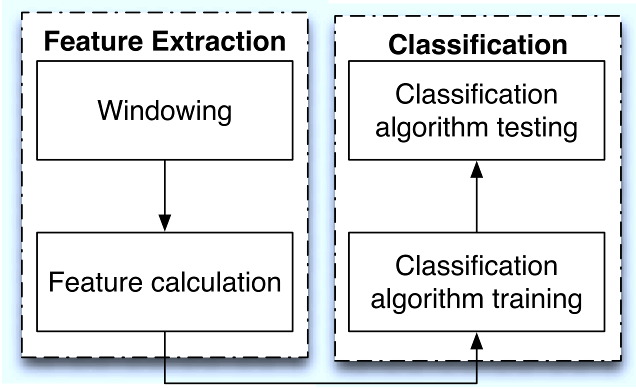

Figure 7: Illustration of the emotion recognition process followed

The total set of features consisted of 101 different parameters widely employed in the literature for MER purposes. Typical examples of such parameters include spectrum's attributes, like centroid, flux and others, Mel Frequency Cepstral Coefficients, energy of the signal, zero crossing rate, strength of beat and fundamental frequency estimation. All the above parameters are waveform-dependent and are not related to any additional properties of a sound event, as presented in the previous Section.

In the current process we used spectral centroid, spectral roll off, spectral flux, spectral variability, energy of the signal, fraction of low energy, zero crossing rate, fundamental frequency, Mel Frequency Cepstral Coefficients (13), strength of strongest beat, beat sum and compactness. Both average and standard deviation were calculated for all the aforementioned parameters. For the feature extraction, a window of 0.01 seconds was used with zero overlap. All the above parameters were extracted using JAudio [45].

5.1 Classification results

For each of the classification algorithms employed, the obtained accuracy from the ten-fold cross validation procedure is summarized in Table 1. It can be observed that the accuracy is below 50% for both algorithms used. Typical results for accuracy classification derived by existing published works are in the range of 82% [19], 83.5% [27], 73% [46] or 67%[47]. More specifically, the SVM classifier achieves a 43.7% accuracy, while A-NN achieves a lower percent of 36.5%. This implies that although the technical features used can offer an almost non-random classification, IADS database's sound events carry a dominant semantic content which is solely responsible for the elicited emotions of the listeners. In addition, while in various works [19, 33] the energy (E) of a signal was strongly and analogously related with the arousal dimension , in Figure 8 it can be observed that there is not such a relation between E and the arousal for the IADS sonic content.

In Table 2, a small number of some selected IADS sound

Table 1: 10-Fold cross validation results for each of the algorithms used

Algorithm	Training accuracy result
SVM	43.7%
A-NN	36.5%

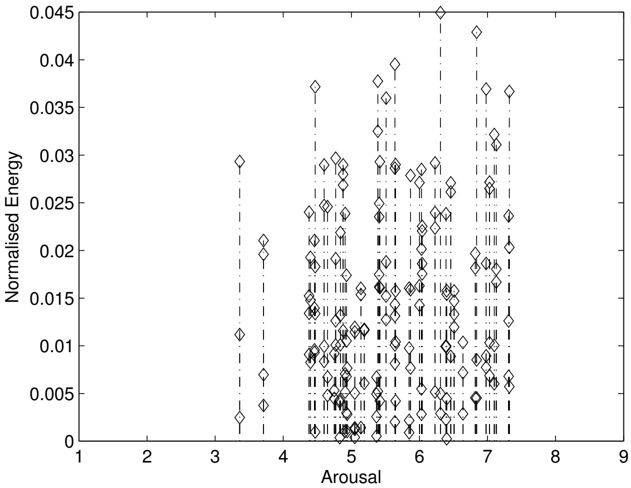

Figure 8: Arousal Vs Normalised energy values for the IADS database

events are presented, along with their semantic content, the corresponding arousal values and the energy E of the waveform. Considering the definition given in Section 4 for a sound event under the proposed scope of affective acoustic ecology, it seems that the characteristics of the waveform cannot produce an efficient classification model. Provided that these algorithms were previously incorporated in several published MER and MIR works with signals' features extracted from tis waveform, it can be inferred that for affective acoustic ecology tasks more parameters of the sound events must be considered. The most prominent in the current work tend to be the semantic content. With respect to Table 2, it seems that this content affects primarily the elicited affective state of the listener.

Table 2: Arousal sorted normalised energy, arousal values and semantic content of selected IADS sound events

Semantic content	E	Arousal value
Air raid alarm	0.0293	3.36
Male scream	0.0210	3.71
Female scream	0.0210	4.46
Bongos	0.0372	4.47
Soda Fizz	0.0009	5.85
Erotic Male	0.0029	6.64
Type Writer	0.0057	7.32

6. CONCLUSIONS

In the current work the concept of affective acoustic ecology is proposed. It is defined as an acoustic ecology extension by introducing the relation between sound events and elicited emotions. A definition of the sound event under this concept was also introduced, forming an audio structure with various attributes related to the source, the environment and the technical characteristics of the sound.

As a first step for investigation, we examined the application of thoroughly used techniques from MER and MIR fields, utilizing only the characteristics of the audio event extracted from its waveform. Such an approach is accustomed in the aforementioned fields for extracting acoustic cues from music signals. Then, we applied multiple classification algorithms, by using as dataset the only existing emotionally annotated sound events database. Although results showed low scores of accuracy, an attentive examination of the sound events' semantic content can reveal that there is a strong relation between it and the elicited emotions.

Considering on one hand the derived results and on the other the definition of the sound event under the affective ecology concept, a future work that would incorporate more aspects of the audio event structure seems to be required for defining an analytic relation between sound events and emotions.

7. ACKNOWLEDGEMENTS

The research activities that led to these results, were co-financed by Hellenic Funds and by the European Regional Development Fund (ERDF) under the Hellenic National Strategic Reference Framework (ESPA) 2007-2013, according to Contract no. MIKRO2-40/E-II-A.

8. REFERENCES

[1] P. N. Juslin and K. R. Scherer, "Vocal expression of affect," in The New Handbook of Methods in Nonverbal Behavior Research, J. A. Harrigan, et al., Eds., ed Oxford, Great Britan: Oxford Unversity Press, 2005, pp. 65-135.

[2] P. N. Juslin and P. Laukka, "Communication of emotions in vocal expression and music performance: different channels, same code?," Psychological Bulletin, vol. 129, pp. 770-814, September 2003.

[3] M. Marcell, et al., "Identifying, rating, and remembering environmental sound events," Behavior Research Methods, vol. 39, pp. 561-569, 2007.

[4] W. W. Gaver, "What in the World Do We Hear?: An Ecological Approach to Auditory Event Perception," Ecological Psychology, vol. 5, pp. 1-29, 1993/03/01 1993.

[5] T. Garner and M. Grimshaw, "A climate of fear: considerations for designing a virtual acoustic ecology of fear," presented at the Proceedings of the 6th Audio Mostly Conference: A Conference on Interaction with Sound, Coimbra, Portugal, 2011.

[6] M. M. Bradley and P. J. Lang, "Affective reactions to acoustic stimuli," Psychophysiology, vol. 37, pp. 204-215, 2000.

[7] R. R. Cornelius, "Theoretical approaches to emotion," presented at the International Speech Communication Association (ISCA) Tutorial and Research Workshop (ITRW) on Speech and Emotion, Newcastle, Northern Ireland, UK, 2000.

[8] P. Rainville, et al., "Basic emotions are associated with

distinct patterns of cardiorespiratory activity," International Journal of Psychophysiology, vol. 61, pp. 5-18, 2006.

[9] S. O. C.-C. H. MacDorman, Karl F., "Automatic Emotion Prediction of Song Excerpts: Index Construction, Algorithm Design, and Empirical Comparison," Journal of New Music Research, vol. 36, pp. 281 - 299, 2007

[10] Z. Shiliang, et al., "Affective MTV analysis based on arousal and valence features," in IEEE International Conference on Multimedia and Expo, 2008, pp. 1369-1372, 2008.

[11] B. Y. Newman, "And Now, Acoustic Ecology," Optometry - Journal of the American Optometric Association, vol. 76, pp. 629-631, 2005.

[12] M. R. Schafer, The Tuning of the World. Michigan: Knopf, 1977.

[13] S. Schnall, et al., "The Immersive Virtual Environment of the digital fulldome: Considerations of relevant psychological processes," International Journal of Human-Computer Studies, Corrected Proof.

[14] R. Cain, et al., "The development and application of the emotional dimensions of a soundscape," Applied Acoustics, Corrected Proof

[15] D. A. Hall, et al., "An exploratory evaluation of perceptual, psychoacoustic and acoustical properties of urban soundscapes," Applied Acoustics, Corrected proof.

[16] M. J. Gang and L. Teft, "Individual differences in heart rate responses to affective sound," Psychophysiology, vol. 12, pp. 423-426, 1975.

[17] Bradley, M. M. & Lang, P. J., "The International Affective Digitized Sounds (2nd Edition; IADS-2): Affective ratings of sounds and instruction manual," Technical report B-3. University of Florida, Gainesville, Fl., 2007

[18] S. Kai, et al., "An improved valence-arousal emotion space for video affective content representation and recognition," presented at the IEEE International Conference on Multimedia and Expo, 2009. ICME 2009, Cancun, Mexico, 2009.

[19] L. Lie, et al., "Automatic mood detection and tracking of music audio signals," IEEE Transactions on Audio, Speech and Language Processing, vol. 14, pp. 5-18, January 2006.

[20] K. R. Scherer, "Which Emotions Can be Induced by Music? What Are the Underlying Mechanisms? And How Can We Measure Them?," Journal of New Music Research, vol. 33, pp. 239 - 251, 2004.

[21] C. Laurier, et al., "Exploring Relationships between Audio Features and Emotion in Music " presented at the 7th Triennial Conference of European Society for the Cognitive Sciences of Music (ESCOM 2009) Jyvaskyla, Finland 2009.

[22] P. Ekman, "An argument for basic emotions," Cognition & Emotion, vol. 6, pp. 169-200, May 1992.

[23] A. T. Ortony, Terence J., "What's Basic About Basic Emotions?," Psychological Review, vol. 97, pp. 315-331, 1990.

[24] Y. Feng, et al., "Popular music retrieval by detecting mood," presented at the Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval, Toronto, Canada, 2003.

[25] L. C. De Silva, "Audiovisual emotion recognition," in Systems, Man and Cybernetics, 2004 IEEE International Conference on, 2004, pp. 649-654 vol.1.

[26] Hevner, K., "Experimental studies of the elements of expression in music," American Journal of Psychology, vol. 48, pp. 246-268, 1936.

[27] A. Wieczorkowska, et al., "Extracting Emotions from Music Data," in Foundations of Intelligent Systems. vol. 3488/2005, M.-S. Hacid, et al., Eds., ed Berlin, Germany: Springer Berlin / Heidelberg, pp. 456-465, 2005.

[28] Li, T., Ogihara, M.: Detecting emotion in music. 4th International Conference on Music Information Retrieval ISMIR, Washington, D.C., and Baltimore, MD, 2003.

[29] C. A. Smith and P. C. Ellsworth, "Patterns of cognitive appraisal in emotion," Journal of Personality & Social Psychology, vol. 48, pp. 813-838, April 1985.

[30] M. Bartoszewski, et al., "Extraction of Emotional Content from Music Data," in 7th Computer Information Systems and Industrial Management Applications, 2008. CISIM '08., 2008, pp. 293-299.

[31] J. A. Russell, "A circumplex model of affect," Journal of Personality & Social Psychology, vol. 39, pp. 1161-1178, 1980.

[32] B. v. d. Laar, "Emotion detection in music, a survey " presented at the 5th Twente Student Conference on IT, Enschede, The Netherlands, 2006.

[33] D. Liu, et al., "Automatic Mood Detection from Acoustic Music Data," presented at the ISMIR (International Conference on Music Information Retrieval) 2003, Baltimore Maryland, U.S.A., 2003.

[34] L. Tao and M. Ogihara, "Content-based music similarity search and emotion detection," in Acoustics, Speech, and Signal Processing, 2004. Proceedings. (ICASSP '04). IEEE International Conference on, 2004, pp. V-705-8 vol.5.

[35] Y.-H. Yang, et al., "Music emotion classification: a fuzzy approach," presented at the 14th annual ACM international conference on Multimedia, Santa Barbara, CA, USA, 2006.

[36] Y. Chia-Hung, et al., "An efficient emotion detection scheme for popular music," in Circuits and Systems, 2009. ISCAS 2009. IEEE International Symposium on, 2009, pp. 1799-1802.

[37] A. P. Oliveira and A. Cardoso, "Modeling Affective Content of Music: A Knowledge Base Approach," presented at the 5th Sound and Music Computing Conference Berlin, Germany, 2008.

[38] 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, pp. 49-59, 1994.

[39] Kostek, B. and Czyzewski, A.,"Representing Musical Instrument Sounds for Their Automatic Classification," Journal of Audio Engineering Society, vol. 49, pp 768?785, 2001.

[40] M. Wang, et al., "User-adaptive music emotion

recognition," in 7th International Conference on Signal Processing, 2004. Proceedings. ICSP '04. 2004, Beijing, China, 2004, pp. 1352-1355 vol.2.
[41] Y. H. Yang, et al., "Mr. Emo: music retrieval in the emotion plane," presented at the 16th ACM international conference on Multimedia, Vancouver, British Columbia, Canada, 2008.
[42] A. Dewitt and R. Bresin, "Sound Design for Affective Interaction," presented at the Proceedings of the 2nd international conference on Affective Computing and Intelligent Interaction, Lisbon, Portugal, 2007.
[43] L. J. Paul, "Granulation of Sound in Video Games," presented at the Audio Engineering Society 41st International Conference: Audio for Games, Vancouver, BC, Canada, 2011.
[44] M. Hall, et al., "The WEKA data mining software: an update," SIGKDD Explor. Newsl., vol. 11, pp. 10-18, 2009.
[45] D. McEnnis, et al., "jAudio: An Feature Extraction Library," in In 6th International Conference on Music Information Retrieval (ISMIR), 2005.
[46] M. Wang, et al., "User-adaptive music emotion recognition," in 7th International Conference on Signal Processing, 2004. Proceedings. ICSP '04. 2004, Beijing, China, 2004, pp. 1352-1355 vol.2.
[47] Y. Feng, et al., "Popular music retrieval by detecting mood," presented at the Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval, Toronto, Canada, 2003.