An Approach for Modeling Affective Acoustic Ecology in City Environments

Konstantinos Drossos, and Andreas Floros,
Audiovisual Signal Processing Lab.,
Dpt. of Audiovisual Arts, Ionian University,
Corfu - Greece,
email:{kdrosos, floros}@ionio.gr

Abstract

Urban sonic ecology represents a major field of research interest for exploiting the relations raised through sound between human populations and a city environment. Recently, the concept of emotional city has boosted the ideas and efforts towards exploiting the particular characteristics of an urban soundscape as a means for conveying emotions. The result of these efforts is the concept of affective acoustic ecology. In this work we aim to provide a fundamental perspective towards modeling affective acoustic ecology, based on a novel definition of sound events, the primitive components of a soundscape. The proposed model considers both acoustic signal-dependent features and semantic content as the potential soundscape parameters that trigger emotions.

1 Introduction

In recent years urban environments seem to undergo a thorough research regarding the characteristics and the quality of their ambient soundscape. The latter has been identified as an important factor to physical comfort [1] and to the impression of the overall environment [2]. In addition, urban acoustic environments have been recognised as complicated systems, related to physical and psychological factors and they seem to consist mostly from SE's with their semantic content ranging from anthropogenic sounds to geophysical sounds [1–3]. Their importance in acoustic comfort has been already assessed by existing research studies, e.g. [1].

Sound Events (SEs) represent an essential component of soundscapes. An SE is defined as a general form of sound which can convey information regarding the nature of the source, the environment and the sound producing mechanism [4]. SE is an external stimulus. As such, it can evoke emotions [5]. On the other hand, one of the primary means associated with hearing and used by humans in order to convey emotions is music, a structured form of sound. Its foremost utilization was to extend and mimic voice characteristics in various social events [6]. Recent findings from Music Emotion Recognition and Music Information Retrieval disciplines present a connection between music's technical characteristics and emotions [7]. Emotion recognition, categorization and music retrieval based on emotion are reported as well [8–11].

Regarding SE's and their connection to receiver's elicited emotions, a scarcity of researches can be observed [12]. Only recently there are some published works that investigate the association of emotions and SE's [4, 12, 13]. In these words, a new proposed model is introduced which suggests the extension of the acoustic ecology concept to affective acoustic ecology [4]. The proposed model of affective acoustic ecology seems to follow a newly suggested framework for understanding and improving urban soundscapes [14] by presenting similar characteristics for the core structure of a sound- scape, i.e. SE's.

In the present work we aim to provide a stand- point in the direction of modeling affective acoustic ecology with respect to cities' ambient sound environment. We utilize findings from previous carried work [4, 12] and the pre-annotated SE's database IADS [15]. We employ a novel definition of SE's as the primitive elements of the soundscape and various signal and semantic based characteristics for the SE's.

The rest of the paper is organized as follows: in Section 2 an overview is included regarding soundscapes, emotion recognition from sound and affective acoustic ecology. In Section 3 the current modeling approach is presented. Section 4 concludes the paper and provides some reference for future work.

2 Existing research

2.1 Soundscapes

The term "soundscape" was initially introduced by P. Schaffer, referring to the "auditory properties of a landscape" [2]. Since then, various studies have been conducted regarding the quality [16], time complexity [3], categorisation according to noise type [17], relation to emotions [18] and other aspects of a soundscape.

The dependency of soundscape's quality to the semantic context of perceived audio stimuli has been shown in [16], stating that it is a subjective individual experience. This outcome clearly depicts the importance of the semantic context as a factor for defining the interaction with the audio environment, with respect to soundscapes. In [3], an investigation of the temporal complexity of a soundscape was performed. It was shown that a short-term characterization of a soundscape demands a 60 minutes recording [3] and thus the dependency of the duration of an audio stimulus is pointed out.

The categorization of a soundscape according to the noise type is performed in [17]. According to this work, a new categorization for urban soundscapes is proposed and regards two types of noise according to their technical characteristics, i.e. a) background noise, and b) transport noise. For this classification, five types of soundscapes were put forward, ranging from very quiet to very loud areas [17]. Additionally, in [18] the researchers employed a 2D model for evaluating listeners' emotion when they were surrounded by a soundscape, pointing out the connection of emotions and soundscapes. Their model lies in the continuous models abstract category, presented later in this paper.

In recent published research efforts, the connection of the sonic environment and the acoustic ecology has been stressed out [4]. This interdependence is apparent in real and synthetic worlds, where audio stimuli are present in almost every activity and audio-based interaction helps the optimized immersion into the virtual activities, respectively [19]. In the above work, the concept of affective acoustic ecology has been also introduced, which connects the acoustic ecology with the affective state of the listener.

Recently, a new framework for urban sound- scape improvement has been demonstrated [14]. In the corresponding published work, a soundscape is regarded as being composed by a mix of different sources. The attributes of the sources are both low level, i.e. technical characteristics like loud- ness, and high level, e.g. dominance. This dis- crimination of sources' attributes is in accordance to the one proposed by the affective acoustic ecology, presented later in this paper.

2.2 Emotions Modeling

Emotion recognition from sound can be considered as a feature extraction and classification task, with various techniques and algorithms employed for each chore [4], and discriminated in two major categories: a) emotion recognition from music, and b) emotion recognition from sound events.

Both of these categories share the emotion classification task in which different models exist and used. These can be grouped in two abstract categories: i) discrete, and ii) continuous models. Typical models of the former group are the basic emotions and group of adjectives [20, 21] and in the latter lie the dimensional representations/models of emotional states (arousal, valence, dominance etc.) [22].

Discrete models use specific words for referring to particular emotions, as illustrated in Figure 1. Although there is a scattering of information in the discrete models regarding the verbal description of emotions due to different words used for the same emotion [6], these models seem to offer and attractive basis for neuroscience researches [4].

In contrast, continuous models provide the representation of emotion as a resultant of continuous values from two or more dimensions that represent affective states like arousal or valence [4, 12]. Due to this approach, the resulting values can be mapped to specific verbal descriptions of emotions [4, 7], as illustrated in Figure 2.

Focusing on data with music content, typical utilised algorithms for classification are Sup- port Vector Machines (SVM), Gaussian Mixture Model (GMM), Decision Trees and Artificial Neural Networks (ANN) [7, 12, 23], whereas common extracted features regard energy, rhythm and timbral characteristics of the signal [4]. Published researches demonstrate relatively high accuracy results that can reach up to 85% [7]. Applications of emotion recognition from music include, but not limited to, new retrieval methods based on emotion and mood instead of the legacy "Artist / Album / Genre / Year" classification and affective music composition [8, 24]. Regarding emotion recognition from SE's, published researches are likely to utilise the continuous emotional model and accuracy results reach up to 88% for arousal [12] and 50% for valence [25]. ANN and SVM were amongst the employed algorithms for classification.

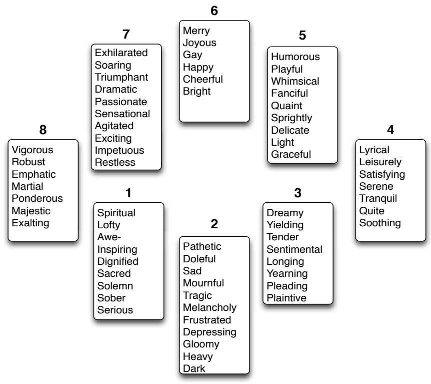

Figure 1. List of adjectives with 8 classes according to [21].

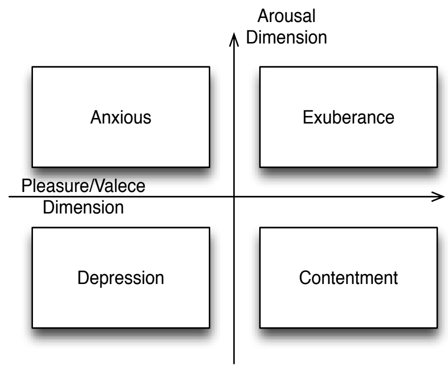

Figure 2. Arousal - Valence plane with mapped discrete emotions

2.3 Affective Acoustic Ecology

Following a previous published work [4], the concept of Affective Acoustic Ecology (AAE) enhances the Acoustic Ecology by inserting the affective state of the listener. Its main element is the SE, defined as an audio structure that contains:

1. a sound waveform

2. manifestation of source's instantaneous spatial position, relative to the receiver

3. duration

4. indication of sound's producing mechanism, e.g. impact, friction and other

5. evidence of vibrating objects' state, i.e. solid, liquid etc.

6. semantic content

AAE can be viewed as a superset of the Acoustic Ecology, by also considering emotional responses. These can be inferred from Sound Event Emotion Recognition, based on the aforementioned enhanced SE form.

Apart from 6, all the other SE's attributes can be measured with acoustic cues. A following work showed that it is possible to identify the arousal of the listener elicited from SEs by solely regarding acoustic cues [12]. The accuracy results demonstrated an accuracy of 85%, independently of the semantic content of the SE.

3 Proposed Approach

A soundscape contains both music and non- musical non-linguistic sounds. Considering the former kind, there is an ongoing research mentioned briefly in the previous section of this work. Regarding the latter kind, a listener can focuses on its "musical" characteristics, e.g. timbre, pitch etc.

While this could be valid and useful for musical sounds, it seems that it is rather the case that can provide useful information for audio interaction in a urban soundscape environment when referring to SEs. Also, this obscuration of information stands true for most of sound environments [4].

As a solution, "everyday listening" is used in a spontaneously manner which allows the concentration and focusing on various attributes of the source [4, 26]. These information can be safely considered of paramount importance in a urban soundscape since it could save the listener from hazard, e.g. the movement of a car moving to- wards the listener, or affect the quality of everyday life, e.g. urban ambient noise.

Thus, the information of "everyday listening" can provide to the receiver the identification of the sound sources and its immediate relation with them. In addition, and as demonstrated in [26], the listener can discriminate the nature of the source, its size and also can deduce data about the surroundings of the source. Such particulars are impulsively processed by the receiver and are likely to elicit emotions, especially when these interactions are viewed under a fashion of continuous appraisal of the stimuli [5].

As shown in previous section, emotion recognition from SEs can demonstrate relevant high ac- curacy results [12]. In addition, these outcomes are on a non-semantic context base which renders them applicable to all SEs. In an other work [4], a significant impact of SEs' semantic context to the receiver's elicited emotion has been revealed. In that work, relevant low accuracy results have been obtained.

Thus, these facts combined with the certainty that soundscapes are composed from SEs seem to imply that the receiver in a soundscape both analysing the semantic context and the technical characteristics of SEs. Moreover, the outcome of the appraisals from both semantic context and acoustic cues is affecting the receiver's emotion. This process can be viewed in Figure 3.

Consequently, and according to the definition of AAE, it seems that the concept of the affective state of the receiver can be studied with the application of the AAE in urban soundscapes investigation.

4 Conclusions & Future Work

AAE and the associated sound event structure seem to be eligible for application in soundscape investigation. Obtained results confirm the concept of AAE and are likely to suggest its incorporation in further researches.

Although already obtained results show a relevant high accuracy for emotion recognition from SEs, especially when the arousal of the listener is considered, a need for more investigation seems to be apparent when the valence is regarded. Also, there seems to be a lack of investigations with respect to continuous emotional responses when SEs are used.

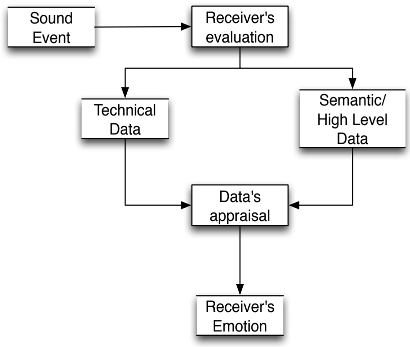

Figure 3. The process of emotion elicitation from SEs in a soundscape

References

[1] Kang J, Zhang M. (2010). Semantic differential analysis of the soundscape in urban open public spaces, Building and Environment, Vol. 45, No. 1, International Symposium on the Interaction between Human and Building Environment Special Issue Section, pp.150–157.

[2] Liu J, Kang J, Luo T, Behm H, Coppack T. (2013). Spatiotemporal variability of soundscapes in a multiple functional urban area, Landscape and Urban Planning, Vol. 115, No. 0, pp.1–9.

[3] Torija AJ, Ruiz DP, Ángel Ramos-Ridao (2011). Required stabilization time, short-term

variability and impulsiveness of the sound pres- sure level to characterize the temporal com- po- sition of urban soundscapes, Applied Acoustics, Vol. 72, No. 2–3, pp. 89–99.

[4] Drossos K, Floros A, Kanellopoulos NG., Affective acoustic ecology: towards emotionally enhanced sound events. In: Proceedings of the 7th Audio Mostly Conference: A Conference on Interaction with Sound, AM ’A12, 26–28 Sept. 2012, Corfu, Greece, New York, NY, USA: ACM; 2012. pp. 109–116.

[5] Scherer K, Dan E, Flykt A (2006). What Determines a Feeling’As Position in Affective Space? : A case for appraisal, Cognition & Emotion, Vol. 20, No. 1, pp. 92–113.

[6] Juslin PN, Laukka P. Communication of emotions in vocal expression and music performance: different channels, same code?, Psychological Bulletin, Vol. 129, No. 5, pp. 770– 814.

[7] Lu L, Liu D, Zhang HJ (2006). Automatic mood detection and tracking of music audio signals, IEEE Transactions on Audio, Speech, and Language Processing, Vol. 14, No. 1, pp. 5–18.

[8] Zhang S, Tian Q, Jiang S, Huang Q, Gao W., Affective MTV analysis based on arousal and valence features. In: Multimedia and Expo, 2008 IEEE International Conference on, ICME 2008, 23–26 June 2008, Hannover, Germany. p. 1369–1372.

[9] Li T, Ogihara M., Content-based music similarity search and emotion detection. In: IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP ’04, 17–21 May 2004, Montreal, Canada. pp. V– 705–8 vol.5.

[10] Yang YH, Liu CC, Chen HH., Music emotion classification: a fuzzy approach. In: Proceedings of the 14th annual ACM international conference on Multimedia. MULTIMEDIA ’A06, 22-228 Oct, 2006, Santa-Barbara, CA, U.S.A., New York, NY, USA: ACM; 2006. p. 81–84.

[11] Yeh CH, Lin HH, Chang HT., An efficient emotion detection scheme for popular music. In: IEEE International Symposium on Circuits and Systems, ISCAS 2009, 24–27 May, 2009, Taipei, Taiwan. pp. 1799–1802.

[12] Drossos K, Kotsakis R, Kalliris G, Floros A. Sound Events and Emotions: Investigating the Relation of Rhythmic Characteristics and Arousal. In: Proceedings of the Fourth International Conference on Information, Intelligence, Systems and Applications, IISA2013, 10–12 July, Piraeus, Greece, 2013.

[13] Schuller B, Hantke S, Weninger F, Han W, Zhang Z, Narayanan S., Automatic recognition of emotion evoked by general sound events. In: IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP ’A12, 25–30 March, 2012, Kyoto, Japan. pp. 341–344.

[14] Jennings P, Cain R. (2013). A framework for improving urban soundscapes, Applied Acoustics, Vol. 74, No. 2, Applied Soundscapes: Re- cent Advances in Soundscape Research, pp. 293–299.

[15] Bradley MM, Lang PJ., 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.: NIMH Center for the Study of Emotion and Attention. 2007.

[16] Hall DA, Irwin A, Edmondson-Jones M, Phillips S, Poxon JEW. (2013). An exploratory evaluation of perceptual, psychoacoustic and acoustical properties of urban soundscapes, Applied Acoustics, Applied Soundscapes: Recent Advances in Soundscape Research, Vol. 74, No. 2, pp. 248–254.

[17] Lebiedowska B. (2005). Acoustic background and transport noise in urbanised areas: A note on the relative classification of the city sound- scape, Transportation Research Part D: Transport and Environment, Vol. 10, No. 4, pp. 341– 345.

[18] Cain R, Jennings P, Poxon J. (2013). The development and application of the emotional dimensions of a soundscape, Applied Acoustics, Applied Soundscapes: Recent Advances in Soundscape Research, Vol. 74, No. 2, pp. 232–239.

[19] Schnall S, Hedge C, Weaver R. (2012). The Immersive Virtual Environment of the digital fulldome: Considerations of relevant psychological processes, International Journal of Human–Computer Studies, Vol. 70, No. 8, pp. 561–575.

[20] Ekman P. (1992). An argument for basic emotions, Cognition & Emotion, Vol. 6, No. 3–4, pp. 169–200.

[21] De Silva LC., Audiovisual emotion recognition. In: IEEE International Conference on Systems, Man and Cybernetics, 10–13 Oct, 2004, Hague, Netherlands, pp. 649–654.

[22] Smith CA, Ellsworth PC. (1985). Patterns of cognitive appraisal in emotion, Journal of Personality and Social Psychology, Vol. 48, No. 4, pp. 813–838.

[23] Li T, Ogihara M., Detecting emotion in mu- sic. In: Proceedings of the International Sympo-

sium on Music Information Retrieval, IS- MIR2003, 26–30 Oct, 2003, U.S.A.

[24] Casacuberta D. (2004). DJ el Niño: expressing synthetic emotions with music, AI & SO- CIETY, Vol. 18, No. 3, pp. 257–263.

[25] Schuller B, Hantke S, Weninger F, Han W, Zhang Z, Narayanan S., Automatic recognition of

[26] emotion evoked by general sound events. In: IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP '12, 25–30 March, Kyoto, Japan, pp. 341–344.

[27] Gaver WW. (1993). What in the world do we hear? an ecological approach to auditory event perception, Ecological Psychology, Vol. 5, pp. 1–29.