Temporal Characteristics of Emotionally Annotated Sound Events

K. Drossos, A. Floros :: Audiovisual Arts Dept - Ionian University - Greece :: {kdrosos, floros}@ionio.gr

Introduction

Sound can transmit or evoke emotions, either in its structured form (i.e. music) or in a non linguistic, non musical structure. Although many studies and researches concern the connection between music and emotions, there are little regarding sound events and emotions.

In this poster a study of the temporal/rhythmic characteristics of emotionally annotated sound events is presented. The main aim is the possibility of sound events emotionally categorisation, regarding only their aforementioned characteristics and the sound events used are of the IADS database [1].

IADS

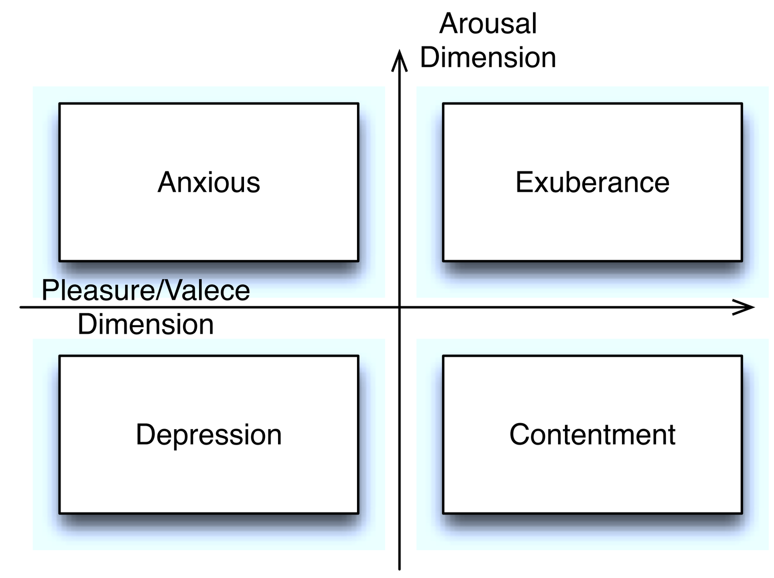
The sound events database used is the International Affective Digital Sounds [1] as the only available sound events database which has its events emotionally annotated. Each IADS file includes its arousal and valence rating and their semantic content span across different activities, like:

- Applauses
- People vomiting
- Air raid alarm
- Gunshots
- Fizz water's sound

Emotions Model Used

The emotions model used is the Arousal - Valence plane (AV plane). A continuous model, in contrast to discrete ones (like the basic emotions model), in which the emotion is represented as a resultant of two states (arousal and valence).

Thus, each emotional state can be represented as a set of two values. In order to achieve better consistency with discrete emotional models, values for similar emotional states can be grouped and assigned to specific emotions as shown in the next Figure.



In the present work, the values on the AV plane were clustered to four classes, each for every quadrant of the AV plane.

Previous Researches

Existing researches, e.g. [3], have used time characteristics amongst other features for emotion classification of music signals. In the field of emotion recognition from sound events, there is no other research concerning solely the temporal characteristics, to authors' knowledge.

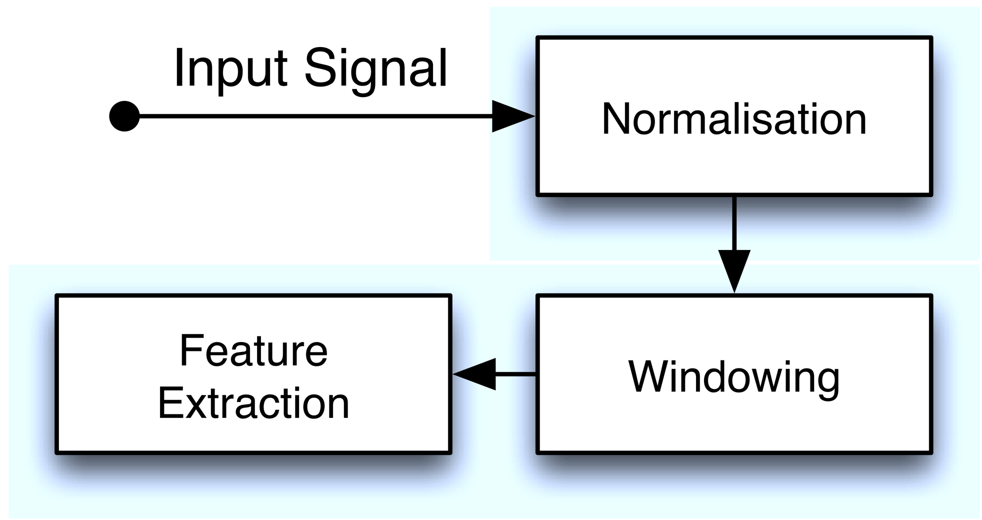
Typical parameters used include:

- Spectrum's attributes, e.g.:
  - Centroid
  - Flux
  - Roll off
- Energy of the signal
- Zero crossing rate
- Strength of beat
- Fundamental frequency estimation
- Spectral flatness

The accuracy of the classification for the aforementioned researches reaches to 80%. Nevertheless, in recent researches regarding exclusively sound events and not music, classification results using various characteristics did not exceed 60%.

Features Extraction & Pre-processing

Prior to sound events emotion categorisation, feature extraction and pre-processing tasks were conducted according to the next Figure. **Windowing** process **values** were: width 1sec, 20% overlap. Extracted features are presented to the following Table.

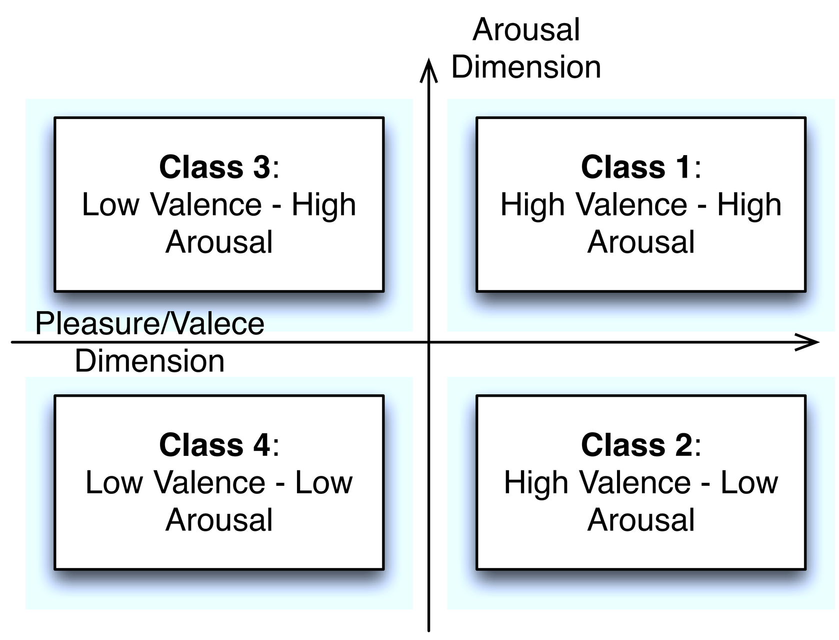


Extracted Features	
Pulse clarity	Onsets
Event density	Fluctuation

From each feature a set of statics measures was calculated. These are:

- Mean value
- Gradient
- Skewness
- Standard deviation
- Kurtosis

In addition, the ground truth data, obtained by the emotional annotation of the IADS database, was added using 4 clusters / classes. The resulting feature set consisted of 21 features.



Results

Classifier	Accuracy
OneR	43,11%
J48	41,91%
Random Forest	46,1%
NB Tree	44,91%

The results obtained using WEKA [4]. For each classifier used, the achieved accuracy is listed.

The input parameters were the extracted features and the classification class was the Arousal-Valence clustering, based on the annotation of the IADS database.

As the results show, the maximum accuracy achieved using the Random Forest classifier.

The accuracy values are below 50%. This fact clearly depicts that using solely the temporal characteristics of the sound events can not lead to classification on both dimensions of the AV plane.

Conclusions & Future Work

There is not a good estimation of the emotional classification for sound events when only temporal characteristics are used. Possible reason could be the strong semantic content of the IADS database.

Nevertheless, from empirical evidence, there might be a strong connection with only one of the dimensions in the AV model. Thus, more research in the connection between the temporal features and the elicited emotions seems to be needed.

References

[1] 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

[2] Garner, T., & Grimshaw, M. , "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

[3] 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.

[4] M. Hall, et al., "The WEKA data mining software: an update," SIGKDD Explor. Newsl., vol. 11, pp. 10-18, 2009.