AUTOMATIC MUSIC PLAYLIST GENERATION USING AFFECTIVE COMPUTING TECHNOLOGIES

Darryl Griffiths, Stuart Cunningham and Jonathan Weinel

Creative and Applied Research for the Digital Society (CARDS), Glyndŵr University, Wrexham, UK

S09001315@mail.glyndwr.ac.uk

{s.cunningham | j.weinel}@glyndwr.ac.uk

ABSTRACT

This paper discusses how human emotion could be quantified using contextual and physiological information that has been gathered from a range of sensors, and how this data could then be used to automatically generate music playlists. I begin by discussing existing affective systems that automatically generate playlists based on human emotion. I then consider the current work in audio description analysis. A system is proposed that measures human emotion based on contextual and physiological data using a range of sensors. The sensors discussed to invoke such contextual characteristics range from temperature and light to EDA (electro dermal activity) and ECG (electrocardiogram). The concluding section describes the progress achieved so far, which includes defining datasets using a conceptual design, microprocessor electronics and data acquisition using MatLab. Lastly, there is brief discussion of future plans to develop this research.

KEYWORDS

Digital music; playlist generation; affective computing; fuzzy logic; neural nets

1. INTRODUCTION

Affective computing is a research area that involves the study and development of intelligent systems and devices. Such technology is ideally able to recognise and interpret the emotional state of humans, determine their surrounding environment and adapt its behaviour to them by outputting an appropriate response for those emotions. Emotional content can be gathered and estimated by measuring the users physiological attributes such as facial expressions, movement, EDA (Electro Dermal Activity) and ECG (electrocardiogram). Context awareness will adjudicate "where the user is" by quantifying real-world data and device data such as temperature, humidity, noise, accelerometer data and compass bearings. Such measurements should disclose the users situation, and by combining this information with affective data may build a picture of "what they are doing in that situation".

Within the field of affective computing, this research aims to devise a system for automatic selection of musical playlists. It is intended that this will be achieved through the use of an array of sensors, which suitably interpret human emotional state and context, in order to inform the selection of audio sound files. The perceived benefits of such a system could elicit a more personal and unique user experience by providing motivation, joy and relaxation to their everyday lives. This framework could also be generalised and adapted to other personal and unique experiences.

2. RELATED WORK

This section describes the work that has been done in the field of affective computing, such is relevant to this research project. In particular I will look at existing developments of automatic

emotional recognition systems that detect human emotional responses through the use of sensors and intelligent systems to generate music playlists. Music retrieval methods are also discussed, which compliment automatic playlist generators by extracting audio features to determine musical properties such as melody and similarity in order to serve the user's needs and satisfaction within multimedia applications. In the section below, I identify some key existing research and methodologies which is likely to inform this project as it develops.

2.1. Affective Playlist Generators

Current research in this field explores automatic recognition of emotion evoked by general sound events [1]. This looks at eliciting emotional states from general sounds from different areas of the daily human environment using a sophisticated model to establish a reliable 'ground truth'. The 'empathy machine' [2], is an application of affective computing in mediating live human-to-human interactions. This system uses 'automatic facial expression recognition' (FER) combined with a rule-based approach to identify the emotional state of a users conversation partner. This then generates non-disrupted 'emotional music' to complement the user's conversational partners expressive state. The addition of music to the interaction, much like the addition of music to a film, compliments the visual emotional signals that the user receives when observing the facial and gestural expressions of the associate. Other research seeks to develop a genetic algorithm based on emotional recognition [3], is a discrete emotion recognition system that collects physiological signals by combining biosensors and music content to elicit four different affective states. Reynolds et al. [4], highlights the need for contextual and environment information, which defines the listeners scenario by means of location, activity, temperature and lighting, etc. This information is then aggregated to create a personalised automatic playlist generator for large music collections. Chung-Yi Chi [5], uses a reinforcement learning approach to emotion-based automatic playlist generation by collecting the users behaviour in music playing, such as rating, skipping and replaying in order to learn the users current preferences. Reinforcement learning is adopted to learn the users preferences, which are used to generate personalised playlists.

2.2. Audio Feature Extraction

Current research in this field includes melody-based retrieval in audio collections by means of a mid-level representation [6]. This approach supports audio, as well as symbolic queries and ranks results according to salient melodic similarity to the actual query. Overall results were positive and fast both with audio and symbolic queries, making the proposed system suitable for retrieval in large databases. A multiple feature model for musical similarity retrieval [7] describes how low-level features combined with appropriate classification schemes are often satisfactory, but in some cases, not being sufficient enough to properly identify similarities between songs due to the "fuzzy" nature of music similarity, which varies subjectively from one person to another. This uses a set of low-acoustic features to mimic human perception with promising results with a data base of 15,000 musical items. Tempo and beat estimation of musical signals [8] is fundamental in automatic music processing and applications. It presents an automatic tempo tracking system that processes audio and determines the beats per minute and temporal beat location from a large multi-genre data base. The algorithm involves extracting onsets, a periodicity detection block and the temporal estimation of beat locations, producing a global recognition rate of 89.7%. Unifying low-level and high-level music similarity measures [9], proposes three distance measures based on audio. First a low-level tempo based measure; second, a high-level semantic based on inference of different musical dimensions that include genre, culture, moods, instruments, rhythm and tempo; third, being a hybrid measure combining the above-mentioned with two existing measures: a Euclidean distance based timbral, temporal and tonal descriptors and a timbral distance based on a single Gaussian Mel-Frequency Cepstral Coefficient (MFCC) modelling.

The examples above provide a range of methodologies for interaction between human emotion and music. Fundamentally, this research seeks to achieve similar aims, so it is likely that I will draw upon some of these models to reinforce this project. The examples above have also provided a positive insight of how the proposed system framework could be planned and implemented.

3. PROPOSED SYSTEM

This project will investigate and evaluate how contextual data and human emotion can be estimated based on data being discretely gathered from a range of different sensors. These sensors will consist of biometric, environmental and mobile device sensors. The biometric sensors will define a users emotional state by measuring emotional properties such as arousal, temperature and activity via skin conductance. The environmental sensors will gather variable data such as outdoor light, sound, temperature and humidity in order to establish boundaries within the system. To compliment such a system, a sufficient range of low/mid level audio feature extractions will be performed on a selection of music tracks in order to define properties such as beats per minute, amplitude range, metadata, etc. to establish the actual content of the data. Expert models of human emotion will be constructed to allow physiological and contextual data to be synthesized and evaluated in order to accurately predict human emotion by classifying such data into a predetermined state. The feasibility of using biometric equipment during the research will be evaluated initially and, if not feasible, sample data sets will be acquired.

The proposed system will acquire physiological and contextual data to determine an emotional state by inputting both entities into a FIS (Fuzzy Inference System) [10]. The defuzzification output will be mapped to a circumplex model of affect in order to quantify the users mood. Such a model usually consists of a horizontal axis representing the valence dimension and the vertical axis representing the arousal dimension [11]. Such a model is illustrated below in Figure 1.

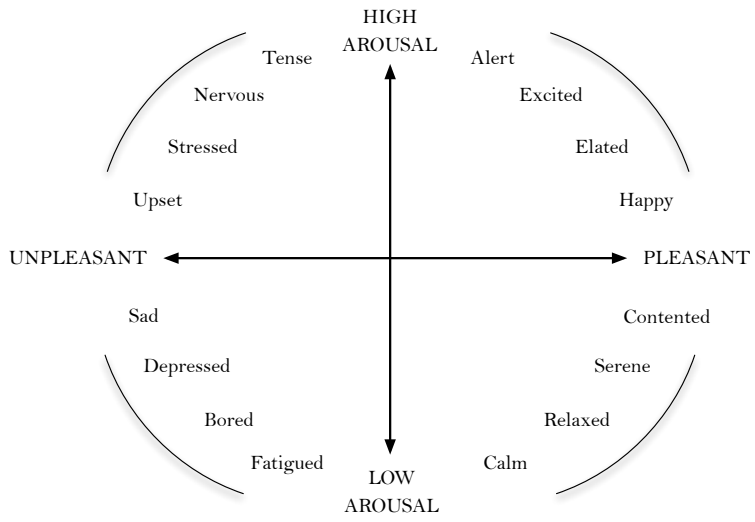

Figure 1. A general circumplex model of emotion.

Music tracks will be defined in corresponding musical categories by performing content analysis before an emotional state has been determined. This can be achieved by extracting low/mid level audio features such as beats per minute, melody content and similarity. Based on these three entities: human emotion, musical content and physical context, a decision will be made, and a music playlist will be generated. A further additional entity may consist of a user training system that will acknowledge an incorrect choice of music for the occasion using a Neural Network [12]. An overview for the system is outlined below in Figure 2.

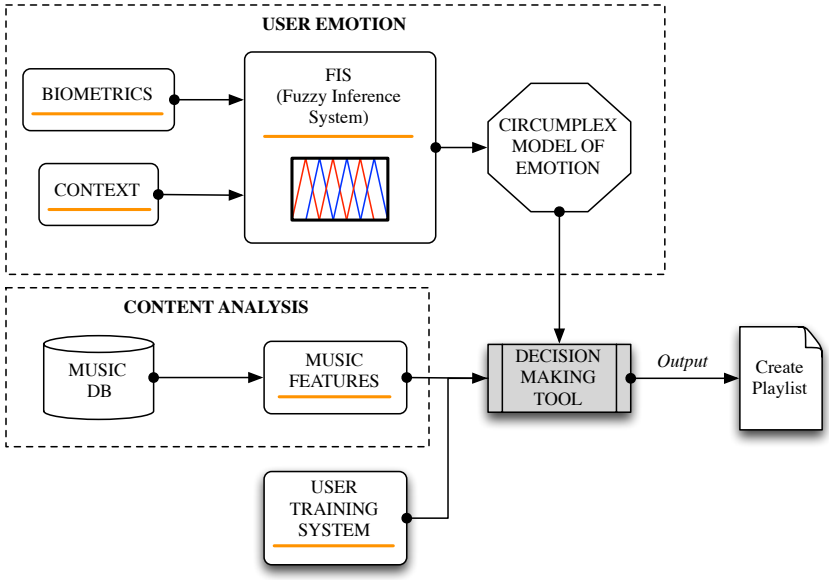

Figure 2. A system overview of how a decision could be made.

User testing will be carried out in order to determine if the system is fit for purpose. An extensive range of qualitative (surveys and interviews) and quantitative (objective data from sensors and statistical tests) methods of testing will be used. This will provide confirmation that the system's initial design principles have been realised as intended, or reveal any errors which can then be addressed and revised accordingly. A further stage of testing, evaluation and revision will then be undertaken.

4. Current & Planned Work

The proposed system was to record and group data acquired from a range of sensors to determine a users affective state. Firstly, a set of hypothetical scenarios of a typical user were created in order to establish how the contextual and physiological sensors could be combined to accurately determine an emotional state. The scenarios ranged from recreational active sports to relaxing on a sunny beach (eg, Figure 3). Consequently, this also defined what datasets will be evaluated within the FIS (Fuzzy Inference System).

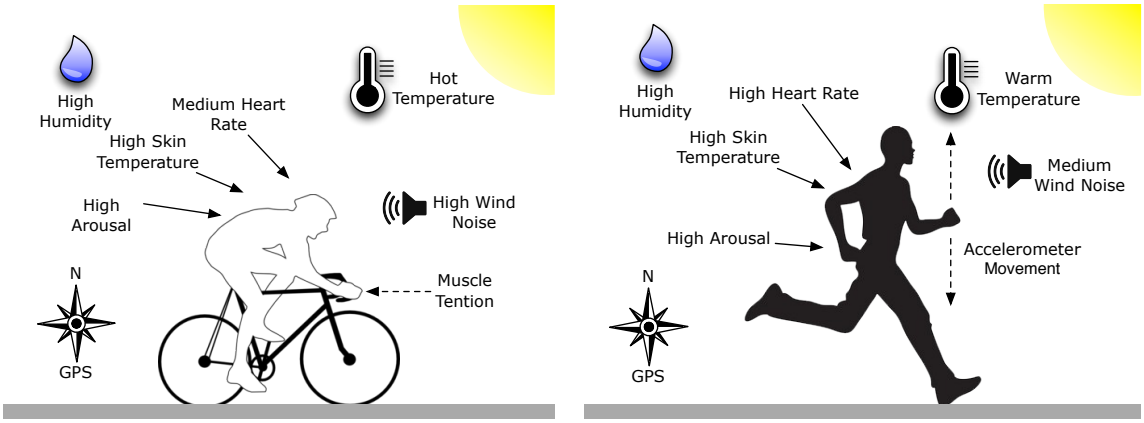

Figure 3. Scenarios: A typical user cycling and jogging.

Progress has also been made by experimenting with external sensors such as temperature, humidity and light using an open-source electronic microprocessor board named: Arduino [13]. By programming the Arduino using the language C++, each sensor was able to be configured and defined to output and display its own unique format i.e. temperature in deg Celsius and humidity as a percentage (eg, Figure 4), etc. Once this has been carried out for each external sensor, they will be combined into one or more portable units for data logging contextual information based on the conceptual user scenarios that were developed initially. The data logging instruments acquired so far are outlined below in Figure 4.

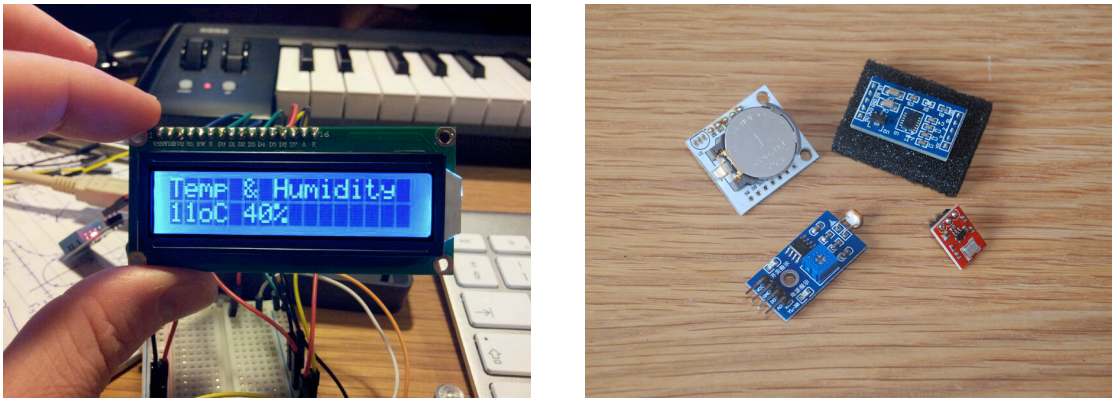

Figure 4. Temperature and humidity readings and Arduino compatible external sensors.

The next step was to capture the sensor data in real time within MatLab (eg, Figure 5) over a serial connection. This enabled the data to be plotted onto a 2 dimensional graph over time (*t*). So far data readings have been acquired from temperature, humidity and light, but more data acquisition is needed from newly attained external sensors such as a sound, a real-time clock, a 1.5 to 6g accelerometer as illustrated in Figure 3. Other sensors such as a CMOS camera and a GPS module will possibly be considered further into the project.

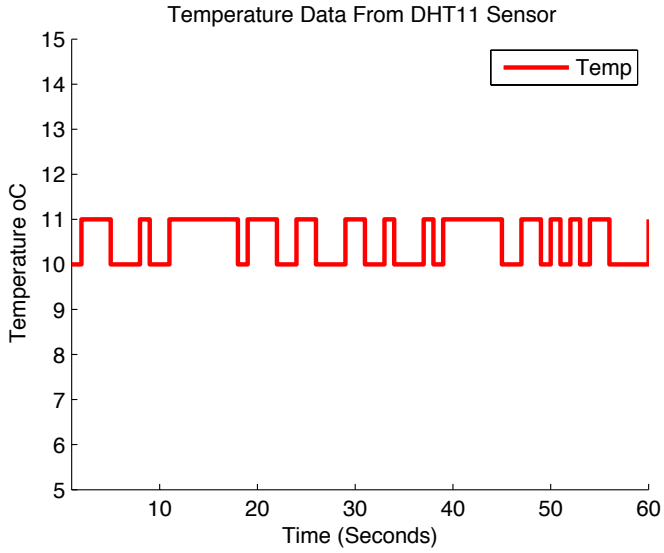

Figure 5. Data acquisition with a DHT11 temperature sensor.

After acquiring a the required data from the contextual domain, data acquisition of the physiological domain will be combined with the contextual data in the FIS (Fuzzy Inference System). At this stage, a good level of knowledge should have been acquired both contextually and physiologically, in order to establish what combination of data sets will be most effective to the overall "decision-making" of the system, as at this stage, this is unknown.

Whilst dealing with the contextual and physiological/emotional domain, a system framework will be designed throughout the course, followed by the implementation of a full prototype "decision making" system, which will include: a user testing phase, system revisions, further user testing phases and further system revisions until the system is fit for purpose.

REFERENCES

[1] B. Schuller, "Automatic recognition of emotion evoked by general sound events", in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE Int. Conf.*, Tech. Univ., Munchen, Germany, 2012, pp. 341-344.

[2] N. Kummer, "The EMPATHY MACHINE", in *Systems, Man, and Cybernetics (SMC), 2012 IEEE Int. Conf.*, Univ. British Columbia, Canada, 2012, pp. 2256-2271.

[3] N. Xiaowei, "Research on genetic algorithm based on emotion recognition using physiological signals", in *Computational Problem-Solving (ICCP), 2011 Int. Conf.*, Chongqing Three Gorges Univ., China, 2011, pp. 614-618.

[4] G. Reynolds, D. Barry, T. Burke, E. Coyle, "Towards a personal automatic music playlist generation algorithm: The need for contextual information", in *Proc. of the 2nd. Audio Mostly Conf.: interaction with sound*, Fraunhofer institute for Digital Media Tech., Germany, 2007, pp. 84-89.

[5] Chung-Yi Chi, "A reinforcement learning approach to emotion-based automatic playlist generation", in *Tech. and Applications of Artificial Intelligence (TAAI), 2010 Int. Conf.*, 2010, Nat. Taiwan Univ., Taiwan, pp. 60-65.

[6] M. Marolt, "A mid-level representation for melody-based retrieval in audio collections", *IEEE Trans. on multimedia*, vol. 10, pp. 1617-1625, Dec. 2008.

[7] E. Allamanche, *et al*., "A multiple feature model for musical similarity", in *ISMIR 2003 Conf. Proc*., Germany, Oct. 2003. pp. 1-2.

[8] M. Alonso, B. David, G. Richard, “Tempo and beat estimation of musical signals”, in *ISMIR 2004 Conf. Proc*., France, 2004, pp. 1-6.

[9] S. Joan R *et al*., “Unifying low-level and high-level music similarity measures”, *IEEE Trans. on multimedia*, Barcelona, Spain, Vol. 13, pp. 687-701.

[10] C. Schmid, “Introduction to fuzzy techniques.” Internet: http://www.atp.ruhr-uni-bochum.de/rt1/syscontrol/node111.html, May. 9, 2005 [Mar. 26, 2013].

[11] J. Posner, J. Russell, B. Peterson, “The circumplex model of affect.” Internet: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2367156/, May. 5, 2008 [Mar. 26, 2013].

[12] L. Smith, “Introduction to neural networks.” Internet: http://www.cs.stir.ac.uk/~lss/NNIntro/InvSlides.html, Aug. 4, 2008 [Mar. 26, 2013].

[13] Arduino, “Homepage.” Internet: http://www.arduino.cc, 2013 [Mar. 26, 2013].