Audio Engineering Society

Convention Paper 7657

Presented at the 126th Convention
2009 May 7–10 Munich, Germany

Elevator: Emotional Tracking using Audio/visual Interaction

BasillisPsarras [1], Andreas Floros[2], and Marianna Strapatsakis [3]

[1] Dept. of Audiovisual Arts, Ionian University, Corfu, 49 100, Greece
billhaze85@gmail.com

[2] Dept. of Audiovisual Arts, Ionian University, Corfu, 49 100, Greece
floros@ionio.gr

[3] Dept. of Audiovisual Arts, Ionian University, Corfu, 49 100, Greece
maristra@ionio.gr

ABSTRACT

The research interest on modeling everyday human emotions and controlling them through typical multimedia content (i.e. audio and video data) has recently increased. In this work, an interactive methodology is introduced for detecting, controlling and tracking emotions. Based on the above methodology, an interactive audiovisual installation termed "Elevator" was realized, aiming to analyze and manipulate simple emotions of the participants (such as anger) using simplified emotion detection audio signal processing techniques and specifically selected combined audio/visual content. As a result, the human emotions are "elevated" to pre-defined levels and appropriately mapped to visual content which corresponds to the emotional "thumbnail" of the participants.

1. INTRODUCTION

Emotional expression represents one of the most important aspects of human everyday life and an important factor in communication. Invoking emotions through music has recently attracted the research interest from both the analysis and synthesis sides [1] (i.e. for algorithmically extracting emotions from music signals and for parametric music synthesis taking into account the desired emotional target). Moreover, a number of published works have investigated the retrieval and definition of the emotional information contained in human speech [2], aiming to analyze and determine the fundamental characteristics of speech that also carry the affective voice content.

This work goes a step further focusing on identifying, controlling and representing the intensity of human emotions through combined audiovisual means and audio signal processing methods. More specifically, in this work we define three different and distinct processes:

a) Emotions' intensity identification (or tracking) that is performed in real-time through audio recording

and appropriate audio signal processing and information retrieval techniques.

b) emotional intensity control, a process that aims to "elevate" the emotion intensity to a specific value (hence the term "Elevator" in the title of the work) and

c) emotional representation that attempts to convert the elevated emotional state into an audio/visual snap-shot (a kind of emotional "thumbnail") that uniquely and efficiently describes the current emotional conditions.

For the purposes of this work, the above three processes were realized and integrated within an interactive audiovisual installation framework. Interactive audiovisual installations represent a new form for realizing complex human-oriented experiments that most commonly involve human-machine interactions, while they are also employed by modern artists as a new artistic expression approach [3]. Hence, new terms and ideas originating from the general concept of interaction are nowadays frequently used to provide novel means of audio and visual production, where the audience is actively participating in the production process [4], [5].

"Elevator" (see Fig. 1) represents a typical example of an interactive audiovisual installation designed and developed during this work. The main purpose of Elevator is to enable and control in real-time specific human emotions using appropriately combined audiovisual content, capture the elevated emotion and finally create a visual output that correspond to the intensity of the emotion. For the purposes of this work, we focus on a very common emotion that extensively characterizes every being in nature: anger.

Figure 1 Elevator logo

The rest of the paper is organized as following: In Section 2, a brief summary of the relationship between audio signals in general and the emotions is presented, focusing mainly on existing methods that are employed for extracting the affective audio content. Next, the analytic description of the "Elevator" interactive installation is provided in Section 3, followed by a brief analysis of the functional and behavioral observations made during an installation exhibition. Finally, Section 4 concludes this work and accents further interaction and audio/visual enhancements that may be integrated in the "Elevator" platform in the future.

2. EMOTIONS AND AUDIO OVERVIEW

Although the scientific outcomes of the first research efforts for studying and analyzing the impact of audio and music signals to human emotions were published in the late 1930s [6], the research on the relation between music and emotions has increased substantially recently [7] due to the fundamental question: "why is music so closely related to emotions?".

The initial approach on studying music emotions was to identify (and possibly quantify) the emotional variations that are induced by specific features and parameters of music signals, such as tempo and rhythm [8]. Additional experiments have considered systematically varied compositions that were heard by subjects that rated the perceived emotional feelings [9].

In general, in order to analytically extract the relationship between music and emotions, efficient and accurate modeling of emotions is required. Towards this aim, there are a number of theories and interdisciplinary approaches for emotion research. For example, a simplified approach is to use a limited set of "basic" perceived emotions, such as happiness, anger, sadness fear and tenderness, which can be clearly distinguished by the human subjects. Although such an approach oversimplifies the overall concept, it has been shown that it provides meaningful result for perceived (and not induced) emotions.

An alternative affective modeling approach is to express emotions in multidimensional spaces as vectors of basic emotional parameters. For example, in [10], the activity-valence space was considered for analytically expressing discrete emotions as points in the two-dimension space.

Provided that musical emotions are modeled based on one of the above theories, an analytic mapping of the modeled emotions to audio / music features must be established. Recent works have investigated the above

issue by experimentally-derived weighted mappings between emotions and audio features using regressive models [11]. As it was mentioned previously, this work is focusing only on elevating and controlling a basic emotion, the anger. Hence, simplified, one dimensional affective models need to be employed that provide the ability of measuring the intensity of anger. As it will be presented in the following paragraph, such a simplified model was derived here by performing a number of subjectively controlled experiments that resulted into a low-complexity mapping of common audio signal parameters with the anger intensity value.

It must by noted that although most of the previously published studies have exclusively considered the impact of music signals to emotions, the relationship of the human audible events in general and the emotions they induced has not been yet exploited [9]. In this work, audio signals are not constrainedly music content, but they are synthesized in terms of audio events appropriately placed in time and 3D space and additionally integrated with visual output. This renders the parameterized emotional elevation and control a very flexible process, allowing the employment of simplified emotional to audio mapping.

3. ELEVATOR DESCRIPTION

During this work, the "Elevator" interactive environment was typically designed for and installed in a close room with dimensions 2x3x3 meters. A microphone installed at the roof of the room was responsible for recording and providing the audio input and feedback to the emotional level detection subsystem. Audio and video playback was performed from typical stereo loudspeakers and high-resolution video back-projection projection systems.

The complete architecture of the installation is shown in Fig. 2. The software implementation of all the required real-time algorithms (such as anger level detection and visual thumbnail synthesis) was performed on the Processing open source platform [12], which represents a powerful software sketchbook and professional production tool used in many fields of audio and image signal processing, science/technology and arts. The selection of Processing induced significant reduction on the final real-time implementation complexity and allowed the concurrent processing and control of both audio and visual playback processes.

The purpose of the video projection systems is to display appropriate visual content that is obtained from a multimedia database developed during this work, in order to elevate the anger feeling by the pre-defined value. This visual content is mainly static, (it was pre-designed and mapped to specific anger intensity values). However, an additional adaptive visual synthesis algorithm was also implemented for creating more appropriate visual representation by altering parameters such as brightness, colors etc. The visual content representation is combined with relevant audio content, also stored in the multimedia database, which is reproduced by the loudspeakers shown in Figure 2.

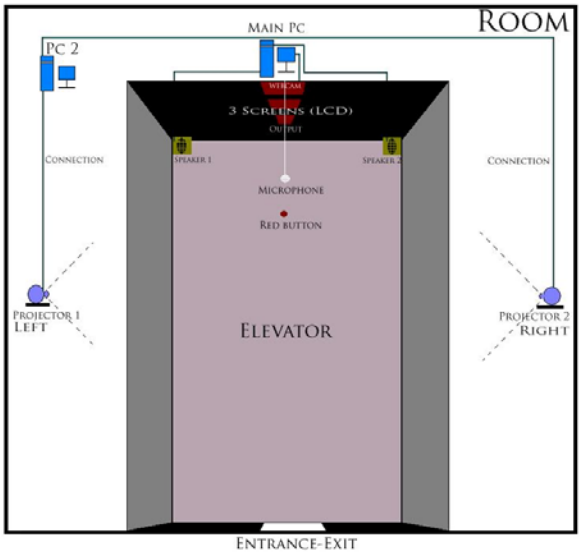

Figure 2 The Elevator installation architecture

The anger elevation process is adaptive, with the necessary feedback provided by the audio signals captured by the microphone installed in the room. In order to obtain adequate audio feedback, the participants are encouraged and motivated to express their thoughts prior to entering the installation room.

The emotional (anger) detection level subsystem is responsible for processing the microphone input providing feedback that contains single anger intensity values. More specifically, as it was previously mentioned, the emotional modeling technique employed is applied on an one dimension vector space that corresponds to anger intensity values ranging from zero (no anger) to 255 (highest modeled anger value). The instantaneous anger intensity value is derived by

detecting specific spectral / envelope characteristics of the audio input that were found to be optimum for the purposes of this work. The selection of the above characteristics was based on a number of tests that were performed by recording speech with different, pre-defined levels of anger using qualified actors.

Figure 3 shows a typical snap-shot of the visual representation projected on the left wall of the installation, while in Figure 4, the direct view to the central wall of the installation is displayed. In the latter case, the three LCD monitors are responsible for creating the complete anger visual thumbnail.

Figure 3 The installation left wall back projection system

Figure 4 The installation central wall

4. RESULTS

During this work, a number of subjective tests have been performed for verifying the accuracy of the emotional control and visual representation processes. For realizing this sequence of tests, a number of participants have used and interact with the Elevator installation. For each participant, the anger elevation value was pre-defined, while the initial anger-related affective condition of all participants was considered to be zero (that is zero anger intensity value). In order to achieve the pre-defined anger-elevation level, appropriate audiovisual content was presented (as explained in the previous Section), and the anger visual snapshot was finally rendered.

Fig. 5 shows a typical visual output produced by the Elevator installation. As it was mentioned previously, the visual output strongly depends on the measured anger value and can be considered as the emotional thumbnail of the participant affective condition that was evoked by the controlled and combined audio and visual reproduced signals.

In order to verify the accuracy of the final visual output representation, the visual emotional thumbnails produced by the elevator installation in all test cases described previously were given to human subjects for grading the anger intensity they believed that each particular thumbnail represented. Six different anger values were available, ranging from zero (no anger) to five (highest anger intensity value). In practice, each of these values was linearly mapped to the pre-defined levels of anger intensity. Hence, these tests assessed the accuracy of the achieved anger elevation value and the corresponding emotional thumbnail mapping. After summarizing the results obtained through these tests, it was found that nearly the 90% of the human subjects had graded the emotional thumbnails in fine-accordance with the original pre-defined anger elevation level.

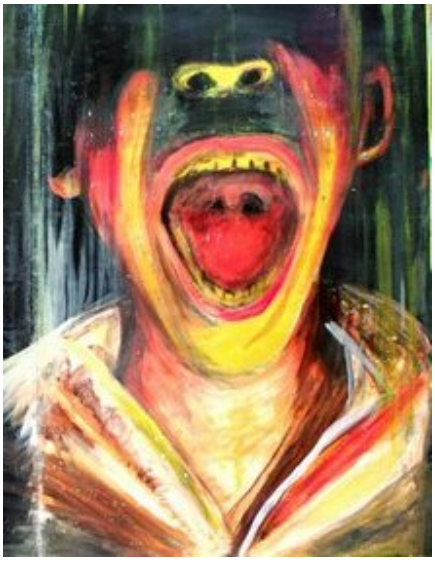

Figure 5 A typical anger emotional thumbnail

5. CONCLUSIONS

In this work, an interactive audiovisual installation called “Elevator” is presented which aims to detect, control and track human emotions (focusing particularly

on anger, a very common everyday life emotion). The detection and control of anger is based on a novel approach which employs the reproduction of appropriate and dynamic visual and audio content for elevating the desired anger intensity value to the participants. The above anger elevation process is adaptive, with the appropriate feedback provided by the anger detection level subsystem which employs signal processing techniques in the frequency domain for deriving the instantaneous anger intensity value.

The achieved measured levels of anger are finally mapped to adaptive visual content that represent the thumbnail of the emotional state of each participant. After a sequence of subjective tests, it was shown that the accuracy of the achieved anger elevation value and the corresponding emotional thumbnail mapping is significantly high. This conclusion also verifies the efficiency of the overall anger detection and control process.

It is the authors' near future intention to further exploit the derived results and conclusions of this work for developing an adaptive version of interaction that takes into account wide-spread and commonly met audio and visual-based emotion-triggering conditions (i.e. visual content and sounds in noisy urban environments), as well as to consider enhanced 3D / multichannel audio and video playback techniques.

6. REFERENCES

[1] A. Friberg, "Digital Audio Emotions: An Overview of Computer Analysis and Synthesis of Emotional Expression in Music", In Proc. of the 11th International Conference on Digital Audio Effects (DAFx-08), Espoo, Finland, Sept. 2008.

[2] M. Kienast and W. F. Sendlmeier, "Acoustical analysis of spectral and temporal changes in emotional speech", In Proc. of the ISCA ITRW on Speech and Emotion, Newcastle, Sept. 2000, pp. 92 – 97.

[3] D. Birchfield, K. Phillips, A. Kidané and D. Lorig, "Interactive Public Sound Art: a case study", In Proc. of the 2006 International Conference on New Interfaces for Musical Expression (NIME06), Paris, France, 2006.

[4] D. Birchfield, D. Lorig, and K. Phillips, "Network Dynamics in Sustainable: a robotic sound installation", Organised Sound, 10, 2005, pp. 267-274.

[5] S. Boxer, "Art That Puts You in the Picture, Like It or Not", New York Times, April 27, 2005.

[6] K. Hevner, "Experimental Studies of the Elements of Expression in Music", American Journal of Psychology, Vol. 48, 1936, pp. 246–286.

[7] I. Wallis, T. Ingalls and E. Campana, "Computer Generating Emotional Music: The Design of an Affective Music Algorithm", In Proc. of the 11th International Conference on Digital Audio Effects (DAFx-08), Espoo, Finland, Sept. 2008.

[8] R. Gundlach, "Factors Determining the Characterization of Musical Phrases," American Journal of Psychology, Vol. 47, No. 4, 1935, pp. 624-643.

[9] P. Juslin and J. Laukka, "Communication of Emotions in Vocal Expression and Music Performance: Different Channels, Same Code?", Psychological Bulletin, Vol. 129, No. 5, 2003, pp. 770-814.

[10] J. Russell, "A circumplex model of affect", Journal of personality and social psychology, Vol. 39, 1980, pp. 1161 – 1178.

[11] A. P. Oliveira and A. Cardoso, "Emotionally-controlled music synthesis", 10th regional conference of AES Portugal, Lisboa, 13 – 14 Dec. 2008.

[12] http://www.processing.org (last visited March 3rd, 2009).