Published in the MM '15: Proceedings of the 23rd ACM international conference on Multimedia, 26-30 October 2015, Brisbane, Australia, which should be cited to refer to this work.

DOI:https://dl.acm.org/doi/10.1145/2733373.2807965

Movie's Affect Communication Using Multisensory Modalities

Joël Dumoulin, Diana Affi,
Elena Mugellini, Omar Abou Khaled
HumanTech Institute
University of Applied Sciences
Fribourg, Switzerland
[name.surname]@hes-so.ch

Marco Bertini, Alberto Del Bimbo
MICC
University of Florence
Florence, Italy
[name.surname]@unifi.it

ABSTRACT

The goal of the system presented in this demo is to make possible for the visually and hearing impaired audience to live empathetic viewing experiences using their home theatre. In this work we suggest the incorporation of new emotion communication modalities into the standard television, to provide the targeted audience with sensations that they do not have the opportunity to enjoy because of their disability.

Categories and Subject Descriptors

H.5.1 [**INFORMATION INTERFACES AND PRESENTATION**]: Multimedia Information Systems—*Video*

Keywords

Sentiment analysis; video analysis; affect communication

1. INTRODUCTION

There are lots of competitors in the television market, that need to differentiate themselves providing new functionalities to their products and to respond to the needs of consumers. Increasing the immersion of viewers in movies is one of the axes of the evolution of television experience. Hearing and visually impaired people are a special and challenging audience. In fact, some solutions exist to provide these people with the semantic content of the video using for instance subtitles for the hearing impaired and audio-description for the visually impaired, but solutions that bring them the emotions contained in the movies are not so common yet. In this paper, emotion communication modalities compatible with televisions are presented in order to assure more immersion and empathic sensations to the viewers.

2. RELATED WORK

Many new technologies have appeared adding synchronised devices to the home theatre aiming to enhance the TV viewing experience. Second screen apps such as Syfy Sync [1] are developed to work together with Philips Hue lights in order to change the room lightning (color and intensity) at key moments (particular scenes, such as extreme weather, or intense violence) of the TV shows. More advanced features, such as Microsoft's IllumiRoom [3] project, augment the area surrounding a television screen with projected visualizations to enhance the traditional living room entertainment experience. The audio cue is also being enhanced by working on 3D sound audio system embedded in the new generation TVs. Researches also focused on emotional subtitles [4] in order to express emotions in originally static text. These techniques do enhance the TV viewing experience but do not offer any substitutes for the malfunctioning senses of the impaired viewers.

MM'15, October 26–30, 2015, Brisbane, Australia.
ACM 978-1-4503-3459-4/15/10.
DOI: http://dx.doi.org/10.1145/2733373.2807965 .

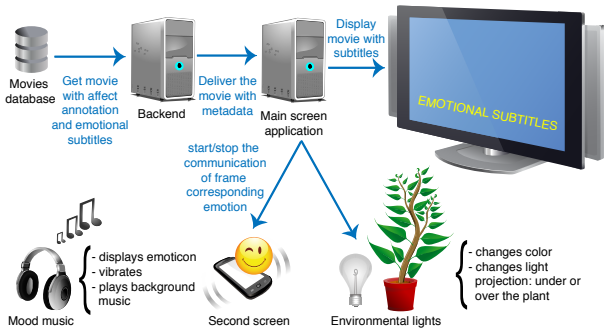

Figure 1: System setup

3. SYSTEM OVERVIEW

The system is composed of a backend application, a main screen application and a set of different emotion communication devices: the TV monitor displaying the movie and the special subtitles, environmental lights, and a second screen application.

In the database, each second of each movie is annotated, showing seven probability values corresponding to the seven basic emotions: *anger*, *disgust*, *fear*, *happiness*, *sadness*, *surprise* and *neutral*. The highest value denotes the emotion mostly perceived in the given scene. This annotation is obtained after a previous processing step applied on the movie in the backend. First, multiple visual and audio features are extracted from each frame per second and over one second long audio clips respectively. Then a previously trained classifier on emotion annotated video datasets, is applied on these features in order to recognise the emotions for each sec-

(a) (b) (c) (d)

Figure 2: Emotion communication via Philips Hue lights: a) TV rear mounted LightStrip b) Blooms setup c) Lighting from beneath for negative emotions d) Lighting from above for positive emotions.

ond of the movie. For classification, a Random Forrest Tree is used after training it on the Acted Facial Expressions in the Wild dataset (AFEW) and the FilmStim dataset. The best result of 32% of accuracy is obtained when using audio features inspired from the INTERSPEECH 2010 challenge and deep learning based visual features addressing both visual content and sentiment [2].

When watching a movie, first the movie is loaded with its emotional annotation and subtitles from the database to the backend application. Then, the video file and the subtitles are sent to the main screen application where the movie is played. While playing, the main screen application sends a command to all the emotion communication modalities to display the emotion corresponding to the current frame.

3.1 Emotion communication devices

Multiple devices and techniques are used to convey emotions to the users while watching a movie. The backend application synchronises these devices with the video player in order to present the right emotion at the right time:

Environmental lights Two Philips Hue Blooms and one Philips Hue LightStrip are used for expressing emotions via lights (see Figure 2). The strip LED light, attached to the borders of the TV screen, changes colours according to the current emotion, while the Hue light, in addition of changing colours, modifies the projection type. In fact, two Hue LED's are used in order to enlighten a plant in two different ways: either from above or from beneath it. This difference of light projection is used to differentiate between positive and negative emotions respectively. As for the colours red is used for anger, green for disgust, blue for fear, yellow for happiness, grey for sadness and pink for surprise.

Second screen A mobile application is developed on the Android platform to accompany the viewers during their movie. The application is controlled by the backend that sends at each second, the corresponding emotion that needs to be presented on the second screen. A *translator layer* is implemented that takes the emotion value and fires accordingly three different events:

1. Displays on the screen the corresponding smiley face (see Figure 3).
2. Plays a background music that is lyrics free and with low volume so it will not disturb the viewer from not hearing the movie's audio but just enough to set a certain mood.
3. Starts a vibrations that will stimulate the viewer - who is supposed to hold the mobile phone in his hand, about some event occurence. The different emotions have different stimuli or vibration patterns as shown in Figure 4.

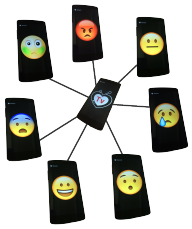

Figure 3: Emoticons shown on a second screen.

Emotion	Vibration pattern
Anger	
Disgust	
Fear	
Happiness	
Sadness	
Surprise	

Figure 4: Vibration patterns for each emotion

Monitor display On the monitor displaying the movie, subtitles are added as well. These subtitles change font, colour and size according to the emotion they convey. CSS classes are created for each emotion and are attributed for each sentence in the movie. The subtitles are generated automatically using the Google Speech Recognition API and saved in a WebVTT file (Web Video Text Tracks format).

The association of colors and emoticons to emotions has been evaluated with a survey conducted on 126 men and 87 women aged between 17 and 60 (avg. 24).

4. CONCLUSION

The designed system presents multiple techniques and devices that can encode an emotional information and deliver it to the television audience. Each of these proposed techniques do target a different sense in order to ensure their effectiveness on hearing and visually impaired viewers. With this solution, the TV viewing experience is enhanced by delivering more immersion for the global users and better content delivery for the targeted audience.

5. REFERENCES

[1] The internet of things: Wifi lights + interactive tv. `https://www.gloto.com/2015/05/the-internet-of-things-wifi-lights-interactive-tv/`. Accessed on May 2015.

[2] T. Chen, D. Borth, T. Darrell, and S.-F. Chang. DeepSentiBank: Visual sentiment concept classification with deep convolutional neural networks. *arXiv:1410.8586*, 2014.

[3] Microsoft. Illumiroom. `http://research.microsoft.com/en-us/projects/illumiroom/`. Accessed on May 2015.

[4] J. Ohene-djan and R. Shipsey. E-Subtitles : Emotional Subtitles as a Technology to assist the Deaf and Hearing-Impaired when Learning from Television and Film. In *Proc. of ICALT*, pages 2–4, 2006.