Emotionally Aware TV

Mohammad Soleymani
Imperial College London
Department of computing
London, UK
m.soleymani@imperial.ac.uk

Maja Pantic
Imperial College London
London, UK
University of Twente
Enschede, the Netherlands
m.pantic@imperial.ac.uk

Proceedings of TVUX-2013: Workshop on Exploring and Enhancing the User Experience for TV at ACM CHI 2013, 27 April 2013, Paris, France.

Abstract
In this paper, we foresee a TV set which is aware of viewers emotions and its environment. This idea goes beyond personalization by taking into account the time and location dependent factors. We argue that not only we should personalize to the users' taste but also to their current emotional state and context.

Author Keywords
Affect; multimedia; interaction; preference

ACM Classification Keywords
H.5.m [Information interfaces and presentation (e.g., HCI)]: Miscellaneous.

Introduction
The digital age dramaticlly changed the way we generate and interact with multimedia content. Multimedia content used to be only generated by a handful of big companies, record producers and television and radio stations. Today, everybody can easily record, edit and publish multimedia content using handheld devices. At the same time, users are also actively participating in annotating the content. Videos go viral based on their popularity among users and the user generated stars are being introduced to the mainstream media. This shift is likely to change the way we interact with our TV sets. We

will expect TV sets not to only be aware of our taste but also to be able to actively engage the viewers in a non-disruptive and natural content selection, based on environmental and personal variables.

Viewers often experience emotions while watching or listening to audio-visual content [6, 7] which manifest through bodily and physiological cues, for example, facial expressions, changes in body temperature and changes in vocal features. Picard was one of the first who brought up the idea of using affect in multimedia interaction [5]. She envisaged a content player which can sense users' emotional states and deliver the content that matches their emotional state. Such an application also needs an emotional understanding of the content itself. Today, we know that there are personal as well as contextual factors in affective responses to multimedia [10, 7].

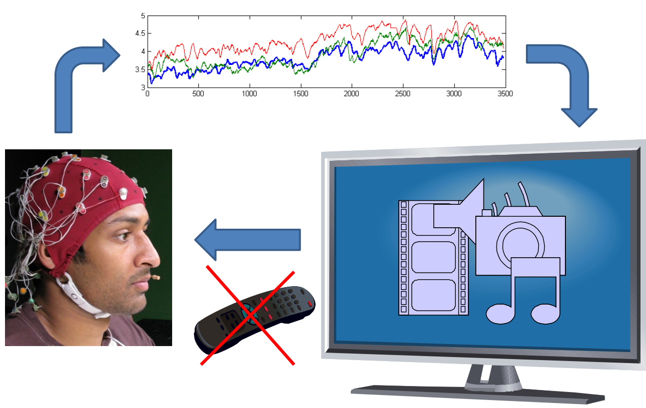

Figure 1: Sensing spontaneous and non spontaneous natural bodily responses will change the way we interact with TV sets.

In order to provide an emotionally aware TV that can recommend content based on a viewer's mood, we need an affective understanding of both the viewer and the content. From the content side, we need affective video indexing which derives representations of video that characterize the emotions that they elicit. There are studies on how to map the content on the most likely emotional reactions of viewers [3, 9]. There have also been studies on representations that can be used as high level descriptors for emotional representation [1]. Affective experiences are personal and context dependent which means it depends on the experience of a single viewer in a certain situation. We do not feel the same in response to a funny video clip being in different mood, fatigue level, time, and location [10]. Therefore, we need sensors to navigate the TV set through viewers' desired content. The sensors can detect viewers' spontaneous responses and provide it to the recommendation engine. The recommendation engine will pick up the sensed information and combine it with contextual factors such as time of the day and the personal profiles generated based on user interaction history and social network. This will reduce the users' direct interaction by sensing and predicting what a viewer really wants (see Figure 1). For example, if we only consider personal view history, a TV set might want to recommend a replay of a recent sport event by the viewer's favorite team. However, if we know that the viewer's initial response is frustration due to the result of the match, the emotionally aware TV can shorten the match summary and skip to a less gloomy program.

There have been studies on sensing viewers' emotions via bodily sensors, e.g., EEG and facial expressions [11, 12] or social media, e.g., Twitter [13]. Proliferation of cheap sensors, such as Microsoft Kinect, is already bringing these sensors closer to our TV sets. The next step in gathering information for content recommendation is not out of reach.

Challenges and Perspectives

With the growing interest in commercially produced sensors and cameras, e.g., Microsoft Kinect, Asus Xtion, and Emotiv helmets, TV sets can benefit from viewers responses to create a more natural and personally tailored experience. However, the research in incorporating spontaneous reactions of viewers is still in progress.

Emotional and spontaneous reactions can be subtle and vary from person to person. One key challenge will be to build machine learning techniques which can detect these subtle reactions and automatically consider person dependent factors to improve the reliability of emotion recognition components. The other challenge of such studies is to create a ground truth by looking into the users' mind.

Taking advantage of emotional information is possible in the active mode [2, 8], when users choose a content based on emotional queries or in the passive mode where the emotional information is used in recommendation [14]. Designing the correct interface and interaction is a major challenge for the active approach. Emotional characteristics cannot be translated directly into preference [4] and finding the best way to use the emotional characteristics for content recommendation is another major challenge. Social dimension is also of great importance in the expression and perception of emotions, e.g., laughter tracks are being added to sitcoms to imitate the emotional contagion. Therefore, social signals should be considered for emotional interaction paradigms.

There are also contextual factors such as time, viewers' physiological state, weather, environment, cultural background, mood and personality that can have effects on our reactions. These environmental and contextual factors are not easy to assess or consider. To model the large variety caused by different contextual factors gathering a large number of samples is necessary. Some people might also find such systems intrusive, and they have legitimate privacy concerns. For example, such technologies can be used for surveillance and marketing purposes without users' consent. These concerns need to be addressed by researchers in collaborations with ethics and law experts.

Despite its challenges, we believe that applications related to entertainment and future media will recognize the value of such systems and will deploy them as one of their core components.

Acknowledgements

The work of Soleymani is supported by the European Research Council under the FP7 Marie Curie Intra-European Fellowship: Emotional continuous tagging using spontaneous behavior (EmoTag). The work of Pantic is supported in part by the European Community's 7th Framework Programme (FP7/2007-2013) under the grant agreement no 231287 (SSPNet) and ERC Starting Grant agreement no. ERC-2007-StG-203143 (MAHNOB).

References

[1] Benini, S., Canini, L., and Leonardi, R. A connotative space for supporting movie affective recommendation. *Multimedia, IEEE Transactions on 13*, 6 (dec. 2011), 1356 –1370.

[2] Eggink, J., and Bland, D. A large scale experiment for mood-based classification of tv programmes. In *Multimedia and Expo (ICME), 2012 IEEE International Conference on* (July 2012), 140–145.

[3] Hanjalic, A., and Xu, L. Affective video content representation and modeling. *Multimedia, IEEE*

Transactions on 7, 1 (2005), 143–154.

[4] Hunter, P. G., Schellenberg, E. G., and Griffith, A. T. Misery loves company: Mood-congruent emotional responding to music. *Emotion 11*, 5 (2011), 1068–1072.

[5] Picard, R. W. *Affective Computing*. MIT Press, Sep. 1997.

[6] Rottenberg, J., Ray, R. D., and Gross, J. J. Emotion elicitation using films. In *The handbook of emotion elicitation and assessment*, J. A. Coan and J. J. B. Allen, Eds., Series in affective science. Oxford University Press, 2007, 9–28.

[7] Soleymani, M., Chanel, G., Kierkels, J., and Pun, T. Affective Characterization of Movie Scenes Based on Content Analysis and Physiological Changes. *International Journal of Semantic Computing 3*, 2 (June 2009), 235–254.

[8] Soleymani, M., Davis, J., and Pun, T. A collaborative personalized affective video retrieval system. In *Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. 3rd International Conference on* (sep 2009).

[9] Soleymani, M., Kierkels, J. J. M., Chanel, G., and Pun, T. A bayesian framework for video affective representation. In *Proceedings of the International Conference on Affective Computing and Intelligent interaction (ACII 2009)* (September 2009), 1–7.

[10] Soleymani, M., Larson, M., Pun, T., and Hanjalic, A. Corpus development for affective video indexing. *CoRR abs/1211.5492* (2012).

[11] Soleymani, M., Lichtenauer, J., Pun, T., and Pantic, M. A multimodal database for affect recognition and implicit tagging. *IEEE Transactions on Affective Computing 3* (2012), 42–55.

[12] Soleymani, M., Pantic, M., and Pun, T. Multimodal emotion recognition in response to videos. *Affective Computing, IEEE Transactions on 3*, 2 (2012), 211–223.

[13] Tang, A., and Boring, S. #EpicPlay: crowd-sourcing sports video highlights. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, ACM (New York, NY, USA, 2012), 1569–1572.

[14] Tkalčič, M., Burnik, U., and Košir, A. Using affective parameters in a content-based recommender system for images. *User Modeling and User-Adapted Interaction 20*, 4 (September 2010), 279–311.