Affective multimedia interaction grounded on a cognitive science approach : interpreting indirect measures of emotion and modeling the affective relationship to multimedia contents.

Olivier Villon and Christine Lisetti
Affective Social Computing Group, Sophia Antipolis, France

September 26, 2006

Abstract

Build an effective affective interaction with multimedia contents (as music, images and videos) require a suitable knowledge of the user. In this proposal, we introduce two problematics related to user modeling, and propose associated models. First we introduce the Embodied Affective Relationship (EAR) as a model aiming at formalizing association between multimedia contents and emotional experience of individuals. Then we introduce the PsychoPhysiological Emotion Map (PPEM) as a parametric model of emotion interpretation from a 3rd person approach (physiological signals) formalizing inter and intra-individual differences.

1 Introduction : human centered multimedia based on indirect emotion measure

Human Centered Multimedia is dependant upon an accurate modeling of the user. As illustrated in the figure 1, we propose an effective affective multimedia interaction in the context of affective computing. We formalize the *emotion elicitor* as the situation which elicits emotion, and the *emotional experience* the evaluation, made by an individual of this situation, and the *1st person or 3rd person measure of emotion* made of psychological and physiological components. Theses two components are the expression/measure of the emotional experience. Considering the physiological and psychological evaluations as the output of a system evaluating this elicitor, it is possible to isolate two important challenges for Human Centered Computing. The challenge (1) is to be able to understand the inter-individual differences between the elicitor and the affective experience. As pointed out in [5],[6],[7] the way individuals evaluate elicitors could be modeled as an *embodied affective relationship* (EAR), dependant of the personal history of each individual. Two users with different background will experience different emotions for the same multimedia contents. Then, the challenge (2) consists on being able to combine psychological (using a 1st person approach, close to emotional subjective experience) and physiological (using a 3rd person approach) component according to individuals, we addressed in [8], [9]. In this context, modeling affective relationship with multimedia contents (EAR) combined with an accurate and continuous emotional sensing (PPEM) could enable multimedia interaction based on emotion. Such multimedia interaction could be applied into interactive art (music, visuals), domotic, multimedia retrieval, Human Computer Interaction ([6],[7]). We anticipate an novel form of multimedia interaction, based on multimedia contents modification, driven by emotion of user/spectator.

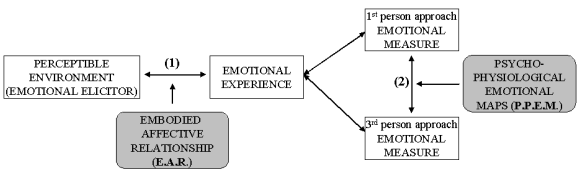

Figure 1: The proposed scenario for an affective multimedia interaction. The EAR model and the PPEM model combined together could allow an affective interaction with the perceptible environment made of multimedia contents.

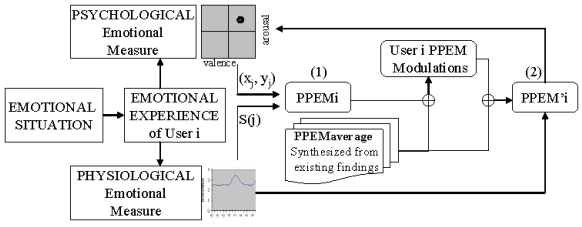

Figure 2: Psycho Physiological Emotional Map (PPEM) construction and use.

2 The Embodied Affective Relationship : E.A.R.

Despite advances of computer devices dealing with multimedia contents, cognitive science models of emotion, and interactive art set up, usable computing tools to simulate human emotional evaluation of multimedia and to design multimedia according to user's emotion are still really poor. We consider that knowledge from cognitive science (i.e. formalizing perception, cognition, and emotion) could contribute to some aims of computer science (i.e. regarding human centered multimedia, personalized content delivery, multimedia indexing and retrieval, interactive art). An *implementable* conceptual model based on a critical survey of literature about the human capability to associate affective experience to media is needed, as well as a computational framework of this model. We proposed that the affective experiences each individuals feel and/or express while experiencing the Perceptible Environment (P.E.) (here extended from music to multimedia) are produced in real-time, on the basis of memorized relationships of the form (emotional experiences; P.E. experience) previously generated by the phylogeny or previously produced by our daily affective experiences with the P.E. We called this the EAR. This is a human functional (simplified) subsystem we model to accounts for our ability to associate affective experience with media, and thus learn and use theses associations while we experience emotion. The contents of the EAR could be different at the inter-individual level. Moreover, the contents of the EAR of an individual could change over time according to new affective experiences with the PE. The structures of the EAR are mainly the Long Term Affective/Associative Memory (LTAM), the Short Time Perceptual Buffer (STPB), and the Short Time Affective Buffer (STAB), which comes from the reduction of the computational (not focused on implementation) model of emotion of Sander and Koenig [3]. This reduction has been made to focus on the possibility of implementation.

3 The Psycho-Physiological Emotional Map : P.P.E.M.

A continuous interaction, e.g. in interactive art, could not be broken by an interruption (e.g. "do you feel amused ?"). Physiology is an interesting candidate for a continuous emotion-based interaction. However, to interpret the physiological activity it is necessary to extract an appropriate emotional semantic from the physiological signal, using a reference to 1st person approach [8, 9]. **Representation of emotion**. The possibility of mapping is demonstrated between some discrete emotions (and affective labels) and position in valence and arousal space (see [2]). Despite several open problems, as the semantic of the affective label, and the exact positions in the space ([4]), "All emotions can be located in a two-dimensional space, as coordinates of affective valence and arousal" ([1]). **PPEM structure**. As shown in figure 2, we propose the PPEM as a descriptive mode of representation for the psychological links to physiological features (Skin Conductances Responses and Heart Rate Variability). We define the PPEM associated to a subject i (single subject form, see fig. 2.1) as a group of specific patterns (S), represented as sets of features values derived from physiological signals, and a psychological part denoted by a coordinate or a dynamic (x, y) into the dimensional affective representation valence*arousal space (which coordinates are convertible into discrete emotion). Once created, a PPEM could be used by a recognition system using the map made of elements. To be able to tailor interpretation without building a complete PPEM for each user, and taking benefits of previous approaches, we define the *parametric* form of PPEM, referred to as $PPEM'_i$ (see equation 1, and fig. 2.2). The psychological output is based on the modulation of a virtual subject PPEM, which represent the pshychophysiological links of the average population found in an experiment and/or in the literature ($PPEM_{\text{average}}$). Inter-individual differences ($dx_{j,i}$ and $dy_{j,i}$, related to personality) are considered as subject i modulation of ($PPEM_{\text{average}}$) output, for the pairs $((x_j, y_j), S(j))$. Intra-individual differences ($dx_{j,i,c}$ and $dy_{j,i,c}$, related to mood and body state), as showed with "Day-dependance" phenomenon, are considered as subject i modulation due to specific conditions c.

$$PPEM'_i = \{((x_j + dx_{j,i} + dx_{j,i,c}, y_j + dy_{j,i} + dy_{j,i,c}), S(j))\} \text{ with } \; j = 1, \ldots, N \tag{1}$$

4 Conclusion

We introduced the PPEM and EAR model, which focus on the possibilities to simulate multimedia contents and emotion human association, and use this simulation to drive the media selection/design according to the affective/emotional experience of user indirectly measured. This could have potential into any HCI system using multimedia manipulation controlled by computer. These models rise several questions regarding emotion indirect interpretation, emotion representation, multimedia contents formalization and the notion of affective memory which may be discussed.

References

[1] P. J. Lang. The emotion probe. studies of motivation and attention. *Am Psychol*, 50(5):372–385, May 1995.

[2] James A Russell. A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6):1161–1178, 1980.

[3] David Sander and Olivier Koenig. No inferiority complex in the study of emotion complexity: A cognitive neuroscience computational architecture of emotion . *Cognitive Science Quarterly*, 2:249–272, 2002. special issue on Desires, Goals, Intentions, and Values: Computational Architectures.

[4] K. R. Scherer. Emotion. In W. Stroebe (Eds.) M. Hewstone, editor, *Introduction to Social Psychology : A European perspective (3rd. ed.)*, pages 151–191. Oxford: Blackwell, 2000.

[5] Olivier Villon. Conscious dynamic of affective state and embodied relationship to music. Master's thesis, Université de Bordeaux 2, Bordeaux, June 2002. Supervisor : Antonio Camurri.

[6] Olivier Villon. Expérience affective en situation d'écoute musicale : expressions et relation à la musique incorporée. Master's thesis, LIMSI-CNRS, Université de Paris XI, Paris, September 2003. Supervisor : Antonio Camurri.

[7] Olivier Villon and Christine Lisetti. Earmultimedia : Approche modèle utilisateur du phénomène d'évaluation affective de l'environnement perceptif. In *Colloque des Jeunes Chercheurs en Sciences Cognitives (CJCSC 2005), 2-3-4 May 2005*, page 225, Bordeaux, May 2005.

[8] Olivier Villon and Christine Lisetti. Toward building adaptive user's psycho-physiological maps of emotions using bio-sensors. In Dirk Reichardt, Paul Levi, and John-Jules C. Meyer, editors, *1st Workshop on Emotion and Computing at KI2006 : Current Research and Future Impact, Bremen, Germany, June 19th, 2006*, pages 35–38, June 2006.

[9] Olivier Villon and Christine Lisetti. A user-modeling approach to build user's psycho-physiological maps of emotions using bio-sensors. In *IEEE RO-MAN 2006, The 15th IEEE International Symposium on Robot and Human Interactive Communication, Session Emotional Cues in Human-Robot Interaction*, Hatfield, United Kingdom, September 2006. IEEE.