Institut Eurécom
Department of Multimedia Communications
2229, route des Crêtes
B.P. 193
06904 Sophia-Antipolis
FRANCE

Research Report RR-07-200

Toward a Computational Model of Subjective Affective States Associated with Multimedia Contents

July 13th, 2007

Olivier Villon

Tel : (+33) 4 93 00 81 51
Fax : (+33) 4 93 00 82 00
Email : Olivier.Villon@eurecom.fr

[1]Institut Eurécom's research is partially supported by its industrial members: BMW Group Research & Technology - BMW Group Company, Bouygues Télécom, Cisco Systems, France Télécom, Hitachi Europe, SFR, Sharp, STMicroelectronics, Swisscom, Thales.

Toward a Computational Model of Subjective Affective States Associated with Multimedia Contents

Olivier Villon

Abstract

Adapting media (images, sounds, videos) according to the affective/emotional experience of user is a challenge for Human Computer Interaction, Computer Mediated Communication, Interactive Art and Personalized Content Delivery. In this paper, we propose the outline of a user computational model based on a survey of the literature about human capability to associate affective experience to media. We also propose applications using such user model for media selection and design according to user estimated emotion.

Index Terms

Computational Model of Emotion, Affect, Emotion, Neurosciences, Multimedia Affective Interface, Interactive Art, Inter-Individual differences, User Modeling, Automatic Affective Processing

Contents

1 Introduction **1**

2 Related Work : The individuals' emotional evaluation of multimedia contents from a computing point of view **2**

3 Embodied Affective Relationship model **3**
3.1 Model outline and use in HCI 4
3.2 Natural cognition basis of the model 7
3.2.1 A subset of emotion study 7
3.2.2 The LTAM : a phylogenetical and ontogenetical associative memory linking emotion and perceptible environment . . 7
3.3 Artificial cognition formalization of the E.A.R. toward implementation 8
3.3.1 LTAM content 8
3.3.2 LTAM structure 8
3.3.3 E.A.R. learning algorithm : filling the LTAM 10

4 Discussion **12**

1 Introduction

Humans have the capacity to be highly engaged into creative processes which can lead into the design of perceptible artifacts providing individuals' emotions. Music and film are examples of such highly advanced emotional communicating artifacts which are sought by many in order to experience affective states. Modeling how individuals embed emotion into such artifacts, and how individuals interpret such artifacts in terms of emotions is nowadays more accessible to research. Affective science make advances in the understanding of how human generates emotion while experiencing multimedia contents (e.g. understanding the emotion elicited by music, colors, etc...). Computing have produced new hardware and software tools to control multimedia contents enabling the creation and/or the control (e.g. MIDI specification, MPEG-7 norms) of such perceptible artifacts (e.g. music, video, odors) which can be experienced with Human Computer Interaction, Computer Mediated Communication, Interactive Art and Personalized Content Delivery. However, the cross-fertilization of the understanding of human perceptible environment emotional processing and computer-based control and analysis of multimedia contents into computing tools which simulate human emotional evaluation of multimedia and which help the analysis control and design of multimedia adapting to user's emotion, are still poor. In this paper, we thus aim at understanding the individual's relationship to computer controlled media, and formalizing it in order to make accurate tools toward an automated user's emotion-based multimedia selection, modification and design. To achieve this goal, we describe a user-model ('Embodied Affective Relationship' -EAR- proposed previously [1,2]) of the affective relationship one can have with multimedia content(e.g. music, sounds, colors, a live-performance video or any interactive interface content)and how such user-model may be used to select and/or modify this multimedia content. As shown in Figure 1 our problematic seeks to answer whether we can model (and later use to tailor and adapt multimedia to User x) the EAR (1) that user (someone who listen music, a guitar player or a spectator of an immersive video performance) use to generate an affective state, an emotion, we can measure (3) in presence of perceived multimedia environment (4). Given that cultural and personal background of each individual can be different, we aim at consider the high level of subjectivity which may characterizes such affective relationship (i.e. the emotional measure (3) may be different from one individual to another one, for a similar environment (4)). In the next section we discuss related research at computing level by giving examples of the state of art in artificial cognitive systems for new personalized media content delivery. We next present possible application of the proposed model, the natural cognition basis of the model, and a formalization of one main component, namely the Long Term Affective Memory.

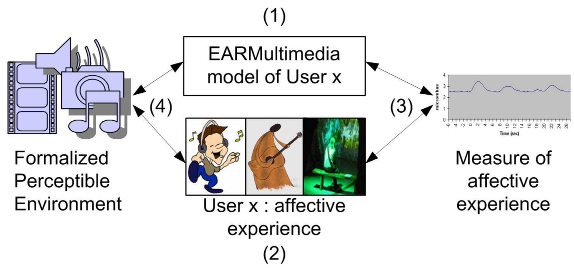

Figure 1: The problematic of the EAR

2 Related Work : The individuals' emotional evaluation of multimedia contents from a computing point of view

Over the past five years, some domains in HCI and Multimedia research (interactive art, bio-feedback, personalized multimedia content delivery and affective computing) have started to acknowledge and include in multimedia environments the knowledge regarding the affective state that the user can experience in presence of specific media mainly using Multimedia Indexing and Retrieval (MIR) techniques. Personalized Multimedia Content Delivery (PMCD) aims at selecting media among increasingly growing collection of media data. The criterion for selection can be any form of categorization, from perceptive (i.e. visual contents with blue color or auditory content with high beat per minute average), to cognitive (i.e. contents with beach scenes, art from a specific singer), to social (the more listened song among a network of people), to also affective (an arousing image, a liked song) on which we will focus here. PMCD is mainly based on two technical methods: collaborative filtering (CF) and content-based filtering (CBF). In CF, the recommendation mechanism only works with the categorization of multimedia elements (e.g. like or dislike of a song) of other users, and does not perform in depth analysis of the contents of multimedia elements, nor of the relationships between the multimedia elements and its belonging categories. CBF, using Multimedia Indexing and Retrieval (MIR) techniques, is a perceptive and cognitive means of categorizing and recommending multimedia. Multimedia indexing aims at generating the metadata (e.g. using color for images [3]), whereas multimedia retrieval aims at retrieving media which match specific metadata or share properties with other media (e.g. melody similarity). Some attempts had been made by measuring affective state of the user to select or modify the multimedia contents (e.g. [4]) explicitly (e.g. with questionnaire), implicitly (e.g. with psycho-physiological devices), or without measuring user's state (e.g. interpreting skipping as an indication of dislike). As they focused on measuring affect, they did not detail however the way they produced the metadata. Recently, [5] uses affective representations for MIR techniques : low-level cues (describing contents) and affective labels are associated. The authors wrote : "the sequences selected should be characterized by the content flow on which an average user is expected to react in a "standard manner" in terms of arousal". This approach support the notion of predefined affective expectations, from cognitive schemes : "For instance, the arousal is expected to rise when the development of a soccer game goes from the stationary ball exchange in the middle of the field and finishes with the score via a surprisingly forward push toward the goal". This schema is intended to produce a shared affective response among the audience. One important limitation remains: the assumption that there exists a common, standard emotional reaction to media regarding intra-individual differences. However, a computer model is needed to relate media content and affect, in the case this relation is highly personal, and not necessarily shared among user. We will see in next sections what to consider from natural cognition (psychology and neurosciences) to enhance existing systems and create new ones which can be based on individual affective relationship to multimedia.

3 Embodied Affective Relationship model

Multimedia contents are usually considered as conveyors of emotion (e.g. films, music, etc...). Be able to retrieve multimedia content which is associated to specific affect or emotion, for a specific user requires first to design a model of the individual affective relationship to multimedia content. To introduce the proposed model, we will consider the emotion communication processes involved in the production of multimedia contents, and the interpretation of these contents in terms of emotion between an artist (or a designer)-spectator(or a listener, a user of an interface) communication. As schematized in 2 an artist or a designer aims at communicate affective information. This is done during a creative process which lead into the production of an artifact (i.e. multimedia contents, e.g. a music, a film, etc...) which we will call an "affective object". According to [6] an 'affective object' has the ability to 'map' an 'emotional data from a person' 'to an abstract form of expression and communicate that information expressively, either back to the subject herself or to another person'. Then, the spectator, the listener or the user of an interface which contain an affective object interpret an emotional message. The fact that the user match or not the affective/emotion intention of the artist should be related to the notion of communication and its underlying processes: how could we describe the process which leads the artist to embed emotion in the affective object ? How could we describe the process which leads the spectator to identify an emotion in the affective object?

We can consider that a common underlying process is used to associate emotional messages (e.g. Joy) to some elements of the affective object. We proposed to formalize the basis of such associations with the Embodied Affective Relationship model ([1, 2, 7]) which stands for a conceptual set of processes and structures. In this model the affective experience each individual feel and/or express while experiencing the Perceptible Environment (P.E., i.e. what we are able to perceive in an affective object) are produced on the basis of memorized relationships of the form {emotional representations; P.E. representations} previously generated by the

phylogeny or previously produced by our daily affective experiences with the P.E. An artist may thus produce a specific affective object to express Joy, using its own implicit associations (fig. 2 the EAR of the artist). Then the spectator experience the affective object (by seeing or hearing it), then perform an implicit analysis of multimedia contents (cognition) and use their own E.A.R. to interpret an emotion (fig. 2 the EAR of the spectator). For [8], "a symbolic form (...) is the point of departure for a complex process of reception (the aesthetic process that reconstructs a 'message.')". Following this formalization, the E.A.R. of individuals could be partially shared or not between different individuals : The E.A.R. is made of universal associations as well as associations learnt and thus dependent on cultural and personal background (e.g. a traditional song for a specific group of individuals may elicit a specific emotion for that group and another emotion for another group).
Among several structure involved into the model, we will focus here on the LTAM component toward its implementation.

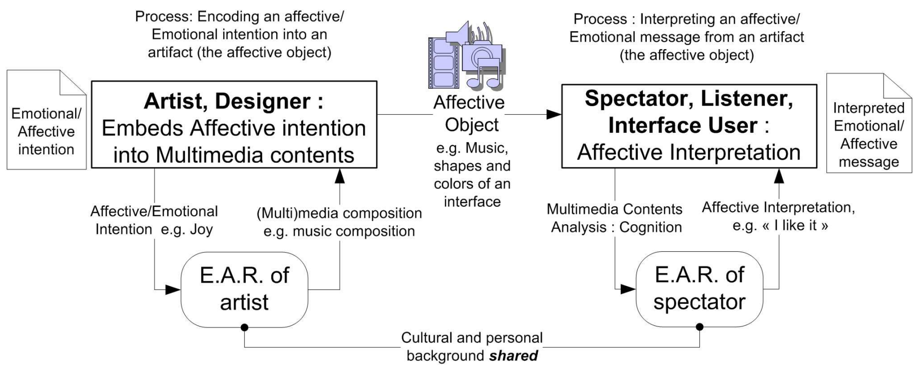

Figure 2: The proposed formalism for an emotional communication trough an affective object (e.g. music ; color of an interface)

3.1 Model outline and use in HCI

We provide an overview (fig. 3) of the use of the *EAR* to Multimedia model by considering three scenarii: Simulation, Poïetic-modification and Poïetic-generation.

Simulation mode (top of fig. 3) aims at simulating the affective evaluation by the user of a P.E. for which the content is formalized. In a *learning* phase, the model extracts different features values from the environment (content formalization, e.g. hue extraction, pitch extraction, musical structure extraction, etc...) and measure the associated emotion for the user (by using for instance psychophysiological measure). Thus we obtain a personal database made of P.E. patterns associated to measure of emotion. The Long Term Affective Memory is updated (LTAM, one main component of the EAR for which the structure and associated process will be detailed in next sections) for this user by adding new relations of the form emotional experience measure; P.E. experience measure - i.e. features from the multimedia and by refreshing existing relations. Then in a *use* phase the system gives an emotional evaluation (valid only for the user on which we built the user model) of incoming multimedia content (arrow (1) show that we feed the EAR model with multimedia content and (arrow 2) an emotional evaluation of the multimedia content is produced). This estimation could be used to select multimedia content on the basis on user estimated emotional evaluation.

Poïetic mode acts on the P.E. perceived by the user. Poïetic is an old greek opposed to Aesthesis. Poïetic refers to fact to embed a specific experience, concept, emotion into an artifact (a sculpture, a painting, a music) and is opposed to the aesthetical experience in presence of specific artifacts, i.e. how the fact to experience an artifact elicit emotions. When the environment is exogenous from the system point of view (like when someone plays an instrument for example), we talk of modification. When the environment is endogenous from the system point of view, i.e. a full compositional control of the environment by the system (as in web interface design), we call the mode "generation".

The **Poïetic-modification** (middle of fig. 3) consists on modifying an existing environment (i.e. exogenous from the system point of view), e.g. a sound output of a musician playing guitar with an audio effect. In this case the system will aim at modify the PE according to wished emotion for the user who experience the PE. Following the example of the musician, we can consider an assistant for artistic application, driven by the artist emotion. During a *learning phase*, the guitar player acts on a sound effect. This sound effect is formalized. In parallel, the emotion of the guitar player is measured. The EAR model is built for this user by associating emotional evaluation and content formalization. Then the system assist the artistic performance in a *use phase*, by driving the effect of the player according to a desired emotion time-based graph. The desired emotion is sent to the EAR model and the model outputs potential content which match the desired emotion. The environment control is done according to this output and drives the effect of the guitar player.

The **Poïetic-generation** (bottom of fig. 3) consists on a complete control of the environment (i.e. endogenous from the system point of view) on the basis on the emotion it may elicit for a specific user. In this case the system will aim at generate an appropriate P.E. according to wished emotion for the user who experience the P.E. In a *learning phase* the system produces random or predefined arrangements of the primitives of the environment (e.g. hue values, sound frequency) into groups of percepts (e.g. a specific shape, a minor chord) which are passed to environment renderers (through environment specification converted to xml). An xml feed is thus sent to renderers as odor system, visual system or sound systems controlled through computer. Then, the user experience this environment and produces an emotional evaluation. This evaluation is sent to the short term memory component of the EAR (which extracts dynamics of the emotional evaluation in the affective buffer and extracts dynamics of the environment in the perceptual buffer[1]). The short-term memory outputs pairs of emotional experience measure; P.E. experience measure valid for the user and are used to update the LTAM of this user. In a *use phase* the environment generation is driven by the emotion of the user, estimated trough the LTAM (by querying emotion associated to specific groups of percepts and their combination). The approach taken by the EAR modeling, is

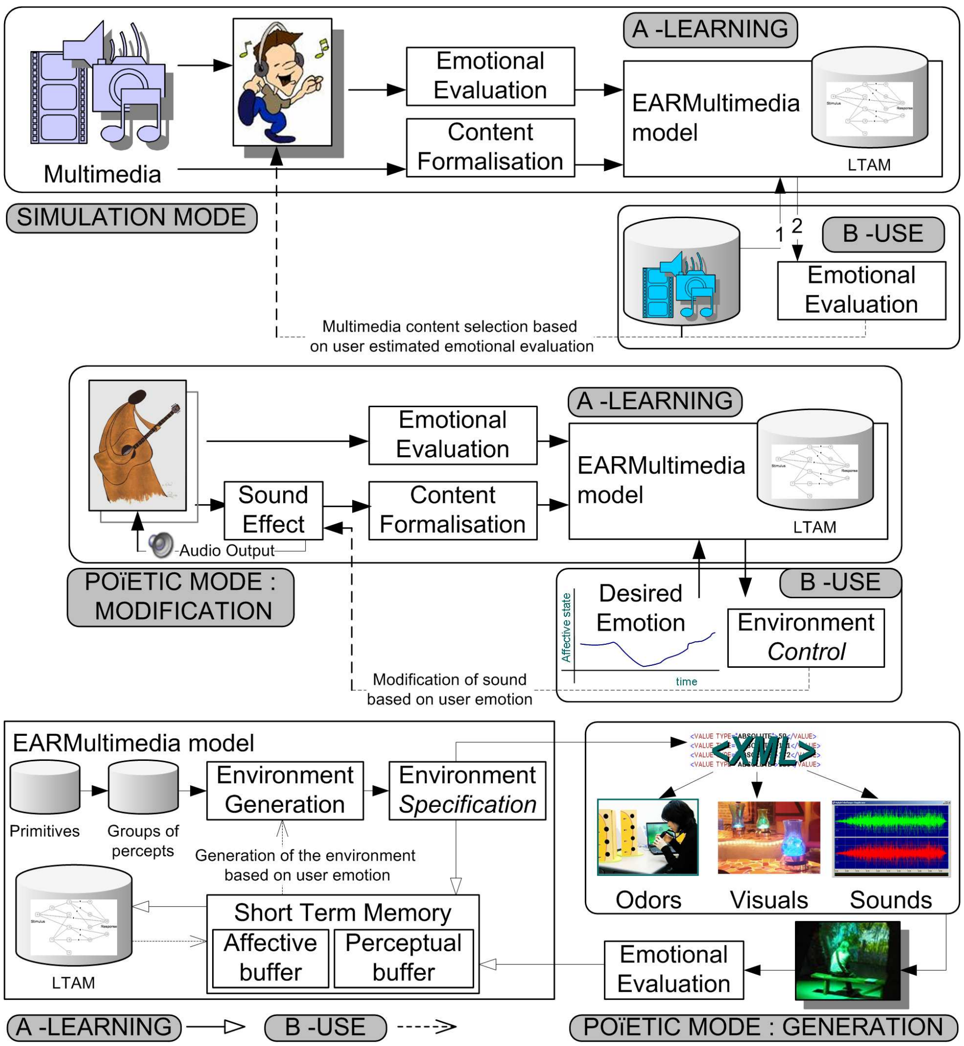

Figure 3: Modes of the model, and examples of potential applications : Simulation, Poïetic (modification) and Poïetic (generation).

[1]The short-term memory components are not described in detail in this article. The short term memory is made of two components : a short term affective buffer (STAB) which decompose affective representations and a short term perceptual buffer (STPB) which decompose environment into groups of percepts and primitives. The short term memory extracts dynamics from the buffers to create structures(e.g. to detect a chord from a set of three notes).

a user-modeling approach. However, rather than consider machine learning technique to associate the measured emotion to the multimedia content formalization, we consider a two steps approach : (1) make an accurate model from cognitive science theories, and (2) fill this user model with data learned from each user ([2]).

3.2 Natural cognition basis of the model

3.2.1 A subset of emotion study

Affective state and emotion is a wide domain which had been scientifically investigated in several directions. Behind affect-related studies, the object of study is not unified, depending of the field of investigation and the focus on (1) emotion expression/measure, (2) generation of emotion/evaluation of stimuli processes or (3) evolutionary description and learning processes. As this paper focus on modeling the affective relationship with media, we will focus on the thematic of "automatic affective processing" ([9]) : "affective processing does not depend on controlled cognitive processing. (...) Organisms are able to determine whether a stimulus is good or bad without engaging in intentional, goal-directed, conscious, or capacity demanding processing of the (evaluative attributes of the) stimulus. Rather, affective processing could occur automatically." However, abstract stimuli as non-figurative music or painting, are processed by high brain structure (newer in the evolution) and so require high level of cognitive processing. This paradox between the human automatic and unconscious affective processing of PE, possible with complex and abstract PE, but resembling stimulus-response of a wide number of species is explained by [10] with music as example : "music recruits neural systems of reward and emotion similar to those known to respond specifically to biologically relevant stimuli, such as food and sex(...) Activation of these brain systems in response to stimulus as abstract as music may represents an emergent property of the complexity of human cognition". As we are able to produce complex emotion onto the PE, it means that cognitive architecture allow us to retrieve complex emotional message with an automatic and passive evaluation of the PE. The cognitive evolution leads in the fact that (1) some abstract and non-evolutionary relevant stimuli are processed using the some similar mechanisms that survival-related brain circuitry, and (2) specific human cognitive process result on the possibility to produce complex emotion using theses mechanisms.

3.2.2 The LTAM : a phylogenetical and ontogenetical associative memory linking emotion and perceptible environment

According to [11], “besides being a factor that can influence memory, emotional information can also be stored as a memory”. The “memory of emotion” (episodic memory, explicit remembering of past emotional situation) is different from the “emotional memory” (storage of affective properties of situations). Emotional memory, in the Ledoux conception, raises the next questions: (1) what and how is encoded into memory after affective experiences (2) how memory is involved into the process of emotion generation? The LTAM is an attempt to formalize responses to these questions by a formalization of the memory contents and the relationship within such emotional memory, i.e. a data structure and processes relating P.E. and emotional experience. [12] built a "explicitly-described computational architecture of the emotion system". Among several useful principles and components of the model, two components are mainly used to build the LTAM : (1) The system of stimulus responses connections which deals with relationships between a stimulus and a response, innate ("evolutionary salient-stimuli", like fear of sudden loud sound), and learnt (conditioned, like fear of a specific object) ; (2) The associative memory store amodal representations, and contains associations "that point toward exteroceptive and interoceptive representations located in different pattern activation subsystems such as the visual, auditory, and internal state pattern activation subsystem". The internal state pattern activation subsystem, "stores in long term memory representations of previous internal state patterns". Thus, perceptions and affective state (the internal state in their model) are related trough pointers.

3.3 Artificial cognition formalization of the E.A.R. toward implementation

3.3.1 LTAM content

Synthesis of neurosciences and psychological studies of affective learning, evaluative conditioning, memory (e.g. [11], [13], [14]) led into the table 1, which contains the type of memorized associations which were experimentally demonstrated (i.e. a formalization of *what* could be stored into the LTAM). All theses opposed categories could be mixed (an affective pair could be e.g. of the form {multimodal complex dynamic pattern subpart of PE; instantaneous direct discrete affective property}.

3.3.2 LTAM structure

According to emotion studies [11, 12, 15] : (1) The emotional memory is the result of a synthetic process using incoming data with previous data stored; (2) Affective learned response could occurs on low-level feature (e.g. color) and on complex pattern (e.g. a painting); (3) Emotional generation could be partly driven by compositionality. Taking together theses statements, we consider that perceptive-affective pairs containing complex patterns of P.E. are not stored as a separate entities, added to the previous contents of the stimuli but are encoded with a synthetic process. This means that the perceptive-affective pairs should be hosted in a structure presenting a high interdependency of the items it contains. Thus, we propose that such perceptive-affective pairs are organized into the LTAM as a network which support the notion of pointers of emotion model of proposed by Sander and

The subpart of the P.E. contained into a perceptive-affective pair could be :	
Multimodal (e.g. color and sound)	Unimodal (e.g. odour)
Complex patterns (e.g. a tiger)	Feature (e.g. a hue value)
Absolute (values or patterns) (e.g. a hue value)	Relative (values or patterns) (e.g. a minor chord at any key)
Static value patterns (e.g. a painting)	Dynamic values patterns (e.g. a song)
The affective property contained into a perceptive-affective pair could be :	
Discrete (e.g. joy)	Dimensional (e.g. high arousal)
Absolute (e.g. high arousal)	Relative (e.g. an increase of arousal)
Static (e.g. fear)	Dynamic values patterns (e.g. joy then fear, or increase of arousal)
Instantaneous (e.g. tiger elicits fear)	Delayed (after a specific pattern)
Direct	inter-items of the P.E. - i.e. ordinal -

Table 1: Description of the affective pairs contained into the LTAM (opposed categories in each columns)

Koenig.The proposed formalization of the LTAM is thus a network of perceptive-affective pairs: a multiple graph (without loops) containing three subgraphs, one perceptual, one affective, and standing for the perceptive-affective pairs (i.e relationships). Formally, the LTAM of an individual n, at time t, is defined by :

$$G_n(t) = ((PV_n, PE_n), (AV_n, AE_n), RE_n) \tag{1}$$

In figure 4, PVn={1,2,3,4,5,6,7,8} (Perceptual Vertices) represents the P.E. and AVn={9,10,11,12,13,14} (Affective Vertices) represents affective states. The more the vertices are situated at the middle of the graph (regarding left and right), the more the level of representation is precise and simple, representing precise features of the PE, and specific elements of the affective experience. At the extremity of the graph, the vertices represent the whole PE and an actual affective state which are related. For example, but not necessarily, it could be a stimulus-response. PEn={(1,6) ;. . . ; (6,8)} (Perceptual Edges) and AEn={(9,12) ;. . . ; (13,14)} (Affective Edges), are respectively the sets of edges relating the multilevel description of the PE, and the multilevel description of potential affective states. For example, the perceptual pattern represented by the edge 7, is made of the values associated to the edges 2, 4, 4 and 5, as they are related by PEn. This representation allows combining a representation of the PE at a rather high level (as structural) with a lower level (as each relative values contained into the structure). REn=(1,9) ;. . . ; (8,12) (Relationships edges) are the perceptive-affective pairs. A pair of the type PE value, Potential affective state is at the middle of the graph (e.g. (1,9)), while a pair of type whole PE, affective response links boundaries of the graph (e.g. (8,12)). Such boundary pair hosts thus also the stimulus-responses. The REn could link any level of representation, i.e. between a whole PE and a single affective potential (e.g. a specific song elicits a arousal augmentation), or between a PE feature and an actual affective response (e.g. a sudden loudness augmentation elicit the fear). Each edge is weighted by a Strength (s) defined by the EAR initialisation and update, corresponding to the probability of occurrence, and the possibility of extinction. This LTAM support both the notion of compositionality and emergence of emotional memory, by defining the REn (e.g. a feature as a specific color embedded into specific patterns could exhibit several affective properties, but these color alone has the potential to elicit all the affective properties in which such feature is involved), or a different affective property.

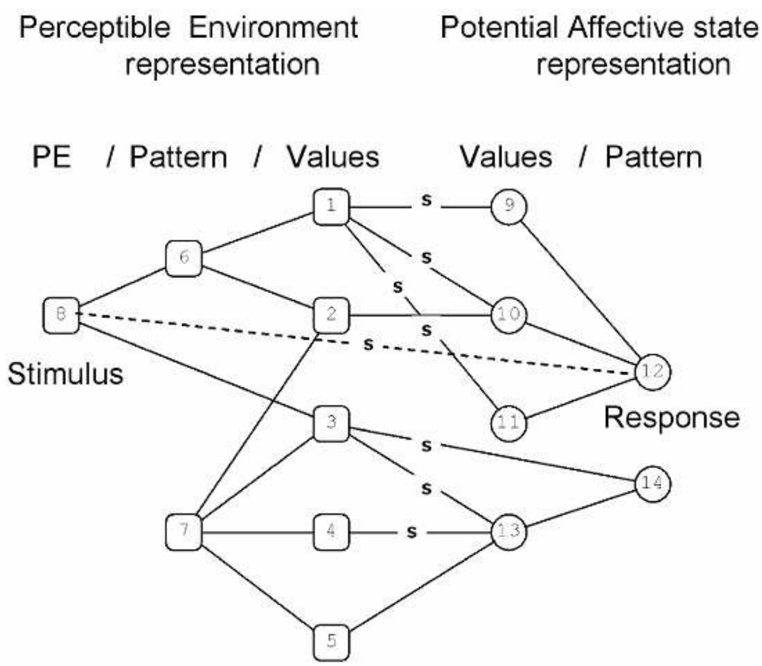

Figure 4: The proposed representation of the Long Term Affective Memory.

3.3.3 E.A.R. learning algorithm : filling the LTAM

The E.A.R. learning algorithm aims to extract part of the E.A.R. of a subject from the affective measures we can perform onto this subject. Each new (intra-individually consistent) affective measures performed are considered. We consider the notion of aesthesis results as a set of pairs made of a formalized PE, which was submitted to a subject, and a formalized corresponding affective experience measured for this subject. To build the EAR, we should consider the role of the STPB. This is done by the fact that the perceptive/cognitive representation embeds formalization of the PE. Then, we should respect the notion of intra-individual consistency. To build the E.A.R, we should consider some aesthesis results for which the affective responses are consistent within an individual. Once we get consistent aesthesis result from an individual, we can build an actual (consistency) and useful (interesting relationship to then be able to manipulate the PE) LTAM. The idea is to be able to start from a measure, and arrive to what serve to this subject to produces such measure. If the EAR is accurately modeled from the measure, we will be able to use it to simulate affective experience for other elements of the PE. The algorithm is on the following form (commentaries are placed after "//"). Not all the rules are provided here, as they still in design phase, but give an overview of the learning mechanism :

Algorithm 1 Form of the EAR learning algorithm.

```
for each perceptual-affective pair from the aesthesis result do
    //stores the aesthesis result perceptual and affective elements
    generates a vertice 'P' in PVn, and a vertice 'A' in AVn.
    places theses vertices at the left and right extremity of the graph
    //generates the perceptual and affective representations
    for each multilevel descriptor of the vertice 'P' do
        search any existing similar vertice in PVn
        for each found similar vertices do
            generate an edge with 'P'
        end for
        for each non found similar vertices do
            generate a vertice
            add it to PVn (placement at the right of 'P')
        end for
    end for
    similar 'for' sequence, for 'A'
    //generates the pointer corresponding to the aesthesis result
    generates an edge relating theses two vertice, into REn, weight = 1
    //generates the possible others pointers
    //and update the existing ones
    for each possible pair of vertices standing for the multilevel descriptor of 'P' and 'A' do
        if the two vertices already existed before the generation of 'P' and 'A' then
            //we confirm the existence of this edge in REn
            increase the weight of this edge
        end if
        //any rules could be added here to increase and decrease weights
    end for
end for
```

The rules to decrease and increase the weights are not dicussed in detail here. They come from ratio between numbers of time the perceptive part of the aesthesis result had been found to be associated to specific emotion measure. Moreover, a process of decomposition of the aesthesis results into more minimal REn edges should be realized using a threshold of compositionality, to stop the process of decomposition. Indeed, internal inference should be done to extend the LTAM amount of information. An example of such inference, using graph data structure could be found in [2], which allow to make prediction with 63% of accuracy. Finally, the minimal requirement of this algorithm is to succeed to create the LTAM as a fully compatible data set with the different aesthesis result. Such coherence is needed to then be able to perform simulation and PE manipulation.

4 Discussion

We proposed the outline of a model which aims at simulate an 'affective memory' for model users' relationship with multimedia contents, based on cognitive science knowledge about human emotional processing. The proposed approach focus on using cognitive knowledge about human emotional processing to design a user model. Our previous work consisted on the definition of the EAR at natural cognition level, its formalization for computing, experimental testing and extension to multimedia.

More generally, this model may help to take into account personalized user affective relationship with the multimedia contents, in any interactive designs, where we expect an affective response of the user. The model presented is a tradeoff between cognitive modeling accuracy, need of implementation, and descriptive status of the model. Moreover, it could embed several level (from whole PE, patterns to values) and types (absolute, relatives values, etc. . .) of representation, both at perceptual and affective one. This is mainly interesting as we don't know precisely what an individual extracts in the perceived environment, and on which level of affective property it create the perceptive-affective pairs.

Recent work includes implementation and testing using multimedia content, using the presented possible modes. The Simulation mode may be complementary with approaches of personalized multimedia content delivery. The Poïetic mode may extends any form of interactivity based on emotion (e.g. interactive art).

Acknowledgments

Thanks to D. Sander for useful explanations about the computational architecture of emotion and to C. Lisetti for her comments on the proposed approach. Thanks to Antonio Camurri, Gualtiero Volpe and Barbara Mazzarino of Infomus for interesting discussion about the basis of the proposed model.

References

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

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

[3] Aigrain, P., Zhang, H., Petkovic, D.: Content-based representation and retrieval of visual media: A state-of-the-art review. Multimedia Tools and Applications **3**(3) (1996) 179–202

[4] Field, A., Hartel, P., Mooij, W.: Personal dj, an architecture for personalised content delivery. In: Proceedings of the tenth international conference on World Wide Web, (1-5 May Hong-Kong), New York, ACM Press (2001)

[5] Hanjalic, A., Xu, L.Q.: Affective video content representation and modeling. Multimedia, IEEE Transactions on **7**(1) (2005) 143– 154

[6] Scheirer, J., Picard, R.W.: Affective objects. Technical report, Mit Media Laboratory Perceptual Computing Section (2000) Technical Report No. 524.

[7] Villon, O., Lisetti, C.: 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, Bordeaux (2005) 225

[8] Nattiez, J.J.: Music and Discourse: Toward a Semiology of Music (Musicologie générale et sémiologie, Paris 1987). Princeton, NJ: Princeton University Press (1990) Translated by Carolyn Abbate.

[9] Houwer, J.D., Hermans, D.: Automatic affective processing. Cognition and Emotion **15**(2) (2001) 113–114

[10] Blood, A.J., Zatorre, R.J.: Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. Proc Natl Acad Sci U S A **98**(20) (2001) 11818–11823

[11] Ledoux, J.E.: Memory versus Emotional Memory in the brain. In: The Nature of Emotion, Fundamental Questions. Oxford University Press (1995) 311–312

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

[13] Houwer, J.D., Baeyens, F., Field, A.P.: Associative learning of likes and dislikes: Some current controversies and possible ways forward. Cognition and Emotion **19**(2) (2005) 161 – 174 Special Issue : Associative Learning of Likes and Dislikes.

[14] Gewirtz, J.C., Davis, M.: Using pavlovian higher-order conditioning paradigms to investigate the neural substrates of emotional learning and memory. Learning and Memory **7**(5) (2000) 257–266

[15] Robinson, W.S.: Colors, arousal, functionalism, and individual differences. Psyche **10**(2) (2004)