Université de Nice - Sophia Antipolis

UFR SCIENCES

École doctorale « Sciences et Technologies de l'Information et de la Communication » de Nice - Sophia Antipolis

THÈSE

pour obtenir le titre de

Docteur en Sciences

Discipline: Informatique

présentée par

Olivier VILLON

Modeling Affective Evaluation of Multimedia Contents: User Models to Associate Subjective Experience, Physiological Expression and Contents Description.

Thèse dirigée par Prof. Bernard Merialdo

soutenue le 29 Octobre 2007 devant le jury composé de

Jean-Claude Martin	Maître de conférence	LIMSI-CNRS	Rapporteur
Thierry Pun	Professeur	Université de Genève	Rapporteur
Bernard Merialdo	Professeur	Institut Eurécom	Examinateur
René Garcia	Professeur	Université de Nice Sophia-Antipolis	Examinateur
Lionel Martin		STMicroElectronics	Examinateur

"Les exigences de l'art et les connaissances opératoires des artistes ont de tout temps stimulé et inspiré la recherche scientifique et l'innovation technologique."

Jean-Claude Risset

Abstract

Ten years after the foundation book "Affective Computing" which proposed concepts and approaches related to the measure and the use of emotion in human computer interaction scenarios, several decades after the first psychological and physiological study on emotion and hundreds years after philosophical study about aesthetical emotion, very few automated systems are nowadays able to *manipulate multimedia contents* (selection, design, creation) according to the *felt affective states and emotions* (which can be measured by different means) by an individual. Among several limitations, we consider in this thesis that one important problem is the amount of inter-individual differences both in the indirect measure of emotion and in our affective evaluation of multimedia contents. In this thesis we propose enhancements of computer possibilities to manipulate media contents on the basis of felt affective states by modeling affective and emotional associations to multimedia contents and by automating the process of interpretation of emotion from indirect measure. This enhancement is achieved by the design of two user models which can take into account user's specificities toward a better adaptation. First we introduce the Embodied Affective Relationship to Multimedia Contents (EAR) as a model aiming at formalizing association between multimedia contents and emotional experience of individuals. This model is then presented in a practical way toward systematic affective handling of multimedia contents. Then we introduce the Psycho Physiological Emotion Map (PPEM) as a parametric model of emotion interpretation from physiological signals taking into account inter and intra-individual differences. Our technique aims at psychologically interpreting physiological parameters (skin conductance and heart rate), and at producing a continuous extraction of the user's affective state during Human Computer Interaction. An experiment is presented to estimate emotion from physiological signals. Both models are built upon an engineering cognitive science approach, i.e. implementing psychological and neuroscience knowledge to design a computer system. Finally, an experimental Application Programming Interface built upon the proposed models is presented to enable novel form of affective state and emotion driven-human multimedia interaction.

Résumé

Dix ans après le livre fondateur "Informatique Affective" qui proposa des approches et concepts associés à la mesure et à l'utilisation des émotions pour l'interaction homme- machine, plusieurs décennies après les premières études des émotions en psychologie et physiologie et des centaines d'années après les études philosophiques sur les émotions esthétiques, très peu de systèmes automatisés sont aujourd'hui capables de *manipuler les contenus multimédia* (sélection, design et création) à partir des *états affectifs et des émotions ressentis* (qui peuvent être mesurés de façons différentes) par un individu. Parmi plusieurs limitations, nous considérons dans cette thèse qu'un problème important est la quantité de différences interindividuelles rencontrées dans la mesure indirecte des émotions ainsi que dans l'évaluation affective des contenus multimédia. Dans cette thèse nous proposons des améliorations quant à la possibilité informatique de manipuler des contenus multimédia sur la base des états affectifs ressentis en modélisant les associations affectives et émotionnelles générées avec les contenus multimédia et en automatisant le processus d'interprétation des émotions sur la base de mesure indirectes. Cette amélioration est réalisée à travers la proposition de deux modèles utilisateurs qui prennent en compte les spécificités de l'utilisateur en vue d'une meilleure adaptation.

Nous présentons d'abord le modèle Relation Affective aux Contenus Multimédia Incorporée (Embodied Affective Relationship to Multimedia Content - EAR) visant à formaliser les associations entre les contenus multimédia et l'expérience émotionnelle des individus. Ce modèle est ensuite présenté de façon pratique en vue d'une manipulation affective systématique des contenus multimédia. Ensuite, nous présentons le modèle Cartes Psycho-Physiologiques Emotionelles (Psycho Physiological Emotion Map - PPEM) constitué d'une représentation paramétrique servant à l' interpretation émotionelle à partir de signaux physiologiques et prenant en compte les différences inter et intra-individuelles. Notre technique vise à interpréter psychologiquement les paramètres physiologiques (conductance cutanée et battements de cœur), et à produire une extraction continue de l'état affectif de l'utlisateur durant une interaction homme machine. Une expérience visant à estimer l'information émotionnelle à partir des signaux physiologiques est

présentée. Les deux modèles sont construits à partir d' une approche issue des sciences et de l' ingénierie cognitives, c' est à dire utiliser et implémenter les connaissances issues des disciplines de la psychologie et des neurosciences afin de construire une système informatique. Enfin, une interface de programmation expérimentale (Application Programming Interface - API) construite pour et sur la base des modèles présentés est décrite pour permettre la mise en place de nouvelle forme d'interaction homme-multimédia dirigée par les états affectifs et les émotions

Acknowledgements

First of all, I would like to thank all the members of my jury. I am really thankful to Professor Jean-Claude Martin (LIMSI-CNRS, University of Paris XI) and Professor Thierry Pun (University of Geneva) to have accepted to be reviewers. Next, I would like to thanks Professor René Garcia (University of Nice), to have accepted to be the president of this jury and Lionel Martin (ST Microelectronics) to follow this thesis at each stage in the framework of the PACALab. Next, I want to thank Professor Bernard Merialdo (Institut Eurécom) who allow me to finish this thesis in good conditions by succeeding to C. Lisetti in the direction of this thesis.

In particular, I would also like to thank Professor Antonio Camurri and other members of the Infomus Lab of University of Genova (Donald Glowinski, Gualtiero Volpe, Barbara Mazzarino...) and other participants of the Humaine Summer School(Ben Knapp, Carol Krumhansl) who provided a unique opportunity to perform an experiment with spectators during an artistic representation.

I want also to thank several Professor who particulary guided my interest for research in computational aesthetics, from art history and aesthetics to cognitive science : Marlene Belly (ethnomusicology, univ. Rennes II), Marc-Michel Corsini (AI, univ. Bordeaux II) and Rene Garcia (neuroscience of emotion, univ. Bordeaux II).

Thanks to Jean Claude Risset to have opened a unified view of art, science and technology (AST) in france and for advising me to study separatly art, science and computer science to then be able to fully combine theses field in research.

Next, I also thanks all the members of Institut Eurecom and especially my team mate in the research adventure : Fabio Valente, Vivek Tiagi, Marco Paleari, Amandine Grizard, Federico Matta, Slim Trabelsi, Remi Trichet, Emilie Dumont, Jihanne Benhour and other PhD students...

Moreover, I would like to acknowledge that this work was partially founded by ST microelectronics in the framework of the Region *Provence Alpes Côte d'Azur* (PACA) PACALab. Thanks to Mercedes and Frederic for allowing me to use their Disklavier piano and for their hospitality during the redaction of this thesis. Thanks to Jean, my grandfather, who initiated me to electron-

ics and thus mainly contributed to my interest in scientific thinking. Thanks to Laurent, my brother, for initiating me to computing and programming on his TO-16 (!) and for his support during this thesis. Naturally I couldn't end without thanks my parents Michel and Michèle for their daily cheerfulness, open-mindedness and their endless support for all my orientations which bring me to present this thesis.

Finally thanks ever so much to Alice for coming along with me and sharing so many beautiful experiences, which mainly contributed to the stability needed to end this thesis. Thanks to Alice for her curiosity in life and in people, her patience, determination and understanding since the beginning of my interest in what is now called computational aesthetics, and which sometimes take up a lot of place in our life.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation: Toward an affective state-driven Human Multimedia Interaction

How to adapt media (images, sounds, videos) according to the affective and emotional experience of individuals ? What are the challenges to overcome to enable Human Computer Interaction (HCI) ,Computer Mediated Communication (CMC), Interactive Art (IA) and Personalized Multimedia Content Delivery (PMCD) based on affective states and emotion experienced by an individual in presence of media contents ? What are the methodological problems involved ? How can we access to emotion measure through a computer without direclty requiring individual explicit input (e.g. without asking user) ? What is the required knowledge in the natural cognition domain (psychology, neurosciences, aesthetics) about the human affective processing of media contents to build a computational model of such processing in regards to multimedia handling ? Ten years after the foundation book "Affective Computing" which proposed concepts and approaches related to the measure and the use of emotion in human computer interaction scenarios, several decades after the first psychological and physiological study on emotion and hundreds years after philosophical study about aesthetical emotion, very few automated systems are nowadays able to *manipulate multimedia contents* (selection, design, creation) according to the *felt affective states and emotions* (which can be measured by different means) by an individual.

The main motivations of this thesis are the enhancement of computer possibilites to manipulate media contents on the basis of felt affective states by (1) modeling affective and emotional associations to multimedia contents and by (2) automating the process of interpretation of emotion from indirect measure.

Affective states elicitation by perceptual contents. It is admitted that humans are able to feel affective states and emotions while experiencing perceptual contents. Ranging from natural contents (e.g. a landscape) to

human artifacts (e.g. music), perceptual contents could be experienced with a multimodal perception (e.g vision, audition, smell). This ability leads humans to create a unique practice among all living species: art.

Practice of embedding emotion into perceptible artifacts : Art. Humans are highly engaged into creative processes which lead into the design of perceptible artifacts providing individuals' emotion. Music, film, perfumes are example of such highly advanced emotional communicating artifact which are sought by individuals. The practice of art (e.g. music composition) is partially a process of embedding an emotional intention of the artist ((e.g. create a message of fear) into an artwork artifact (e.g. the musical piece) delivered to a group of spectators for which an affective experience is intented to be felt. Such practice require to (1) be able to manipulate the perceptual contents. This means that (1.1) artists need cognitive abilities to split the perceptual contents into percepts and identify such percepts to compose the artworks with them (e.g. a musician should be able to distinguishes notes and also be able to organize them to build a musical piece). This means also that (1.2) tools are needed to formalize such perceptual contents (e.g. mathematical division of octave in 12 half-tones in music) and to render the formalization into perceptual contents (e.g. musical instruments). The second requirement is to (2) be able to associates affective states and emotions to elements of perceptual contents, e.g. percepts or group of percepts(e.g. when composing a musician need to express in some way his emotion and expressive message into the musical piece produced).

Computer requirements toward an affective state-driven Human Multimedia Interaction.Art is a human practice, and the motivation of this thesis is not to enable computer to make art. However, the elements presented in the artistic pratice are mandatory to enable a computer to manipulate multimedia contents toward an affective state-driven Human Multimedia Interaction. Advances of human comprehension, helped by the cognitive science approach, provides interesting findings toward an accurate model of human perceptible environment emotional processing. However, usable computing tools to simulate human emotional evaluation of multimedia and design of multimedia according to user's emotion are still poor. We illustrate the motivatation of the approach of this thesis to contribute to the affective state-driven Human Multimedia Interaction by using a parallel with artistic practice. In parallel with artistic practice, the computer needs (1) to manipulate multimedia contents. It also needs (1.1) to have a set of vocabulary to identify percepts (e.g. have a formal set of features to recognize). It also needs tools (1.2) to analyze the contents and extracts percepts (i.e. extract features related to perception, e.g. the pitch and musical structure used for multimedia indexing) and to be able to automatically render such formalization (e.g. the Musical Digital Instruments Interface MIDI which enable to make music from specific instruction driven by a computer). Finally, the last requirement in parallel to art practice is (2) the ability to as-

sociates affective states and emotions to elements of multimedia contents, to be able to select, modify or event create multimedia contents of the basis of affective states and emotions. Theses considerations built on both modeling human activity and on solving a computer-based challenge directly motivate one main problematic of the thesis : *how to model abstract representations of affective states associated to multimedia contents ?* The second main consideration to enable an affective state-driven Human Multimedia Interaction is the ability to measure affective states and emotions indirectly. We mean by indirectly the fact to not directly ask an individual how (s)he feel.

1.2 Focus and Contributions of the thesis

We can decompose an affective-state and emotion-driven human multimedia interaction into two main computer tasks to achieve and answer associated research questions. The figure present an overview of the conceptual components invovled in this thesis. 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 problematics.

We first need to focus on (1) the extraction of emotional evaluation of multimedia contents from theses contents. In computer science, this interest is recent and usually takes the form of machine learning task in the field of multimedia indexing and retrieval, with an associated semantic to extract related to emotions data. We then need to focus on (2) the measure of subjective emotion from indirect emotion measure. An indirect measure require an intepretation to extract the emotional content it contains. Gestures, facial expression and physiological signals are example of such indirect measure. In this thesis, we focus on the indirect measure of emotion using physiological signals and especially skin conductance and heart rate. In computer science this is also usually considered as a machine learning problem using methods for features selection and optimization of learning. Moreover, we can isolate two important considerations. First (1) we should be able to understand the inter-individual differences between the elicitor and the affective experience. Two individuals may experience different emotions for the same multimedia contents. Then (2) we should be 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, knowing that inter-individual differences also exist. We aim at connecting the components of the figure 1.1 by mainly taking into account the above-mentionned

issues.

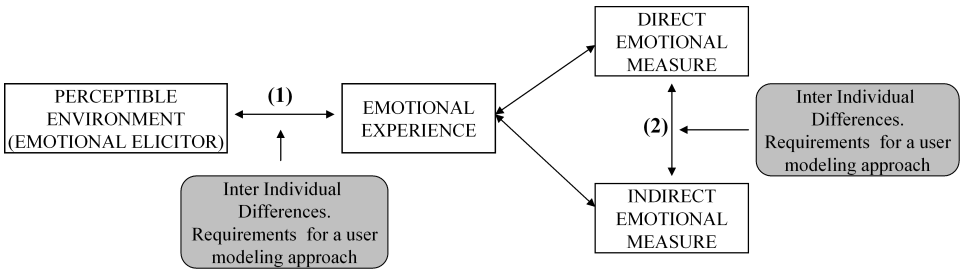

Figure 1.1: Main conceptual components involved in this thesis, and associated problematic to overcome for an affctive state driven multimedia interaction.

Approaching a computer science problem with a cognitive science approach. In the computing domain, the work presented in this thesis takes place in the context of Affective Computing([Picard, 1997], by focusing on the measure and modeling of affective state and emotions from computers, to enhance Human Computer Interaction. It also take places into the KANSEI processing notion whch can be considered as a Subjective Data Mining [**?**], oriented toward affective states and emotions. It also takes place into the field of interactive art by aiming at providing novels forms of interaction with media contents based on the experience felt by individuals. The study of emotion in computing field is relatively new compared to psychology and neurosciences. The recent interest for cognitive approach, which aims at formalizing and using knowledge from natural cognition science to solve artifical cognition problems lead to interesting results and proof to be useful for designing computer systems[1]. The approach taken into this thesis is mainly *a cognitive science approach applied to computing methodological and modeling issues in the topic of affective-state and emotion-driven human multimedia interaction toward the design of a computer-based system.* The approach is illustrated in the figure 1.2. As we need computing models adaptable to

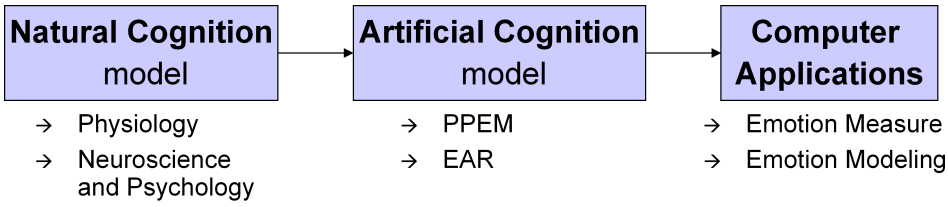

Figure 1.2: The cognitive approach taken in this thesis to design computer models.

[1]For instance, implementation of form recognition in vision computing system from description of human vision system lead to good results)

users, we aim at building models grounded on cognitive science knowledge and approach. The contributions of this thesis may thus be splitted in the design of two user models, as illustrated in figure 1.3:

1. The indirect measure of emotion toward an accurate interpretation close to the subjective experience (P.P.E.M. : PsychoPhysiological Emotional Map approach)

2. Modeling affective relationship with perceptible environment for interactive affective system (E.A.R. : Embodied Affective Relationship to Multimedia Contents model)

These models are proposed as a contribution to enhance affective interaction by building models to overcome inter-individual differences difficulties.

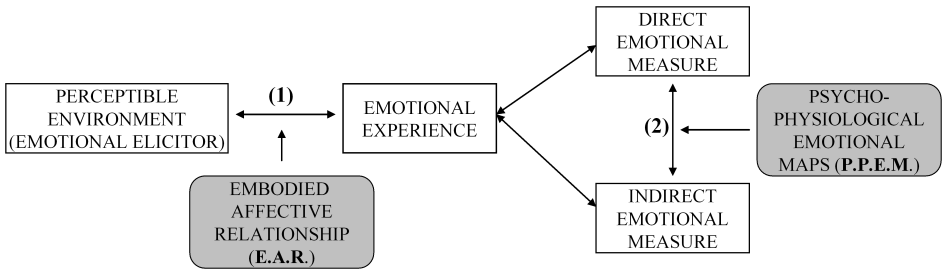

Figure 1.3: The proposed models in the thesis to build an affective state and emotion-driven human multimedia interaction.

1.3 Outline of the thesis

The first chapter constitutes an overview of the state of the art at computing level in the affective state and emotions-driven human multimedia interaction. This includes measure of individual relationship to multimedia content and its application, and the measure of affective state and emotion from physiological signals. We then adopt in the second chapter a cognitive science approach on these computing approaches and extract some problems in regards to natural and simulated cognition studies (i.e. psychological and neurosciences models and empirical data related to emotion). By natural cognition we mean the knowledge about human cognitive information processing, including perceptual and affective processing. By simulated cognition we mean the study which aims at implement models of cognition to test empirical findings and design computing systems. By analyzing knowledge from natural cognition studies in regards with the computing we point out key problems to overcome at computing level. The chapter 4 is dedicated to the description and proposition of an implemented system toward modeling the embodied affective relationship of individuals. We propose a conceptual

model which is based on a critical survey of the literature about the human capability to associate affective experience to multimedia contents, and propose a computational framework of this model. This model focuses on the possibilities to simulate such human association, and use this simulation to drive the media selection/design according to the affective/emotional experience of user. This could have potential interest for HCI system using multimedia manipulation controlled by computer.

In the chapter 5, we develop an approach to recognize emotion from physiological signal which take into account knowledge from physiological research (i.e. type of signal to study, etc..) and include this knowledge into the process of building a user model. The proposed method is based on the psychophysiology field, which tries to describe precisely the links between physiological parameters and their interpretation in terms of psychological meaning. The technique aims at interpreting psychologically physiological parameters as skin conductance and heart rate, and produce a continuous extraction of affective state, based on 3rd person approach. The reference to the 1 st person approach measure allow a tailored interpretation of the physiological measure, closely related to the user own emotional experience. After an introduction about the psychophysiology of emotion, an experiment is presented. The results of the experiment are analysed and a software for real-time analysis of emotion is designed.

In the chapter 6 an experimental Application Programming Interface built upon the proposed models is presented to enable novel form of *affective state and emotion driven-human multimedia interaction*. Finally, the chapter 7 present a sumary of contributions and future works.

Chapter 2

Computing Approaches on Multimedia Contents and Affective Experience and Expression

2.1 Introduction

In this chapter we introduce existing computing systems and approaches related to the measure and modeling of affective states. We first present systems involved in what we could call the *affective gap*. The semantic gap [Smeulders et al., 2000] is usually referred as the research toward associating some indexed content features to some high-level concepts. In the approaches we will present the high-level concept is from affective domain (e.g. find something which is liked by the user). Then we present the approaches related to the emotion measure from user of computing systems. Firstly, we introduce general means to automatically measure emotion and then focus on the emotion measure from physiological signal.

2.2 Systems focusing on emotion and multimedia contents : the "affective gap"

This section aims at federating several artificial cognition systems which tends to embed affective relationship with media into their systems: interactive art, bio-feedback, personalized multimedia content delivery and affective computing. From selection of preexisting media (e.g. video films sequences) to media generation (e.g. real-time automated composition) attempts had been made to manipulate media according to user's emotion criterion.

2.2.1 Personalized multimedia content delivery

Personalized multimedia content delivery aims at selecting media among increasing collection of media data. The criterion for selection could be any form of categorization, from perceptive (i.e. contents with blue color, with

high beat per minute average), to cognitive (i.e. contents with beach scenes, from a specific singer), social (the more listened song among a network of people), and also affective (an arousing image, a liked song). These categories could overlaps. Here we will focus on the last form of categorization ,i.e. a personalized multimedia content delivery based on affective data, as we think it is one of the technologies which could takes benefits of our proposed EAR approach.

In general, personalized multimedia content delivery (will be next referred as PMCD) is mainly based on two methods. On the first hand, Collaborative Filtering (CF) ([Herlocker, 2000]), is a social means of categorizing and recommend multimedia. In the emotion domain, it is based on the comparison of people tastes, likes and dislikes, among a user's network. Once users' profiles are created (relating some multimedia elements with user categorization), the system use the comparison of users profile to recommend to a specific user multimedia elements that might be liked by this user. The method used consists on finding several users with shared tastes for media elements, and then propose to one of this users of this community, new media elements (new for this user, and coming from other users' profile among this community). The affective ratings could be done using the behavior of the user (e.g. skipping), or explicit rates (e.g. using Likert's scale ranging from dislike to like).

The first wide application of CF to PMCD, associated to direct user taste assessment, was realized at the end of 1996, by the company Firefly Network (initiated by Pattie Maes from MIT). They proposed a system based on collaborative filtering to tailor media (here music) delivery for users, which was down in 1998. Nowadays, among existing systems, main online one are LastFm (music, `http://www.last.fm/`), using the CF engine AudioScrobbler (`http://www.audioscrobbler.net`), and Launch at Yahoo (music, videos, `http://launch.yahoo.com/launchcast`). As if CF is a powerful system, it remains limited due to the need of social data. So, the recommendation mechanism work only onto the categorization of the multimedia elements (e.g. like or dislike a song) by other users, but does not analysis in depth the contents of multimedia elements neither relationships between the multimedia elements and its belonging categories.

On the other hand, the Content-based Filtering (CBF), using Multimedia Indexing and Retrieval (MIR) techniques, is a perceptive and cognitive means of categorizing and recommends multimedia. It is a type of multimedia semantics management, based on generation and use of multimedia contents descriptors (multimedia metadata) mainly for audio (e.g. [Iwahama et al., 2004]) and visual ([Aigrain et al., 1996]) media . Multimedia indexing aims at generate the metadata (e.g. using color for images ([Aigrain et al., 1996], or musical events, based on beats onsets, for sound [Gao et al., 2004]), while multimedia retrieval aims at retrieve media which match specific metadata (e.g.) , or share properties with other media (e.g.

melody similarity [Carré, 2002], rhythmic similarity [Foote et al., 2002]). Few attempts had been made to incorporate affective dimensions into MIR techniques, by measuring affective state of the user explicitly, implicitly (e.g. with psychophysiological devices), or without measuring user's state. First works were done in the music domain, as the Personal DJ of [Field et al., 2001] in 2001, based on explicit behavior of the user, like skipping songs, and mood ratings. Other work was done using digitized psychophysiological measures (heart beat, skin conductance, etc...) shown to be indicator of emotion ([Healey and Picard, 1998]), like the affective DJ of [Healey et al., 1998]. These two devices are able to select songs according to the user moods. As they focused onto affective measurement, they didn't detail the way they produces the metadata. More complex approaches are done in video, like the recent work of [Hanjalic and Xu, 2005] in 2005, which fully integrate the notion of affective representation with the MIR technique. In this approach, low-level cues (describing contents) and labels (like affective position in the Pleasure-Arousal-Dominance space, [Russell and Mehrabian, 1977]).

However, it remains two important limitations. The first is the notion of common, standard emotional reaction to media (see section 3.3 regarding intra-individual differences, above-mentioned) ; the author 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". So, at the opposite with previous mentioned studies, they not focus of the affective measure, and does not fully takes into accounts the user-dependent affective reaction. They 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". However, it is a work closely related to our proposed approach, despite the notion of E.A.R. as a personal relationship.

Regarding the indexing techniques, traditional MIR techniques are mainly based onto algorithmic and machine learning solutions. However, alternative and robust techniques aims now to embed simulated perception and cognition systems, like the image classification using SpikeNet neuronal package simulator ([Thorpe et al., 2000], [Delorme and Thorpe, 2003]), which could lead in more human-like form of indexing, and are thus interesting for our approach.

To summarize, the CBF is also a powerful approach, as it is really close to the human affective evaluation, i.e. based on the contents of the media, rather than simple metadata descriptors (like the name of an artist, etc...). Work has been also done in the fusion of both techniques: CF and CBF ([Paulson and Tzanavari, 2003], [Basilico and Hofmann, 2004]), to improve the recommendations mainly by adding MIR techniques to collaborative systems ([Kohrs and Mérialdo, 1999]). However, as we seen, existing system does not focus on user simulation and modeling.

The selection of media could be also be augmented by the cognitive science-based modeling of each user's relationship to multimedia content, i.e. use the E.A.R. of each user to improve the recommendation.

This section was oriented to the selection of the media according to individual's emotion. The next section summarize how the media contents could be modified and/or generated using the approach of interactive art.

2.2.2 Toward and Interactive art based on spectator emotion

We can describe the process of art creation/reception by the notion of poïetic (i.e. productive process) and aesthesis (i.e. perceptual process) initiated by Aristotle. These terms belongs now to the field of semiotics, and had been more recently reformulated as the whole process of art creation by [Nattiez, 1990] about music. Using a poïetic process, the artist (producer) embeds a message into an object (the 'trace', neutral level). This trace is then experienced by a spectator (receiver), which "reconstructs a 'message'" trough "a complex process of reception": the aesthesis (see [Nattiez, 1990]).

2.2.2.1 From art selection to art creation by spectator: 1970's interactive art revolution.

Considering the fact that art creation and experience could be related to emotion, the aesthetical process could lead in the construction of an emotional message by the spectator, as well as the artist could engage its own emotion into the art object trough the poïetic process.

Until 1970's, the spectator was passive. It could only experience art and according to the emotion felt decide to orient himself toward specific style, artist. This could be summarized as the unique possibility of the spectator was a selection process. This is for example what we do when we are choosing to attend to our favorite artist concert, for the pleasure it gives to us.

The 1970's have "established the artist either not as producer of objects but as maker of situations wherein which the creativity of the public can be realized" (translation from [Bureaud, 1999]). Such 'oeuvres participatives' (participating pieces) place the spectator in a new position: it let the spectator modify and/or create the art object. This notion is formalized under the term "spectactor" (spectator/actor) ([Dumouchel, 1991]) to describe the new role of the spectator. Tacking back the model of [Nattiez, 1990] we can consider that the spectactor could be involved into poïetic process or verily that the artist and the spectactor are the same person. In this last case, we can consider that such spectator it's the artist for himself, if we succeed to give to him the possibilities to do it. The interactive art, a result of automation and computing techniques applications for art, allow now an actual spectactor status to spectator.

2.2.2.2 Interactive art based on spectator emotion: an ongoing revolution.

Combining interactive art and emotion is a challenging ongoing research. The general interactive art main problem is made of 3 steps.

Firstly the choice of the sensor (like microphone, tactile, lights, etc...) should be interesting for the spectactor to be able to communicate something to the system.

Secondly the choice of actuator to generate and /or modify the artistic object (the 'trace' in the Nattiez model), should be enough advanced to provide an interesting artistic experience to the user.

The third critical choice is to be able to map the sensor and actuator to modify the artistic object in an interesting manner. Computing solution exist to do such work, as patch programming multipurpose environment like for instance PureData (http://puredata.info/), Max (http://www.cycling74.com/products), EyesWeb (http://www.eyesweb.org/), or dedicated interactive software like for instance LiveTrip(http://musike.org/livetrip/). Theses software provide means to sense user action (e.g. movement), but let the programmer free to create the rules linking sensor and actuators, using their own criterion.

Several computing projects attempt to use the spectator's emotions as criterion of interactivity. Theoretically, it means that the emotion should be sensed, and that the effect of specific artistic object should be statistically correlated to the emotion sensed, in order to be able to control emotionally the evolution of such participative piece.

2.3 Modeling a natural cognition phenomenon to perform a computing task.

The proposed approach aims at cross-fertilize natural cognition phenomenon of emotion study and computing.

2.3.1 Contribution of the approach to computing

The affective relationship to media is becoming an important component of HCI. It can enhance computing systems by adding them the knowledge of the affective response the user would have been produced in presence of specific media. Over the past five years, main fields of HCI research, focusing on the modification of multimedia environment, started to add such dimension (see last section)However, if those systems show a growing interest for the affective dimension, we discuss next the need to investigate a core modeling and implementation of the human emotional processing, as a factor of development of such systems.

2.3.2 Focusing on cognitive simulation and contributes to an engineering problem.

To be able to choose, and or modify the media to fit users affective specific requirements, the ongoing systems developments does not try to produce a correct modeling of such human ability, but try to find an engineering solution by considering the mapping between labels and media as a mathematical problem. Actually personalized multimedia content delivery is now an active field of research with the growing collection of media data. Some application now aim at deliver contents based on affective data. Main tendencies are the Collaborative Filtering (see an overview in [Herlocker, 2000]), associated to direct user taste assessment (see online systems like AudioScrobbler -http://www.audioscrobbler.net/ - or Launchcast - http://launch.yahoo.com/launchcast/ -), and the Multimedia Indexing and Retrieval (MIR), or Content-based Filtering, associated to user taste assessment (see [Hanjalic and Xu, 2005]). Theses systems are detailed into section 4.1.

Those applications do not look at cognitive science modeling of user in depth, but primarily search for solution to deliver contents that works, as this is a necessity for multimedia providers, more engineering-oriented than fundamental research/simulation-oriented. Collaborative filtering use the comparison of multiple users' data to select and propose to user personalized content. MIR searches optimal machine learning techniques to associate affective labels with the indexing sets of features extracted from multimedia contents (e.g. [Kang, 2003] which use HMM).

The approach we present aims at formalize several natural cognition principles in terms of logic and mathematic and then embed such formalized principles into software. Aiming at simulate, then implement a cognitive phenomenon had been proofed to be a powerful approach for computing engineering (see [Delorme and Thorpe, 2003]). We consider the fact that such approach is useful for resolving such engineering problem, related to emotions and computing, like previous approaches which were focused onto perceptual engineering problems did (artificial audition, vision, and olfaction). Such approach aims at bringing together the disciplinary humans' status: from subject (during experimentation in psychology and neuroscience), then individual (persons with their own personal and cultural background), until user (of HCI application, with needs).

2.3.3 Contribution of computing to the approach

However, as if such work could serve computing for the above-mentioned reasons, it is the computing development which allows the design of such work. Firstly as underlined in the problematic the P.E. should be formalized in terms of perceptual elements (trough analysis or generation process), and

could be manipulated (modified or generated) with an automated manner (i.e. with computer) to get a full status of multimedia (i.e. with precisely described contents).

Formalizing the environment in terms of perceptible elements does not requires fundamentally computing, especially when is it closely related with the generation of such environment, and so when the environment as the status of artifact (created by human). For example practice of formalizing the contents of the music is an old process, merged with the generation process itself, and related to the perception. However artifacts which use a formalized representation does not always come with the formalization itself, like polyphonic music. In this case computer is needed to be able to retrieve the formalized structure of the artifact. Moreover, artifacts do not always rely on a well standardized formalization (e.g impressionist painting). The formalization of such P.E. in this case is thus comparable to non-artifact elements.

In the case of non-artifact elements of the P.E. (like nature elements), the type of formalization is necessary a perceptual one: the format of the formalization should follow the format of our perception. In this case, computing is needed to be able to reproduce our perceptual processes, like pitch extraction for music, color detection for vision. The engineering progress made in sensors (from old microphone to digital camera and recent artificial nose) and in perceptual simulation (artificial audition [Leman et al., 2001] and vision [Delorme and Thorpe, 2003] cognitive systems) produced by the implementation of cognitive science research allow an actual digital formalization of the P.E. The standardization of such formalization leads into promising automatic formalization, like the MPEG-7 standard, for describing multimedia contents.

Formalizing the P.E. is a necessary basis to answer the statement (1) of the problematic, i.e. the simulation, and the 'indexation' part of the statement (2). However, 'modifications' and 'generation' of the P.E. evoked in the statement necessary require a computer-based manipulation of the P.E. We define the manipulation of the environment as a control of elements of the P.E. (modification), or a synthesis capability (generation). The effectors of computer are now really suitable to P.E. control. From a standard PC screen/loud-speaker to virtual reality systems, it is now possible to control precisely sound and visual displays. Odors synthesis (see DigiScent company, and Exhalia system initiated by France Telecom R&D) will probably be soon possible for end-users. Several standards, like the midi one in music, allow to a synthesis of sound according to computer formalized action, extending traditional scores. Human rules of artifact's production start also now to be embedded into computers. For example The Continuator ([Pachet, 2003]), or SaXex ([Cañamero et al., 1999],) is able to generate music according to humans' improvisation styles. Thus computer allows manipulating multimedia on the basis of the description of its contents.

Secondly, efforts in affective computing allow new direction toward automated emotional sensing. Ongoing research shows that computing starts to automatically interpret behavioral and physiological measures performed on the user of any HCI devices in terms of emotion, embedding cognitive science experimental results ([Lang et al., 1993], [Tsai et al., 2000]) into HCI devices. The embedding and interpretation of this signals into HCI devices had been mainly considered as a unified means of measuring user's emotion with the MAUI framework ([Lisetti and Nasoz, 2002]) . Emotion behavioral measures mainly consist on facial expression, which led into facial expression recognizers systems like ([li Tian et al., 2001], [Bastard, 2004]). Emotion physiological measure mainly consists on heart rate and skin conductance interpretation, which led into experimental systems like [Vyzas and Picard, 1999], [Lisetti and Nasoz, 2004], [Kim et al., 2004])). Moreover, the use of wearable computer for emotional sensing (mainly based on physiology) allow new and more ubiquitous HCI set-up.

Considering together the PE formalization and manipulation possibilities given by computing systems, and the automated measure of individual affective state, computing provides the necessary tools to model our afctive relationship to multimedia contents.Another important requirements is the ability to measure emotion from computer. The next section focuses on the computing approaches which automatically perform indirect measure of emotion especially from physiological sensing.

2.4 Approaches of Systems Performing Physiological Indirect Emotional Measure

2.4.1 Methodology for Measuring User Emotion in Human Computer Interaction from Physiology

Research on emotion recognition from physiological signals has increased during last decade and is getting closer to achieve online recognition (i.e. best approaches tend to a 80% of recognition) either at the inter-individual level (using a common training database for different subjects) or at the intra-individual level (using one specific physiological training database for each subject). It is therefore becoming feasible to aim at building user-models including the user's emotions from that recognition process. However approaches takes diferent directions especially about the methodology involved to build classifiers and test emotion recognition.

We can consider that the generic process of emotion interpretation from physiologiy is made of two components (see figure 2.1). The direct measure is related to self-report, i.e. when we directly ask user about the emotion (s)he feel. In this case we have to interupt the user in its current situation and him to express the emotion in a given abstract emotion reresentation (see section

3.4 for an overview of possible emotion reresentation from natural cognition studies). The indirect measure is related to an interpreted measure in terms of emotion. The physiological measure is an example of such measure in the sense it need a method to translate physiological signal into an abstract emotion representation.

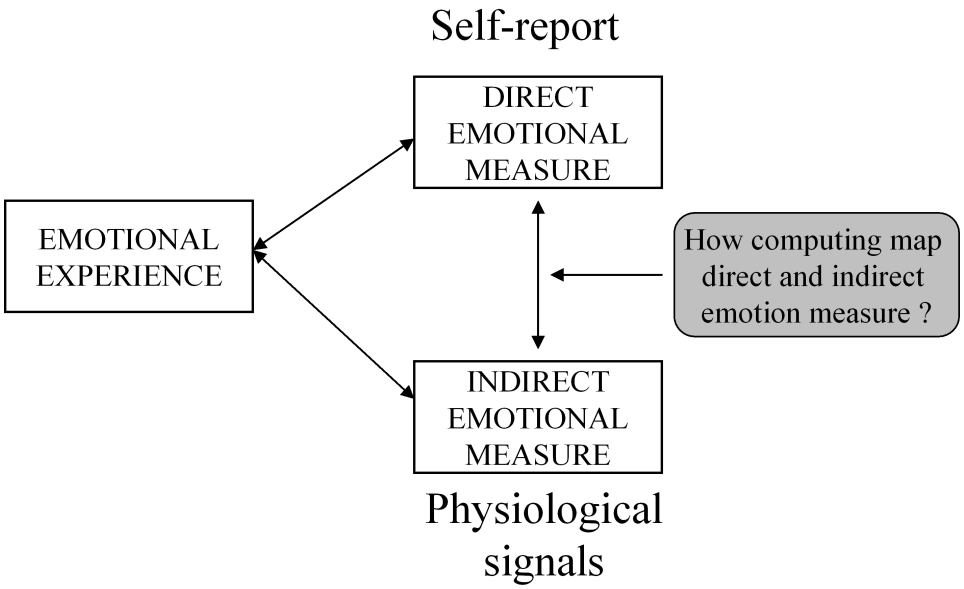

Figure 2.1: Computing approach on emotion interpretation from physiological sensing : Indirect physiological measure versus self-report direct emotional measure.

During last decade, an increasing interest for interpreting users' emotional subjective experience on the basis of physiological signals has led to various approaches. Among these approach, we can consider two main points : (1) considering the emotion recognition as a machine learning problem with minimal interest for physiology itself and (2) focusing on single user or multiple users experimental methodologies and models. We document theses notions in the sections

2.4.2 Machine learning approach and interest for physiology

Several systems presented recently use skin conductance (SC) and heart rate (HR, see section 3.4 for a detailled definition and discussion about theses physiological signals) and aims at extracting online, or near to real-time, a meaning of physiology in term of psychological parameters. Thoses approaches are mainly [Picard et al., 2001], [Lisetti and Nasoz, 2004], [Kim et al., 2004], [Anttonen and Surakka, 2005], [Haag et al., 2004], [Changchun et al., 2005]. Among thoses approach, we can point out some is-

sues. In this section, we will focus on the fact that most of theses approaches mainly consider the emotion recognition as a machine learning problem with minimal interest for physiology itself. They thus focus on performance and optimization of learning algorithms. However we can identify potential issues in the task of emotion recognition from physiological using such approach :

- ≻ The choice of features is not always grounded on physiological knowledge (e.g. mean, derivative)
- ≻ The aproach does not consider (nor embed) psychophysiological mappings from existing literature
- ≻ The psychophysiological mappings found are not precisely described (black-box)
- ≻ The output of the classifier could only be of one type of emotion representation (e.g. discrete or dimensional)

Firstly, the choice of feature set from physiological signal is a problematic approach as we know that SC and HR are not only modulated by emotion. For example, [Lisetti and Nasoz, 2004] and [Picard et al., 2001] used SC and HR signal with preprocessing as min, max, average, etc... of the signal during the stimulus presentation. The problem with such approach is that features such as Skin conductance Responses (SCRs) and Heart Rate Variablity (HRV, see section 3.4 for a detailled definition) in spectral domain seem to be cues of emotion (see section 3.4.3.3 and 3.4.3.4), while SCL and HR in time domain are more related to other parameters (see 3.6). Also, a suitable preprocessing is needed, before any machine learning or other means to build a classifier This is the approach recently taken by [Kim et al., 2004],using SCRs and HRV in frequency domain.

Secondly, as if theses systems justify the use of these physiological signals refering to previous psychophysiological experiment in the litterature, they don't re-use these result for the modelisation of links between psychological and physiological measures. Actually, [Lisetti and Nasoz, 2004] tested different machine learning techniques, [Kim et al., 2004] used Support Vector Machines, and [Picard et al., 2001] used SFFS-FP and k-NN, directly between their discrete emotionnal categories and the physiological features (pre-processed or not). Once the machine learning model is created, authors present their recognition scores of emotion, at inter-individual level (i.e. with a common training database for different subjects). This approach underlie two problems :

(1) As if such machine learning system are effective and robust to the differences in physiological signal interpretation, the limitation here is that we can't establish a common parametric model (embedding previous research) to describe precisely the link between psychological and physiological aspect of emotion.

(2) A machine learning system at inter-individual level couldn't point out qualitatively the inter-individual differences specificities of subjects, regardings the link between psychological and physiological. In other words, we can't model the difference from one subject other regarding the psychological semantic of a physiological signal, and we can't describe precisely the intra-individual differences (thoses described as "Day-dependance", in [Picard et al., 2001]).

2.4.3 User-Dependency and Subjectivity of Stimuli Methodologies

In this section we focus on the different exisiting approaches toward emotion recognition in regard with the chosen methodology. Some authors focus on single user or multiple users experimental methodologies and models and test the recognition at inter or intra individual level. We consider two important notion in regards with the methodlogy involved : the *user-dependency* and the *subjectivity* of stimuli used for learning and testing emotion recognition techniques.(1) The *user-dependency* of psycho-physiological data collected means that we choose to keep track of the specificity of individuals' responses or that we ignore such specificity). The (2) the *degree of subjectivity* of the stimuli used to elicit emotions means that stimuli with high level of agreement in terms of what emotional experience they elicit among a population can be chosen versus stimuli without such an agreement. We provide a brief review of existing approaches in terms of these two issues (summarized in Table 2.1 Whatsoever the approach, learning and testing is made on what we refer as a set of a psycho-physiological pair. The psychological part is made of a discrete representation (e.g. 'joy','fear') or a dimensional representation (e.g. 'high valence','low arousal') of emotion. The physiological part is a set of feature values derived from the physiological measure performed on the subject.

User dependency : Considering the individual's psycho-physiological pair or not consider them (by mixing or averaging the pairs) as potentially different among users.

- ≻ In a *user-dependent approach*, a set of psycho-physiological pairs built by collecting data of one unique user (often recorded during several days) is used. A machine learning is then typically performed between the psychological evaluations in terms of emotion and physiological measures performed on the user, and recognition rate estimation is achieved on the database for that unique user.

- ≻ In a *user-independent approach*, a set of psycho-physiological pairs is built by collecting data from several users, and it is then averaged (or mixed, i.e. grouping pairs of different users) after normalization among the population studied. A machine learning algorithm is then performed between psychological and physiological representations of emotion of the averaged (or the

mixed) studied population. Recognition rate is acheived on the database for any user, and still in the context of the experiment.

Subjectivity of stimuli : Considering or not consider the subjective evaluation of stimuli (by focusing or not on stimuli with a high level of agreement according to the emotion elicited across a population).

- ≻ In a *subjective rating of stimuli approach*, the user is requested to produce (via mental imagery) or to estimate subjectively the psychological emotion evaluations of stimuli, either with a discrete label (e.g. joy, fear) or on a dimensional spatial representation (e.g. in a 2D-space represent high arousal with a dot). This subjective estimation of the stimuli (self-report) is used for the training and testing.
- ≻ In a *social agreement of stimuli approach*, the user is not requested to estimate subjectively psychological emotion evaluations of stimuli, as a database of pre-validated stimuli chosen to have a high level of agreement among a representative set of a population is used.

	User-dependency		**Subjectivity of Stimuli**	
Authors	User Independent	User dependent	Social Agreement	Subjective Rating
[Picard et al., 2001]		✔		✔
[Haag et al., 2004]		✔	✔	
[Kim et al., 2004]	✔		✔	
[Lisetti and Nasoz, 2004]	✔		✔	
[Anttonen and Surakka, 2005]	✔		✔	
[Wagner et al., 2005]		✔		✔
[Changchun et al., 2005]		✔	✔	
[Villon and Lisetti, 2006]	✔	✔		✔

Table 2.1: Description of approachs of emotion recognition from physiological signals, from the user dependency and the subjectivity of emotion estimation.

The specificities of these approaches for emotion recognition are tested experimentally in section 5.

2.5 Conclusion

In this chapter, we shown the approach of existing systems related to physiological-based emotion measure and multimedia delivery system based on emotional information. This state of the art will allow us to consider what kind of knowledge from natural cognition could be interesting to consider to understand how to enhance such existing computing research systems.

Chapter 3

Analysis of Emotion Processing of Multimedia Toward a Computing Approach

3.1 Introduction

In this chapter, we provide a state of the art of natural cognition knowledge in a form of a critical survey of the literature to propose requirements for a computer science approach on an affective state and emotions human multimedia interaction. We first place our problematic within the affective science domain. Then we describe psychological and neuroscience knowledge regarding affective states and emotion elicited by multimedia contents. We finally discuss the requirements to adopt a psychophysiological approach for the extraction of emotion from the indirect measure of physiological signals involved in affective information processing.

3.2 Natural cognition Research knowledge needed for computing approach

We propose here a list of research question which we considered in this thesis at natural cognition level.

1. Perception, enaction, representation
   - ≻ What is perceived, built, stored during perceptual process?
   - ≻ How ?
2. Formalization of perceptible environment
   - ≻ Identification of cognitive variables.
3. Emotional evaluation and memory

- ≻ How the emotional memory (with LeDoux [LeDoux, 2000] definition) is used while one perform an emotional evaluation (and not how emotion modulates the memory encoding and retrieval) ?
- ≻ What is stored ?

4. Emotional system update, plasticity, learning:
   - ≻ Description of the phases of encoding and consolidation of the affective relationship to percepts.
5. Attention/emotion interactions within perceptual processes.
   - ≻ How the perception is modified by affective values of perceived groups of percepts ?
   - ≻ Which impact this has onto emotional evaluation?
6. Emotional measure: How could we interpret affective state from physiological signal of an individual, especially in reference to 1st person approach?
   - ≻ What are the robust results of interpretation of signals measured using 3rd person approach on the basis of 1st person reference?

3.3 Affective States and Emotion Elicited by Multimedia Contents

3.3.1 A subpart 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. The fields working on emotion are mainly: sociology ([Bourdieu and de Saint-Martin, 1976], clinical and experimental psychology [Winter and Kuiper, 1997], [Scherer, 1984], neurosciences ([LeDoux, 2000]), and more recently computing ([Picard, 1997]), with different explanations of same focus, or with few sharing of focus. Finally, and depending of the field, the level of organization studied is not always the same.

As this thesis study the ability to associate affective experience to PE (and so multimedia contents), we will focus on what has been done in the generation of emotion/evaluation of stimuli processes, while experiencing P.E. The study of the mechanisms involved into the production of emotion (behaviorally, physiologically or consciously measured) in specific situations

lead into different models according to the nature of the situation, and level of brain processing considered.

By situation we mean the immersion of an individual into its physical, social, goal-oriented and time-dependent environment. We can consider internal situation of the individual (its short term perceptions, emotion), as a recent history of this individual, and goal-orientation the expectation of the individual from this situation. The physical environment is mainly the PE, while social component is made of the social net made between individual and the other individuals involved into the situation. Table 3.1 simply summarize several causes of emotion eliciting, according to different examples of situation. In this conception of emotion generation process, we take position regarding the fact that logically, the type of situation in which the individual is involved could help to define the cause of emotion. The type of situation could engage individual action or not. We defined the individual state, as active or passive as counterpoint of emotion theories related to individual's engagement into actions ([Ortony et al., 1988], [Gratch and Marsella, 2004]), to express the fact that emotion could, or not, be built without a goal's fulfillment logic.

As if these causes reflect the entire emotion phenomenon generation, and are combined during emotion generation, only the last one is suitable to understand how someone could have emotion while listening to music, evaluate affectively a place, look at a painting, without acting, and without be embedded directly into a social situation. So, in such case, the main process involved in the P.E. affective processing, should be compatible with such a rather passive process. In the case of evaluation of PE, the minimum requirement is the last cause. For example looking at a painting, or listening alone to a music, engage an interpretation of physical environment (and so the PE), without active state (despite perception and cognition), and without any social situation. This defines the *subpart of emotion study in which we are interested in.*

Of course one can evaluate the PE in a specific social situation, and performing action. In this case the main emotion felt and or expressed by an individual will be the resultant of these different causes, according to the type of situation and individual state. Nevertheless, it remains that before to be combined to more complex model of emotions, we should study the process of human ability to associate affective experience to media, as specific above-described situations (like listening of music) allow considering it as a an independent process (i.e. listening to music could elicit powerful emotion without action or social situation). As this thesis focus on modeling the affective relationship with media, and for the above-mentioned reasons, we will focus on the thematic of "automatic affective processing" ([Houwer and Hermans, 2001]). The Figure 3.1 presents the use of EAR during evaluative process of environment. The EAR process firstly evaluate affectively the PE, and could (1) be followed by social and goal-oriented pro-

Situation		**Individual state**	**Cause of emotion**
Example	*type*		
Not succeeding to perform an action	Physical, goal-oriented	active	Immediate or recent action
Agoraphobia	Physical, social	Active or passive	Immediate or recent social situation
Looking at a painting, a dangerous animal, a graphical design	*Physical*	*passive*	*Use of previous phylogenetical heritage and ontogenetical learning of relationship to PE*

Table 3.1: Different causes of emotion according to the situation

cess generating emotion (e.g. and individual fail an action, in presence of people) or (2) directly produce emotion without involving (bypassing) other process (e.g. when someone listen to music alone). Finally, this passive and P.E. based generation of emotion doesn't mean only simple affective state are elicited, but also that complex emotion could occurs. Main examples could be found in aesthetical emotion, found to be blends of primary emotions ([Cupchik, 1994]) and musical emotion : for [Zentner et al., 2005] '"emotions in everyday-life and in musical contexts differ in relative frequency of occurrence".

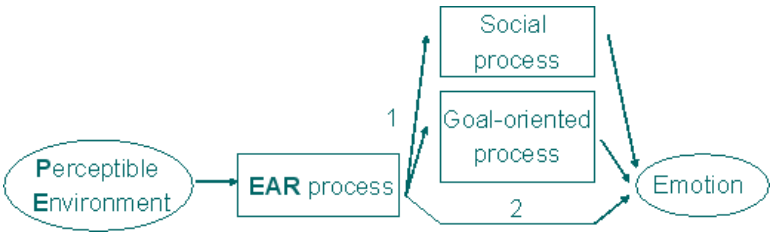

Figure 3.1: The representation of the E.A.R. process, during an emotional generating situation

3.3.2 Automatic and passive affective evaluation of P.E. : need for an unified theory

As we defined in the last section, we focus here on works relating to automatic and passive affective evaluation (which could be simple valence and arousal values or complex blends of emotions) of PE. [Houwer and Hermans, 2001] made a review of several model of emotion, regarding automatic affective processing. They point out that several models conclude that "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."

Main state-of-the-art models of emotion generation modeling distinguish different level of (brain) processing, and apply to generic or specific type of situation. By level of (brain) processing, we mean the consideration made by different authors to describe the emotion phenomenon at different level of cognition processes, from reactive, to symbolic/cognitive level.

Psychologist [Scherer, 1984] proposed that the generation of emotion is the "result of successive outcomes of a series of Stimulus Evaluation Checks" (SECs). Then, [Leventhal and Scherer, 1987] proposed the Component Process Model (CPM), a psychological multilevel model describing the processing level for SECs, made of sensory, schematic and conceptual levels. The CPM is a powerful approach as is embed different level of brain processing and apply to different situation. As if it was mainly based onto psychological approach, ongoing work relates it to computational neural model of emotion ([Sander et al., 2005]). The levels are recruited concurrently to generate emotion. Among this complete system, the schematic level, associated to the sensory one is an ideal candidate to describe the affective evaluation of P.E. While the schematic level deals with learning (ontogenetic) the sensory deal with innate (phylogenetic).

This notion of level serves to explain the evaluation of the incoming stimuli. In this evaluation, and as stated previously, we do not integrate all the checks from the conceptual level into the evaluation of the incoming stimuli (goals, conscious plan, and social factors like norm compatibility) to define the domain of the E.A.R. However, this should not mean that while the E.A.R. is constituted (i.e. not then we are evaluating stimuli using EAR) that only schematic or sensory level processing are involved. High cognitive conceptual processing (like extracting the notion of same 'object' from different perceptions) could enter in the EAR constitution, but are effectively not needed during the EAR using phase, denoted by the SECs of Scherer. The SECs of conceptual level which could be involved into the EAR use phase, could be the 'derived positive-negative evaluations'. [Fellous, 1999], [Arbib and Fellous, 2004] presented a neuroscience hierarchical description

of emotion expression and experience. The model is divided into four behavioral categories: Reflexes, Drives, Instinct-Motivations, and Cognitions. From Reflexes to Cognitions, the specificity is enlarged. As Reflexes are pre-programmed motor patterns, Cognitions could be any form of learning. In this case, the cause of emotion generation for PE could be varied. There are no here distinction between a conceptual level and a schematic one, as learning is mainly done into the Cognitions level (ontogenetic process), while innate (phylogenetical process) relationships to inputs are hosted by other levels. The cognitive level embeds any form of learning and is highly neuromodulated.

These different levels account for the possibility to associate affective experience to P.E., and learn these associations, questioning the notion of conceptual and cognitive process into the emotion generation. Affective evaluation of PE is a common property of all species, in term on motivational values (approach/withdrawal), and for evolutionary stimuli. However, the same mechanism of association should explain the affective evaluation of more abstract stimuli (abstract in the sense of no evolutionary meaning), in term of complex emotions (like aesthetical one), and in a non typical stimulus-response manner (as the famous historical Pavlov's experience making a sound salivating a dog trough conditioning), but as a continuous construction (like while we listen to music).

However such abstract stimuli (i.e. non-figurative music or painting) are processed by high brain structure (newer in the evolution) and so require high level of cognitive processing. So there is a paradox between the human automatic and unconscious affective processing of PE, possible with complex and abstract PE, resembling stimulus-response of a wide number of species and however not present into non-human species. For [Blood and Zatorre, 2001] : "music recruits neural systems of reward and emotion similar to those known to respond specifically to biologically relevant stimuli, such as food and sex, and those that are artificially activated by drugs of abuse. (...) Activation of these brain systems in response to stimulus as abstract as music may represents an emergent property of the complexity of human cognition". It is thus our cognitive architecture which allows us to interpret PE with the status of artifact, as well as non evolutionary relevant stimuli, as emotional, using (but not only) the same associative areas as other species. As we are able to produce complex emotion onto such PE, it means also that cognitive architecture allow us to retrieve complex emotional message into automatic and passive evaluation of the PE.

Thus we take the position that the suitable theory for modeling the E.A.R. should consider that cognitive analysis (not only perceptions) are associated on the same fashion than evolutionary salient stimuli, i.e. trough associative link between perceptions and affective states. However, the cognitive evolution leads in the fact that (1) some abstract and non-evolutionary relevant stimuli are processed using the same mechanisms that survival-

related brain circuitry, and (2) specific human cognitive process result on the possibility to produce complex emotion using theses mechanisms. With this conception, it should be possible to model in the same fashion the different kind of PE configurations of the Table 3.2. We argue that the E.A.R. is the common base to affectively evaluate such configurations of PE.

Type of PE	**Example**
Evolutionary relevant	A smiling face / a snake
For which the affective relationship had been previously constructed, by association or causality	A face of someone we like / be in a place resembling one where we had a setback
For which the cognitive contents is difficult to model	HCI designs / colors / shapes / abstraction
For which the emotion-generation is not related to goals	Listening to music

Table 3.2: Possible type of PE be evaluated by the E.A.R. with the same fashion

The above-mentioned models do not describe enough precisely, in terms of computations, the individual relationship to PE. However a recent model of emotion defined as an "explicitly-described computational architecture of the emotion system" has been made ([Sander and Koenig, 2002]). This model is the result of reviewing relevant papers from neurosciences, using computational analysis and justifying neurosciences as appropriate to study the whole emotional system. It provides a computational architecture of emotion, similar to those like vision, audition, etc. . . , and is an ideal candidate for the implementation of the EAR (see section 4.3.

However as if the model provides an architecture it doesn't focus on formalized and detailed mechanisms, toward implementation. It nevertheless gives pointers toward computational models of precise and specific mechanisms involved into emotion system (see section 4.3Following cognitive science approach, a model formulated in terms of logic and mathematic could easily become an active simulation of emotion phenomenon. However, there is actually no fully formalized theory/model in natural cognition regarding the human cognitive ability to associate affective experience to media, and so let-alone implemented simulation of the mechanisms involved into the emotion generation trough P.E. as defined previously. We saw that it exist several models of emotion, viewed as potentially embedding the fact of associating affective experience to media, not enough precise and formalized to be implemented as a functional use of the E.A.R. Moreover, it exist really precise models (mainly in neurosciences, see section 4.3) which could together lead into a formal and implemented system, standing for the E.A.R.

In next sections, we will point out main problems due to, in our sense,

few focus on fundamental concepts (considering stimuli as a configuration of PE, studying memory and emotion in depth) and on specific experimental methods (the need of experimentally work at intra-individual level) in next sections.

3.3.3 The need of focusing on inter-individual differences in emotional responses to media

Among diversity of methodological approaches used in the study of emotion, one important distinction is to consider the intra-individual differences as field of research, or as a noise (Villon 2003, unpublished work, : How analyze inter-individual differences in the expression of arousal induced by music?). The inter-individual variability study in emotional response for same stimuli across large population has been mainly considered as noise over past decades. This led into normative studies which demonstrate by their results that, for specific stimuli, broad population affective responses could be similar. For example, ([Lang et al., 1997], [Lang et al., 2005]) developed the International Affective Picture System (IAPS), and the International Affective Digital Sounds system (IADS) consisting of hundreds of images and sounds with associated affective values supposed to produce an emotional effect similar for everyone. Their results present little variability (intra-individual standard deviation of affective ratings for each stimulus is low) and so support a model of a shared affective relationship with the presented media among the international community. Such approach is useful for field of research of psychophysiology where we expect standard emotion for specific stimuli.

That said, it remains that media emotional eliciting could vary from one subject to other, as the existence of affective preferences for different media (music, films) among subjects is socially established. For specific stimuli it is probably not very relevant to compute an average of the emotional answers within a population, when a high variability is produced. Anyone can imagine the high level of variability we could measure by asking individual from different social and cultural background, of different ages, to rate a music belonging to the style 'hardtech'. In this view, a model of shared affective relationship with media among the international community is no longer supported, but we should talk of an individual affective relationship.

Adopting this position, we agree [Scherer, 2001]) for whom the psychology of emotion as focused on averaging emotional responses and not considered that each individual could react differently. Experimentally, it means that the problematic of difference of emotion for the same stimulus has been poorly investigated by psychology of emotion.

However, taking the music as example, ([Meyer, 1956]) wrote that "the difference [between non emotional states and emotional ones while listening to music] lies in the relationship between the stimulus and the responding

individual"). More recently, ([Panksepp and Bernatzky, 2002]) commented that "places culture and learning at the very heart of the musical experience". Here we consider that this statement could be extended to other media like video. Actually, people increasingly access wide media databases and create their own affective relationship to media with personalization of media access. Moreover, the social and cultural aspect of music affective meaning is largely modified by personal aspects. Several studies showed cultural differences due to everyday life emotional evaluation of various artistic contents.

To illustrate the *shared* versus *individual* affective relationship to media, we contrast as an example two approaches describing the links between compositional and perceptual process of artworks. The first one is based on Shannon's theory of information communication and encoding expanded to emotion communication and encoding, and the second is our proposed framework expanded from the Nattiez scheme to emphasis the individual property of affective relationship. As shown in Figure 3.2 (adapted from [Shannon, 1948]), with Shannon's theory an ideal listener 'decodes' an emotion (the message) according to the same referent used by the artist (e.g. composer, interpret or DJ). The emotional message building through music media in this case do not take care of the inter-individual differences problem. In this ideal case, it is the same code which is used to encode and decode an emotional message. This is what corresponds to a shared affective relationship.

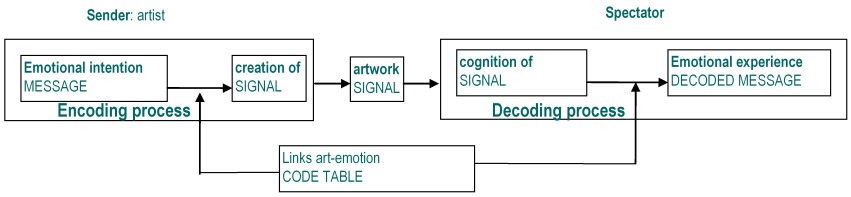

Figure 3.2: Emotional encoding and decoding of artwork, with a unique code table.

At the opposite, for ([Nattiez, 1990]), "a symbolic form...is not some 'intermediary' in a process of 'communication' that transmits the meaning intended by the author to the audience; it is instead the result of a complex process of creation (the poïetic process) that has to do with the form [e.g. the painting] as well as the content [e.g. the emotion] of the work; it is also the point of departure for a complex process of reception (the aesthetic process that reconstructs a 'message.')". Expanding this idea, our approach suggests that the decoding of emotional message is in fact an affective process building on the basis of the spectator's code table, separated from a code table that the sender uses to encode the emotional message. In this case, we do not talk of code table anymore, but can talk of E.A.R. Figure 3.3

represents also the emotional message building through artistic media, but taking care of inter-individual differences problem. The emotional message felt by the spectator could be different than the sender's message due to the difference of E.A.R. of each person. The emotional message is not any more decoded but projected, fully built by the spectator.

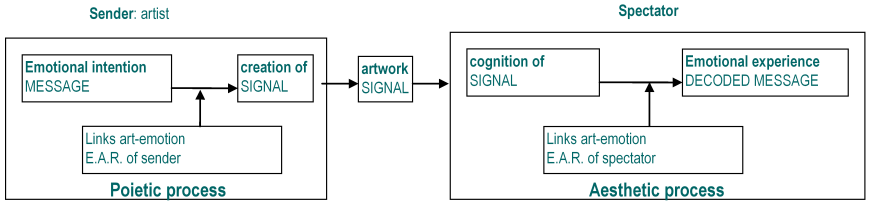

Figure 3.3: Proposed approach of encoding and decoding of artwork, adapted from Nattiez, 1990.

3.3.4 Considering stimuli as a configuration of the P.E.

Experimental studies (mainly in psychology, neurosciences and affective computing) analyze the responses of a subject, or user, to PE. While the affective/emotion related to P.E. (including stimuli, multimedia contents) had been intensively studied, PE had not been systematically described in depth.

Several studies of emotion related to PE use each stimulus as an irreducible entity. In this case, the formalization of the entities is a reference, or a simple cognitive descriptor label. [Bourdieu and de Saint-Martin, 1976], using a sociological approach, mapped everyday life object, home-design elements, or artwork, denoted by reference label (e.g. song title) with properties table of tastes, for some individuals. The IAPS and IADS, already-mentioned, propose a categorization of its items using cognitive descriptors; like 'car accident', 'knife', etc ...

Experimentally, this means that (1) each irreducible entity (stimulus) is displayed on a control neutral or constant background (there is no evolution of a continuous environment) and that (2) no subtle modification of the contents of each stimulus could be done.

At the opposite, other studies of emotion related to PE consider the stimuli as a configuration of generative (synthesis) and/or perceived elements. The approach of formalization of the stimuli is used with media for which the formalization could be part of the generation of the stimulus, i.e. mainly artifacts. The main example is classical occidental polyphonic music, for which score is an accurate formal model of the stimulus, at physical level (the formalization is what is played) and at perceptual level (the formalization is close to what we perceive-despite the timbre-). Moreover, the formalization could also work on cognitive level, like in

the study of [Krumhansl, 2002] which is an interesting example of this approach. Her study relates a cognitive formalized value describing music (the musical tension), with the direct expression of emotion felt. The same approach is often taken by neuroscience of conditioning, interested in intra-individual approach to emotions. To study the retention of conditioned responses to specific stimuli (especially the overlapping of stimuli features, see [Courvile et al., 2004]), an appropriate description in term of perception of the stimuli is needed. The physical characteristics are usually used to represents stimuli in theses experimental paradigms (like for instance, a "red square, measuring 7*7 cm, on a dark blue background (...) ; a red triangle of approximately the same size and displayed against the same color background", see [Marcos and Redondo, 1999a])

To summarize, we can consider the nature of the perceptible environment (music, image, video) as opposed to the deepness of the description of the environment used in protocol of emotion. On one hand, we have studies using a complex, long and not formalized stimulus (e.g. a film of 5 minutes) and on the other hand we have study using a simpler stimulus, short and formalized in term of perception(e.g. a music of 1 minute) from which all the events are known.

We argue here for the need of considering stimuli experimentally presented to subjects, or multimedia design presented to user of HCI devices, as a formalized configuration of the PE in terms of perceptive and cognitive units/primitives. The need of such consideration is motivated by (1) *the need to understand precisely the relationship between the cognitive processing of PE and affective state*, which involve knowing exactly what has been perceived. It is also motivated by (2) *the fact cognition is not a stimulus response system where the input are processed as irreducible entities, but is a dynamic process of analysis* of its environment, including perceptive and cognitive structuring. Finally it is also motivated by (3) *the need to be able to computationally manipulates such PE.*

Theses three conditions define the type of formalization needed. According to the first, the formalization used should be (in)directly related to the emotion process: the variable used to describe the PE should be considered as factor of emotion generation. The deepness of the description of the environment is thus considered as an important advance into emotional phenomenon description (the one which is focused by this thesis). Associating emotional measure of specific -formalized- configuration of the PE is an advance toward the possibilities of HCI on this basis.

According to the second, the formalization should be closely related to perceptive and cognitive process, i.e. the ideal case is the use of simulation of perception, followed by cognitive values extraction related to structure an temporal high level analysis (like auditory grouping : [Darwin, 1997]). This is the approach taken by [Leman et al., 2001] and [Martin et al., 1998] about music, using actual simulation of auditory human system, from perception

brain area to cognitive ones.

The last condition is related the analysis/synthesis paradigm and rises important questions regarding the formalization viewed as a generative process (like music) or an analysis one (like video analysis of films scenes). Thus, a configuration of PE would not necessary be formalized as perceptive/generative units (like in music), but will necessary be formalized as perceptive one. It means that computer could modify the contents in the first case, not in the second.

An important reason to not be interested to formalize the content is that in some case, the formalization seems to be not able to transcript the actual perceived contents of the stimulus. Let's take the example of the van Gogh painting. Obviously, due to the medium, it will not be possible to derivate perceptive/generative units (which would correspond to the paintings traces). However, it is possible to derivate analysis, e.g. with artificial vision system, and multimedia indexing approach. As the formalization done by analysis is related to perception principles, this will be an accurate representation of the PE, in terms of what is perceived by the subject.

The type of formalization could be described as perceptive vs. cognitive, and machine-oriented vs. human-oriented, or creation-oriented vs. perception oriented. Section 4.3 will detail the computing possibilities to formalize and use the PE (mainly Multimedia Indexing and Retrieval (MIR) approach).

3.3.5 Memory and emotion

Memory and emotion are widely studied together in the literature. Main publications study the memory modulation by emotion, i.e. how the encoding and retrieval ([Kensinger, 2004]) of memories is facilitated or inhibited according to the emotional context of encoding ([Pelletier and Pare, 2004] and [Hamann, 2001] for a review) and similarities between emotional context of encoding and decoding (mood-congruence phenomenon, see [Teasdale and Russell, 1983]).

Among memory and emotion studies, a term often used is *emotional memory*. As stated by the neuroscientist [Ledoux, 1995], "besides being a factor that can influence memory, emotional information can also be stored as a memory". Storing emotional information into memory, refers mainly to the episodic memory (explicit remembering of past emotional situation). However, it can also refer to the storage of affective properties of situations. In this case we not refer anymore to declarative and explicit memory, stated as "memory of emotion" by Ledoux, but to an implicit memory (created without awareness), storing affective properties of PE (the "emotional memory" for Ledoux). We will next use emotional memory in this meaning, and not the meaning of memory of emotion (like used by citeHamann2001) We consider this as an important point to consider for the modeling of the

user of any HCI device, i.e. such emotional memory could be a factor of inter-individual differences and should be part of the user model.

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 first question is highly related to affective learning studies , and so to the notion of results of learning (traces) created and stored after an affective experience.

Elements of responses could be found in ([Phelps, 2004]). The article study how the brain systems of 'memory for emotion' encoding (the hippocampus), and 'emotional memory' encoding (the amygdala) are interrelated. An affective experience leads into the formation of a memory for emotion, and for an emotional memory. The memory for emotion could modify the emotional memory while the emotional memory could modulate the encoding and retrieval of the memory of emotion. Moreover, a false memory for emotion, like the fact that someone tells you that a dog is dangerous, will next activate the amygdala when encountering a dog. Despite the fear was the main shown learning capability of amygdala, recent works of [Sander et al., 2003] showed that this structure is actually involved in other forms of emotional learning. So, despite different way to affectively learns, the results seems to be stored into the amygdala.

The second question is related to the involvement of the results of the memorization of emotional experiences (mainly in amygdala brain structure) while we affectively evaluate PE : How the storage/synthesis of emotion experiences into memory/integrate something in memory that will be used to have future emotional personalized experiences ? In fact it places the problematic of emotion generation in front of PE, under learning and memory field of research, which the interest is that we can use principles of memory theories to explain emotion generation from PE. Memory is usually a synthesis of incoming experiences, synthesis of learning. So, asking the question (2) requires knowing how the synthesis is done and what is then used to build emotion while experiencing PE.

We argue here that the questions (1) and (2) are central to model human affective relationship to PE, aiming at integrate the inter-individual differences, and use the model for building interacting systems based on multimedia ; theses questions are central in the notion of E.A.R. and will be discussed in details in next sections.

Attempts has been done to implements system based on emotional memory, mainly to evaluate words emotional meaning ([Yoshida and Yanaru, 1995], [Liu, 2003]). However, more general model (not domain-dependant) are neuroscience one, at low level (basic learning): [Balkenius and Morén, 1998a], or higher one (cognition): [Sander and Koenig, 2002].

3.3.6 Cognitive neurosciences of emotion and conditioning as an appropriate reference for modelling the E.A.R., with respect to intra individual differences

Among different experimental approaches, we consider that neurosciences (and psychology) of associative learning are the best candidates for computational modeling of E.A.R., and its implementation.

Firstly, neurosciences focuses on affective learning (trough classical and evaluative conditioning), studying in detail the possibilities of linkage between our PE and the expression of affective state.

Then, neurosciences use formalized stimuli (with precise description of its contents), so we know what is perceived precisely. This corresponds to the requirement of the P.E. Then, as interested in the concept of learning, neurosciences make repeated affective measure on a single-subject (mainly animal, but also human). The intra-individual approach is a central concept regarding the contents of affective-perceptive linkage, while the rules of such linkage are examined as species-dependent property. This allows an objective consideration of emotion inter-individual subjectivity regarding PE.

Then as examining the perceptive and cognitive processing of information at neural level, cognitive neurosciences approach gave important information regarding the formalization of perceptive to cognitive processing of PE (like describing sound as cells activities until an auditory grouping phenomenon). Thus, the notion of cognitive representation of the environment is mainly viewed as higher process onto perceptual one. This allow considering formalization of environment as a continuous form, from perceptual to cognitive (see for instance the work of [Leman et al., 2001] with music)

Finally, as focusing on brain structures, memory and learning, and cognition/emotion relationships, cognitive neurosciences propose interesting framework to study the affective relationship to PE. As if no detailed and implemented modeling of emotion relationship to PE, with systematic multi-level description of environment, has been proposed, theses approaches combined together places the neurosciences as an appropriate source of information for the design of the E.A.R. model, and so will next be focused on.

3.4 Psycho-physiological measure of emotion

3.4.1 The representation of emotion

Huge debate in emotional modeling is the representation of emotion. The two main opposite approaches are discrete (e.g. labels like joy, anger, fear, etc...) and dimensional (mainly the valence*arousal*dominance space). Both approaches have advantages as we explain in next sections.

3.4.1.1 The affective dimensions

The affective dimensions are continuous, and so are interesting for real-time monitoring of user emotion. Moreover, this mode of representation allow the computation of dynamics in the state of the user, which is needed to represent emotion as an evolutive process. The chosen dimension are the valence and the arousal. Valence (pleasure/displeasure) is related to motivational system of approach/withdrawal. Arousal(calm/excited) is the excitation state of the user, the bodily activation, an intensity of emotion.

3.4.1.2 The discrete emotions

The discrete emotions are interesting when we search to match or avoid a specific emotion of the user, like frustration.

3.4.1.3 Conversion from dimensional to discrete

The possibility to convert from one representation to other is given by a discrete-dimensional model of affective space : the Circumplex model [Russell, 1980]. In 1980, Russel provided a valence*arousal space where discrete emotions can be plotted (see 3.4.1.3), with an appropriate coordinate for each discrete emotion. As we should keep advantage of each dimension,

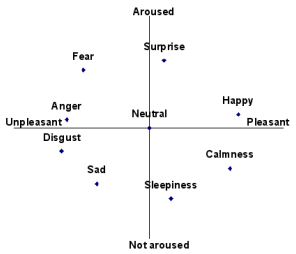

Figure 3.4: The circumplex model (adapted from Russel,1980)

the chosen approach is to focus on dimensional representation (i.e. the system will use a dimensional representation) but which easily convertible into discrete representation.

Nonetheless, the possibility of mapping is prooved between some discrete emotions and affective labels and position in valence and arousal space. As stated by [Scherer, 2000] : "Researchers have been able to locate particular emotion labels in clearly indentified regions of the two-dimensional space, independently of the langage or culture in which these studies have been conducted".

However, following Scherer ([Scherer, 2000]), two important points should be considered. First, the shape of circle of the coordinates in the space should be taken with caution, as increasing the number of adjective change considerably the shape. Secondly, the possibilty of translation form a coordinate to

an affective label is also higly related to the semantic of choosen verbal label, which semantic could vary. Thus, such conversion should be considered as a simplification.

The notion of simplification is actually the goal of such translation from coordinates to classes, but at the same time keeping a more detailled representation (the coordinates).

3.4.2 Respect the inter-individual differences using self-report to study emotion

We formalize the elicitor as the situation which elicits emotion, and the affective experience the evaluation performed by an individual of this situation and made of psychological and physiological components. Theses two components are the expression/measure of the affective experience.

Considering the physiological and psychological evaluations as the ouptut of a system evaluating this elicitor, we can isolate two levels of potential inter-individual differences. The first level is the existing differences between the psychological and physiological component according to individuals, which is what a part of this thesis aim at adressing, as denoted in section 5.

Figure 3.5 presents schematically the inter and intra individual differences between psychological and physiological components. Left subfigure is a well-known example of average statistical rule between skin conductance and the subjective arousal. Inter-indiviudal differences (middle in figure 3.5) has been shown by [Fiorito and Simons, 1994] : some subjects can show a mapping between psychological meaning and physiological measures which is different compared to average population. Intra-individual (within an individual) changes can also occurs between psychological meaning and physiological measures from a day to other ("Day-dependance" phenomenon [Picard et al., 2001], see right subfigure).

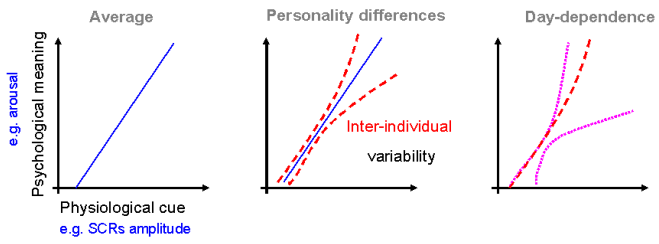

Figure 3.5: Inter and intra individual differences in the psychological and physiological representation associated to emotional situation.

The second level is the existing differences between the elicitor and the affective experience. As pointed out in [Villon, 2002],[Villon, 2003], and next discussed in the chapter 4 the way individuals evaluate elicitors could be considered as being dependent of the personal history of each individual. At

computing level it means that the self-report evaluation of subject, and its analysis which is based on intra-individual methodology instead of averaging self-report affective responses of several individuals, is required for a suitable interpretation of emotion.

3.4.3 Emotionally-specific choice of features from Autonomic Nervous System (ANS) signals

3.4.3.1 ANS description and psychological information contents

Autonomic Nervous System (ANS) is driven by brain structures, related to emotional processing. The ANS role is the regulation of the organism. We can access indirectly to it by several measures like heart activity (mainly the rate) and skin electrical properties (conductance). As we can have indirect access to ANS activity, trough above mentioned type measures, we should be able to extract different kind of information.

Specific peripheral measures of ANS activity gives information about behavior. For example, [Meste et al., 2005] and [Kettunen and Keltikangas-Jarvinen, 2001] heart rate activity contains breathing information, known as RSA.

Therefore ANS is modulated by emotional and other factors((see 3.6). For example, [Friedman and Thayer, 1998] showed that ANS could be modulated by personality trouble. We focus here on the emotional information contained in ANS, and peripherically measured trought the Heart Rate (HR), and the Skin Conductance (SC). These two measures are non-invasive and were found to contains auseful information about ANS activty.

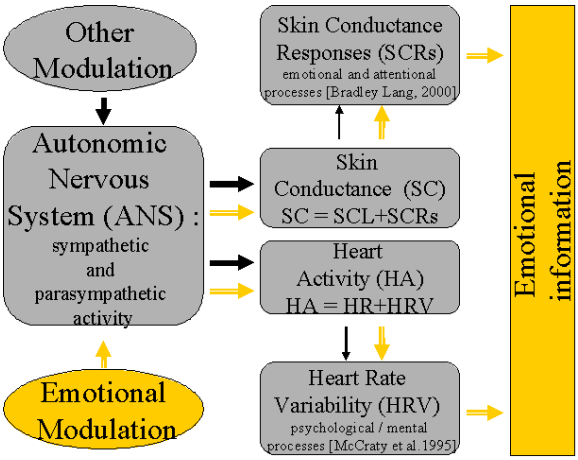

Figure 3.6: Emotional modulation of ANS and peripheral measure of this emotional information.

Heart Rate. The heart beats are autonomously controlled by the sinoatrial node situated at the top of the heart[Meste et al., 2005]. The parasym-

pathetic and sympathetic branch of nervous system modulates the HR (short term modulation), as well as norepinephrine realeased in the bloodstream by the adrenal glands(long term modulation)[Fiorito and Simons, 1994].

Skin Conductance. The skin conductance (SC) is the electrical conductance measure at the surface of the skin. We apply a tension of reference to the skin, and then measure its variations ([Fowles et al., 1981] and [Malmivuo and Plonsey, 1995]). The modulation of the SC is the result of neurons activity, afferent from the the sympathetic system, which controls the sudoripar glands of the skin. The sympathetic system is, as for it, innervated by cortical and sub-cortical structures involved in emotional and motivational system. More precisely, the neural substartes of SC are mainly amygdala and hypothalamus ([Critchley, 2002]).

Both HR and SC modulation are thus peripheral measures of cerebral activity.

3.4.3.2 Emotional information in ANS : Psychophysiology

The possibility of measurement of the emotions starting from physiological data such as the skin conductance, had been proven in experiments by many articles. An interesting philosophical review of the suitable inference of a psychological semantics starting from physiology is given in [Cacioppo and Tassinary, 1990] (see page 6-7 for a specific criticism of SC and HR). It is a fundamental problem for the measurement of the emotions, corresponding to the discipline of psychophysiology of emotion. [Lisetti and Nasoz, 2004] made a review of more than thirty articles associating experimental physiological measurements with their emotional interpretation. This abundant literature legitimates the use of physiological cues for the measurement of the emotions. Recently, [Christie, 2002] and [Christie and Friedman, 2004], [Kettunen, 1999] showed that ANS can discriminate affective states, both in dimensional and discrete representations.

Actually, specific cues contained in SC and HR were found to be related to the expressed emotion of subject.

3.4.3.3 Heart Rate Variablity (HRV) and emotional information

Heart activity analysis allow extraction of heart rate (HR) and related heart rate variability (HRV). HR and HRV has been found related to physical efforts, but also to cognitive activities, as meditation ([Takahashi et al., 2005]) and emotional processing. Direct analysis of HR, like in [Fiorito and Simons, 1994], showed that HR is involved in emotional valence processing ; [Palomba et al., 2000] showed that an HR acceleration occurs when subject are looking at unpleasant films. HRV analysis is a more difficult task, but gives more precise information about the changes of the heart rate. Mood and emotion influences HRV. For example, [Hughes and Stoney, 2000]

found relation between people with depressed mood and their HRV, compare to person with normal mood. [Jonsson and Sonnby-Borgstrom, 2003] showed the effects of picture of emotional faces to HRV. [McCraty et al., 1995] demonstrates that "anger produces a sympathetically dominated power spectrum" while "appreciation produces a power spectral shift toward MF and HF activity". Morevoer, [Izsó et al., 1999] shown high correlation between MF and negative subjective experience ('difficult' situation).

HRV could be assessed in time-domain and frequency-domain (see section below). Several studies focused on emotion in the time-domain, and few in the frequency domain. However, the Frequency domain seems to contain a lot of useful information for emotion, as it reflect how the ANS is modulated by emotional factor (see 3.6). Thus, we will study here the relationship between felt emotion, and the HRV measured in frequency domain.

3.4.3.4 Skin Conductance Responses (SCRs) and emotional information

The SC is an index of subjective arousal, in other words we can say that the psychological semantic of the skin conductance is the arousal. For example, the experiment of [Lang et al., 1993] showed experimentally a covariance between amplitude of SC of subjects exposed to stimuli, with the 1 st person evaluation of these same stimuli. We usually distinguish two major information into the raw signal of SC. Firstly, the Skin Conductance Level (SCL) which represent the basal level, the tonic component of arousal. The SCL reflects the general arousal level of the organism. The second signal is the SCRs which are transitional responses, contained into the raw SC signal, with a characteristic form. They are characterized by fast variations of the SC signal, and doesn't take into accounts the slow changes of the raw SC. Theses responses seems to be correlated with uncounsciouss emotional and attentional processes, which interested us.

Moreover, [Bradley and Lang, 2000](page 213) and [Cuthbert et al., 2000] (table 3, page 106) showed experimentally statistical correlation of arousal and SCRs.

In addition, valence could not be retrieved systematically from this signal, as if some poor correlation were found, and without any discrimination of the valence sign. For example, ([Lane et al., 1997] paragraph *results*, page 1439), found strong correlations from the arousal (subjectives measures) and the SC signal, while for valence, they only found correlations discriminating 'neutral' and 'pleasant or unpleasant' conditions.

So, we should find intra-individual (within-subject) correlations between SCRs (amplitude, duration, number), and the subjects' affective ratings in arousal.

Nature, duration and elicitation of SCRs. The SCR is a charac-

teristic response obtained from the SC raw signal, with a typical duration of 1 to 3 seconds [Malmivuo and Plonsey, 1995]), which can be extended to 10 seconds between the beginning and the complete return to the SC former level (before the SCR) ([Lim et al., 1997]).

SCRs are elicited between 200ms and 1 second after the presentation of a new stimulus (see [Nuechterlein and Dawson, 1998], paragraph *Measures of Electrodermal Activity*). The modification of deep sudoripar glands is extremely fast. It is not only a superficial sudation, but actually the skin resistance, modulated by deep skin layer (see [Malmivuo and Plonsey, 1995], who provide an electrical model of deep skin layer electrical modulation).

The SCRs can be elicited by stimuli of really short duration (see a study with subliminal images, i.e. around 30 ms, inferior to liminal level [Ohman and Soares, 1994]).

The stimuli which elicits SC reactions could be of different nature, duration and modality. For instance :

- [Khalfa et al., 2002] use music
- [Moller and Dijksterhuis, 2003]use odours
- [Palomba et al., 2000] use films

For a deeper description of the features of the stimuli, the synthesis of [Lisetti and Nasoz, 2004] presents a collection of studies using different stimuli. However,we will not list here the kind of stimuli which can the stimuli seems to not follow a specific typology. Indeed, [Fowles, 1986] wrote that "The stimuli that elicit these responses" (i.e. the SCRs) "are so ubiquitous that it has proved difficult to offer a conceptualization of the features common to these stimuli". This means that potentially any situations could elicit emotion, and thus any multimedia contents may elicit emotion.

3.5 Conclusion

In this chapter, we presented a critical survey of the litterature both related to affective state and emotions elicited by multimedia contents and psychophysiological interpretation of affective information from physiological signals involved into affective processing.

This analysis showed that user modeling approach are needed to solve these two associated problematics. We will introduce in next chapters two proposed user models. Firstly the Embodied Affective Relationship to Multimedia contents to model affective information related to multimedia contents in an implementable perspective. Then the PsychoPhysiological Emotion Maps model and approach will be presented to interpret physiological signals in terms of emotion.

Following a cognitive science approach due to the multi-disciplinary context of the affective computing, both models are built upon natural cognition consideration to contribute to the design of the computer-based models.

Chapter 4

Modeling Affective Relationship with Multimedia Contents (E.A.R. Model)

4.1 Introduction

4.1.1 Understand the individual's relationship to computer controlled media toward a multimedia design

In this chapter we argue the position that understanding the individual's relationship to computer controlled media, and formalizing it in depth should be done in order to make accurate tools toward an automated user's emotion-based multimedia selection, modification and design. We focus on the requirements to design a multimedia interactive system based on user emotions, along with an implementation.

Thinking of an interactive and automated emotion-eliciting multimedia system could lead into the enhancement of existing applications. It could expand the interaction of the user in traditional HCI, by modifying the interface multimedia appearance. Imagine, for example, that the color interface, associated to the background music, changes continuously according to the user's emotion. It could also lead in the development of new applications and design of novel forms of interaction, as the dynamic creation and modification of immersive artistic multimedia environment, according to the user's real-time sensed emotion.

4.1.2 Approach at cognitive level

Cognition is construction of the relationship to the environment. Motivation, which is an important subsystem of the cognition, is based on the general mechanism of seeking pleasure and avoiding pain. The motivation is regulated by activity but is also oriented by attributing affective properties to elements of the environment.

This chapter focuses on and studies one of the human cognitive ability which is the capability to associate affective experience with media. This means that the subject is able to link specific perceptive/cognitive properties of its environment with some affective properties. This phenomenon is illustrated by several everyday life behaviors which directly engage people with their own affective relationship with environment they experience. For example, people search (new) music to listen (through concerts, shops, internet portals), experience artistic multimedia performances and buy artworks, choose and arrange elements of their home.

4.1.3 Research question

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 chapter, 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. We present the 'Embodied Affective Relationship' (EAR, which basis had been presented previously [Villon, 2002, Villon, 2003]) 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 4.1 our main interest 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)).

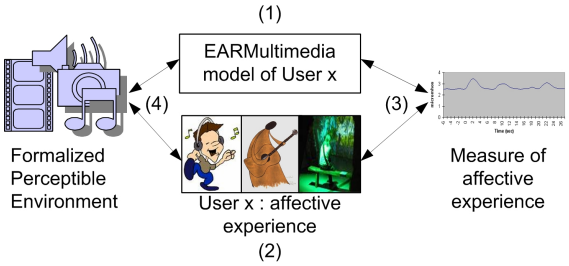

Figure 4.1: The problematic of the EAR

4.1.4 Emotional communication trough affective objects

The proposed approach could participate to model emotional communication trough affective objects. We 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 4.2 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 [Scheirer and Picard, 2000] 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 may 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 propose to formalize the basis of such associations with the Embodied Affective Relationship model ([Villon, 2002], [Villon, 2003], [Villon and Lisetti, 2005a]) 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. 4.2 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. 4.2 the EAR of the spectator). For [Nattiez, 1990], "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 mainly focus here on the LTAM component toward its implementation.

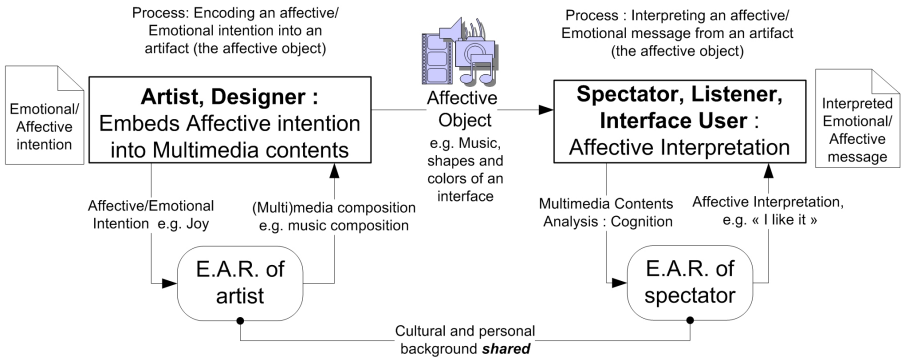

Figure 4.2: The proposed formalism for an emotional communication trough an affective object (e.g. music ; color of an interface)

4.1.5 Model outline and potential use in HCI

We provide an overview (fig. 4.3) of the components to design to build the *EAR* to Multimedia model by considering three scenarii: Simulation,

Poïetic-modification and Poïetic-generation.

Simulation mode (top of fig. 4.3) 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. 4.3) 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. 4.3) 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. 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). 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 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 a model from cognitive science theories, and (2) fill this user model with data learned from each user ([Villon, 2003]).

4.2 An experiment to measure the inter-individual differences involved in emotion communication using media contents

The experimental work presented in this section has been performed upon a set up proposed by members of the Humaine Network of Excellence (http://emotion-research.net)and members of the InfoMus Lab staff (Camurri, A., Castellano, G., Cowie, R., Glowinski, D., Knapp, B., Krumhansl,C., Volpe, G. , http://www.infomus.org) during the Humaine Summer School (September 2006). We proposed an addition to the original set up : the measure of the emotion recognized by spectators. The goal was to measure the emotional communication trough a musical performance in order to study

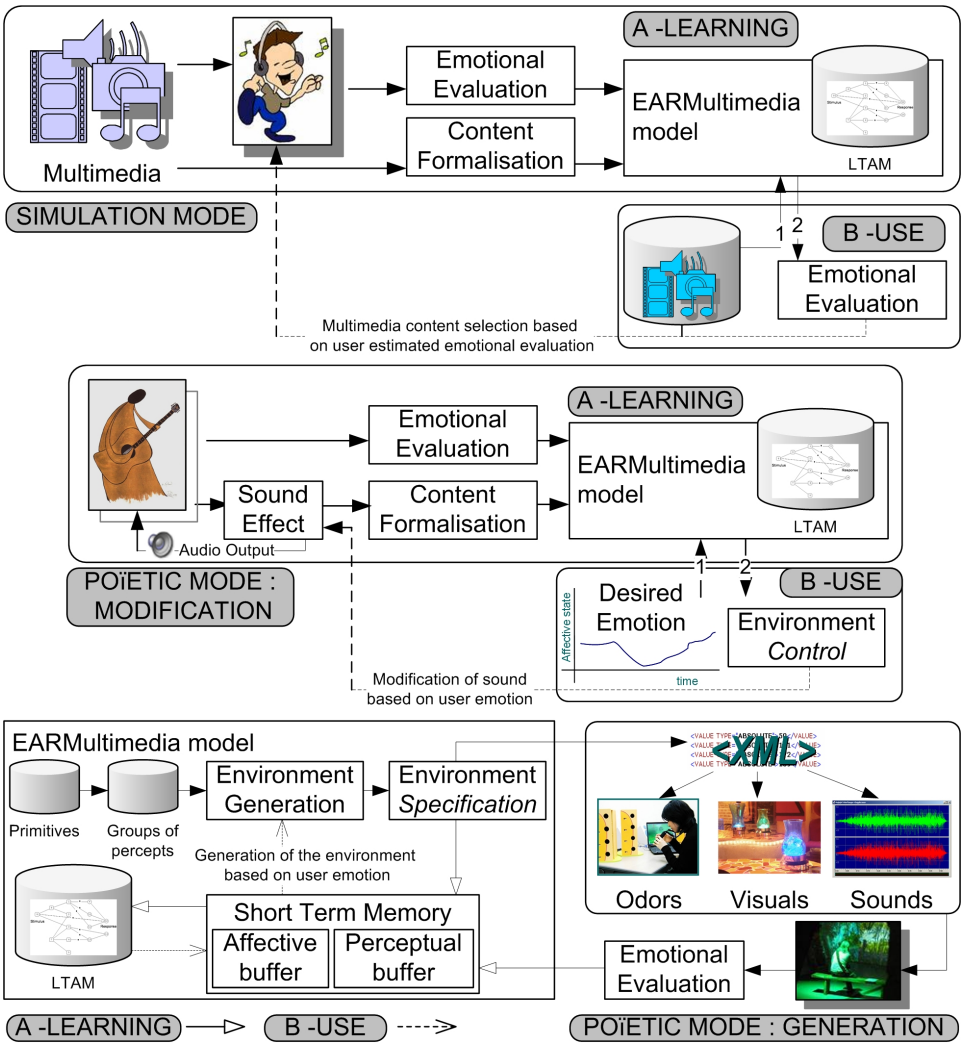

Figure 4.3: Modes of the model, and examples of potential applications : Simulation, Poïetic (modification) and Poïetic (generation).

the inter-individual differences involved and the need for a user modeling approach between the multimedia contents (i.e. music) and the affective information (the emotion recognized by spectators). This experimental work thus introduce our proposed model aiming at associate multimedia contents description and affective information.

4.2.1 Material and Methods

We performed an experiment in a concert hall at the Casa Paganini in Genoa, in the form of a concert-experiment. Figure 4.4 summarizes the set up. We asked a professional violin player (the performer on fig. 1: a semi-finalist of the Paganini International Violin Competition, Diana Jipa) to interpret four times a musical piece (a canon by J.S. Bach - the Composer on fig. 4.4- from the Musical Offering). Before each interpretation, an experimenter proposed a discrete emotion to play to the violin player and asked her to interpret the musical piece following this emotion as a target.

While the composer had its own expressive intention while composing the musical piece, we aimed to test how the performer may convey the emotional target by modulating the original score. Among the spectators, 31 listeners were requested to rate the emotion the player interpreted using a closed choice questionnaire.

The instructions were: "You are going to listen and see a violin player interpret four times the same piece of music. At each interpretation, player will express an emotion. Please fill the following questionnaire after each interpretation: "Which emotion this player aims at expressing?"". The questionnaire was made of four ten-points Likert scales, associated to the emotions Angry, Sad, Joy, and Serenity for each interpretation.

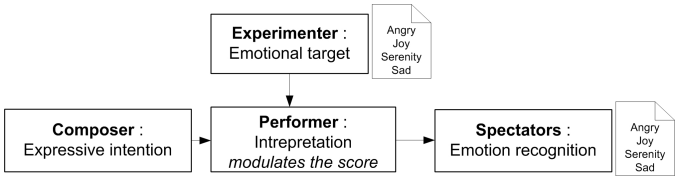

Figure 4.4: The experimental protocol to assess the spectators' recognition of the performer emotional interpretation

The questionnaire was translated in the language of the audience (Italian), and verified according to the recommendations from [Scherer, 1988] to comply with the semantic of discrete emotion labels.

4.2.2 Results

We tested if the emotional message was accurately recognized by listeners. Figure 4.5 presents the results of the study. The four proposed emotions

were recognized by the spectators : each best amount of perceived emotion estimated by spectators matches the emotion proposed to the performer.

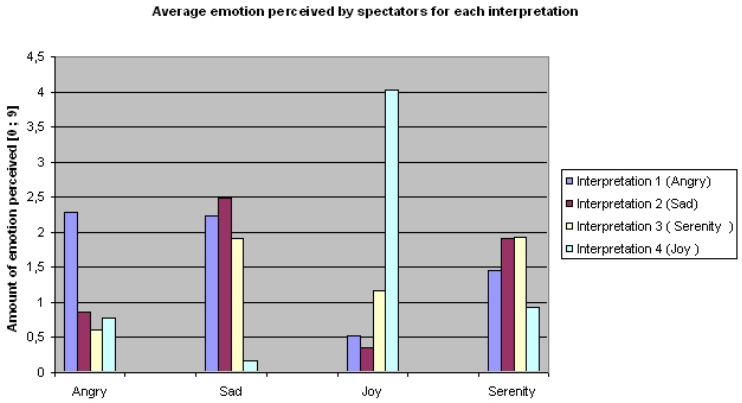

Figure 4.5: Average emotion perceived by spectators for each interpretation.

To analyse the accuracy of the recognition, we considered the task of the audience as a classification problem. Thus we first built a confusion matrix (see table 4.1) by considering the spectator population as a single 'classifier': each spectator response was a test case). To achieve this representation we considered the classified class as the class with the maximum value among the four Likert scales. Using this representation, we found that the recognition rates were respectively 43, 46, 71 and 43 % for the classes Angry, Sad, Joy and Serenity, with a global recognition rate of 50.7 %. However, by considering the whole spectators population, and considering the maximum recognition rate for each emotion class, the recognition was successful for all the emotions.

	Classified			
Target	Angry	Sad	Joy	Serenity
Angry	43	33	7	17
Sad	11	46	7	36
Joy	11	4	71	14
Serenity	7	29	21	43

Table 4.1: Confusion matrix of discrete emotions classification (in percent of cases of target classes), by considering the spectator population as a single 'classifier'

Joy was the best discriminated emotion, followed by Sad, and equally Angry and Serenity. Sad and Joy were reciprocally the best discriminated at intra-class level. False recognitions (e.g. recognizing Joy instead of Sad) are all inferior to true recognitions. However the sum of false recognitions

for each emotion target is superior to true recognition, despite for Joy.

Then, we not considered the population of spectator as a single group, but searched for clusters among the spectators. We again considered the recognized class as the class with the maximum value among the four Likert scales, for each spectator and for each interpretation. Then, we computed the number of emotion classes successfully recognized by each spectator. Table 4.2 presents the results of this approach which consists of the percentage of spectator who recognized at least 1, 2, 3 emotion classes, or recognized the 4 emotion classes.

	Amount of spectators (%)
At least 1 emotion recognized	90
At least 2 emotions recognized	56.6
At least 3 emotions recognized	30
4 (all) emotions recognized	20

Table 4.2: Amount of spectators who recognized at least 1,2,3 or the 4 emotion classes.

Thus, as if only 20% of the spectators succeeded to recognize all the emotions, 90 % recognized at least one emotion, meaning that communication of the emotion from the musical performance worked, even partially. We will consider in the next section the interpretation of these results, in regards of the notion of emotional message communication.

4.2.3 Discussion : the need for modeling affective relationship to media contents

Results show that (1) emotion targets were successfully communicated from the performer to the spectators, if we consider the true recognitions versus isolated false recognitions and even if the global recognition rate is low. Results show also that (2) almost all spectators recognized at least one emotion while only few recognized all the emotions. This result may show that emotional communication worked, but only partially. As schematized in figure 4.4, the performer interprets the score, and thus the expressive intention of the composer, to translate her representation of the target emotion into musical modulations. Then, the spectators listen to this musical interpretation and recognize a specific emotion.

The fact that spectator match or not the target emotion may therefore be related to the notion of communication and its underlying processes as discussed in the section 4.1.4. How could we describe the process which leads the performer to embed emotion in the interpretation, and how could we describe the process which leads the spectator to identify an emotion in the heard interpretation? Following the formalization proposed in the figure 4.2, we applied this to the context of this experiment (see figure 4.6).

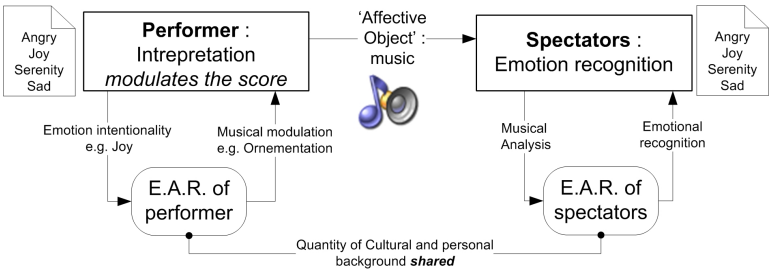

Figure 4.6: The proposed formalism of media affective encoding and decoding applied to the performer expression and spectators' recognition.

The results of this experiment may be explained by the quantity of cultural and personal background shared between performer and spectator (and thus the similarities of the contents of the EAR). Such similarities may explain the amount of spectators who recognize emotion or not. We may consider that 90% of the spectators share one emotion expression one the four of the performer, while only 20% share totally the way of expressing emotions of the performer (in the context of this four emotions).

We consdidered the hypothesis that performer encodes an emotion message (e.g. Joy) into a audio media contents following its own expressive rules (associating music representations and emotions. Then, a spectator decodes an emotional message on the basis on its own expressive rules. Investigating such hypothesis needs a formal and systematic usable computer model of the associations between media contents description and affective information. This is what we present in the next sections.

4.3 E.A.R. conceptual analysis and Multimedia model

The introduced notion of E.A.R., defined as the affective relation anyone could learn, synthesize, stores and then use with its environment is a latent notion, both in the scientific and artistic research, and in the human practice. Previous sections lead into the conclusion that (1) multiple fields of research focus on the E.A.R. without referring at a unique concept (from learning theories, modern aesthetic, to multimedia indexing) but actually have a converging approach, and (2) multiple field of system design (from multimedia affective delivery to interactive art based on emotion) does not rely on a unified theory and modeling of human affective relationship to environment.

However, the positions of the engineer, the artist and the cognitive scientist, regarding the constitution of an accurate model of the proposed notion of E.A.R. are complementary.

Firstly, we could consider that the engineer aims at finding a solution to be able to manipulate multimedia contents (from music artificial artifact to multimedia HCI interface). As mentioned by [Hanjalic and Xu, 2005],

talking about the modeling of emotion in video sequence: "We see the possibility for further improvement (...) in searching for more concrete relations between the affect dimensions (arousal and valence) and low-level features. (...) The relations known so far are rather vague and therefore difficult to map onto reliable models for arousal or valence components". However, the engineer has practical tools to manipulates multimedia contents (video and audio players), to analyze such contents (e.g. MPEG-7), select parts of this contents. Thus, from the engineer point of view, the tools exist to relate the PE to affective state of user, but it remains a need regarding the description of the E.A.R.

Then for the cognitive scientist, the cognition is construction of our relationship to environment (we project information onto our environment and we extract information from this environment). Among this relationship to the environment (including the P.E.), a subpart is realized as an affective phylogenetic and ontogenetic structuring with this environment, due to motivational needs, and related to cognitive activity of categorization. This subpart of the relationship to the environment, constrained to the P.E., correspond to the introduced notion of E.A.R. Aiming at build model of human cognition by bibliographical work and experimentation, cognitive science produced several models. A cognitive neurosciences architecture of emotion (see 5.1.1.1) provide guidelines for an accurate model of emotion, but does not precisely describe the relationships relating affective experience and P.E.; Associative learning theories (see 5.1.1.2) describes with specific focus (on modality, procedural set up, etc...) the learning of new affective relationships with stimuli, along with the relationship of theses learning with phylogenetic heritage. However, several form of associative learning exists, and no unified model at the level of the relationships exists. Thus, from the cognitive scientist point of view, the models provided point toward a precise description of the E.A.R., according to the possibility of merging theses models.

Finally, for the artist, and art theorist, interested in hedonic artworks design, like musician, the E.A.R. could be viewed as the basis which allows the conversion of an emotional intention into an artifact (which represents the P.E.). These human artistic design practices, regarding music, paintings, and films involve a good understanding of the E.A.R. Historically viewed as a non-personalized E.A.R., common within a population, the relationship between emotion and the artworks were described as explicit compositional rules. Art history explicitly stated rules of beauty changing over centuries, from musical notes structures to sculpture inter elements ratio. These rules were guidelines to produce aesthetical experience in any person within a population. E.A.R. is also viewed as an implicit embedded complex process. Art theorists try to understand the reason of artworks aesthetical interest, trough artwork analysis, by extracting structures, internal relationships. The formalization of the artworks in term of perceptible environment

is thus associated to affective experience, and theses relationships tried to be explained. Moreover, inter-individual differences (see 3.3) and less theorized recent practices (as electro acoustic music) lead into highly personal and specific artworks which accounts for a non explicit, and personalized E.A.R. Thus, for the artist and art theorist point of view, the E.A.R. could be an explicit set of relationship between affective experience and artworks, varying for periods and places, or a fully embedded EAR, varying from one person to other, poorly described.

As the main objective of the E.A.R. approach is to emotionally deal with perceptual and cognitive multimedia contents, taking into accounts the personal and cultural background of everyone, we understand the benefits for theses above-mentioned approaches.

In this section, we will firstly discuss the theoretical and computational works which constitute the basis of the E.A.R., i.e. principles of natural cognition, and computer system of artificial and simulated cognition, as the result of the federation of different fields of research. We will details next the system of the E.A.R., both in natural cognition (i.e. how the E.A.R. is constituted and used by human) and in artificial cognition (i.e. how we can model the E.A.R. of an individual and use it to simulate his affective state regarding PE, and how to modify PE to follow specific affective state requirements).

4.3.1 Theoretical and computational basis

We describe in this section the bibliographical requirements for an accurate elaboration of the E.A.R., i.e. what is needed theoretically and from computing to build a computational an implementable model of the human cognitive ability to associate affective experience to media ?

As mentioned in the previous sections, the automatic and passive affective evaluation of the PE is studied without an unified framework and thus no model regarding the process of emotion generation with media is done, let-alone detailed description of the E.A.R. However, several models and theories point toward the E.A.R., and we consider that by combining several approaches together, and with an appropriate conceptual analysis, we can explicitly describe the E.A.R.

According to the key problems identified in the previous section, we will focus on and combine approaches of psychology of emotion and cognitive neuroscience approach. Mainly, we will focus on intra-individual research, taking the position of Scherer regarding the fact that inter-individual difference should be fully embedded into the emotion theory ; and considering inter-individual differences as an indicator of intra-individual personal history and not only a noise. Moreover, we will focus on relating the contents of the media with some affective properties, not only for simple stimuli.

Make an accurate model of the E.A.R. requires studying natural cogni-

tion, artificial cognition, and norm for perceptible environment. We identified at the beginning of the section 3 a list of main theoretical questions to solve. We add to this questions from artificial cognition domain here.

Artificial and simulated cognition: formal and computer science level. (1) Pattern recognition into multimedia contents: how to extract structure into PE? (2) Systems to represents emotion: which reliable implementable representation of emotional state of the user to use? (3) Memory (associative) model: which reliable implemented model of associative memory could stands for the host of learnt affective relationship with PE ?

Norms for perceptible environment. Formalization of perceptible environment for the representation and the synthesis: (1) Which languages exist to represent the environment (like derivate from XML). ? (2) Which systems/languages could be used to synthesize PE from computers (like midi engine, etc...) ?

All theses question together correspond to the needs required to build the model in depth. In the next sections we will address some of them, and just give interesting pointers for other questions as they involve more complex problematic of cognitive science, and are out of the scope of this thesis.

4.3.1.1 Natural cognition principles used in this model

In previous sections we point out why neurosciences are more useful to study the E.A.R. Moreover, the emotional memory, discussed previously lead in a conception of a personalized affective evaluation of PE, according to past experiences. This section discusses principles from neurosciences and psychology of conditioning and associative learning which are the basis of the E.A.R. formalization. The main question of this section is: How we register and use specific elements of the environment with specific emotional states, and what we register and use precisely in the environment ?

4.3.1.1.1 The Sander and Koenig computational model The cognitive neuroscience model of emotion of Sander, Koenig and Kosslyn [Sander and Koenig, 2002] is a computational (not implemented) and functional architecture based on a state of the art of emotion study, from cognitive neuroscience point of view. It is thus a complex model, providing an holistic view of the phenomenon of human emotion.

However, as we seen in section 3.1 we are interested only in a specific subpart of emotion phenomenon: the affective relationship to PE, built on the base of the EAR. So we describes here not the full architecture, but only several (highly simplified for the need of implementation) principles and components among their complete model of emotion, to provide guideline for the design of the E.A.R., extended from what was stated in [Villon, 2003]. The motivation of simulation for engineering justifies the choice of this state-of-the-art model.

Motivated by the design of an accurate E.A.R. model, we can extract from this cognitive architecture, several functional (FP), computational (CP) and somatic-based (SP) principles, which could be viewed as arguments for the design of our model.

The FP1, "to detect relevant stimuli", place the affective properties of PE as an indicator of relevance. Thus, it embeds the emotion within attention, and research of saliency into a cognitive function. This justifies the notion of E.A.R., in the sense that the emotion-based relevance of PE should be based on systems which allow the emotional evaluation. This means also that to simulate the EAR, perceptual systems should take care of the notion of attention. We can for instance, model that the more the stimuli are relevant, the more they are present to our attention.

The FP6, "to permit emotional learning" is directly related to the central aspect of the EAR, i.e. the fact that our evaluation of PE is partly based on previous affective experience with this environment, which are learnt and reused. Indeed, it provides guideline regarding (1) one of the components of the EAR: a data structure which should be able to host a result of the past emotional learning, and regarding (2) the method to fill and use this data structure: firstly, model the result of individual's learnt emotional reaction and, secondly, uses this user model to predict possible emotions of this individual regarding specific PE.

The CP1, "Exogenous and endogenous inputs", and CP2 "Mental representations as inputs" means that we can produce by ourselves the inputs to be evaluated (like mental imagery, etc...). Thus, it indicates that we should take into accounts as inputs any endogenous inputs, which could be elicited by exogenous inputs (a stimuli could elicit the remembering of other stimuli or situations). So, this principle should lead in a network-type memory, when an input form the PE could activate neighbors in term of perceptual similarity, and then feed the emotional system as inputs.

The CP11, "Outputs as Inputs", and SP1, "Internal-state patterns are informative", mean that the affective evaluation of the environment (and its corresponding bodily activity) should be taken into consideration as an input of the emotional system. Taking into account that our present emotional reaction depends on what we recently felt, this CP and SP lead into the conception of an emotion rising as a dynamic and recurrent process. Emotion generation from PE is not a simple dictionary of responses according to identified stimuli, but is a real time process, build dynamically. The current affective state of an individual should be taken into account to accurately model the dynamic of this affective state. So, we need in the EAR, a recurrent input regarding emotional output. Moreover, the CP1 and CP11 together lead into the need to integrate a kind of short term memory buffering structure, both at the perceptual level (we call it Short Term Perceptual Buffer-STPB) and at the affective one (we call it Short Term Affective Buffer-STAB), to accounts for what is actually perceived and emotionally

produced.

The CP3, "Variations in eliciting potential of exogenous stimuli", means that we are genetically prepared to process certain class of stimuli (evolutionary-salient stimuli) which are preferentially able to elicit emotional reaction. This drives the design of a partial phylogenesis basis of the EAR.

The CP5, "Automatic evaluative mechanisms", is an argument to design the evaluation of PE as an automatic process, and thus designs algorithms of evaluation without take care of volitional data.

The CP6, "Emotional expression", places the output of an emotional system as an expression of the evaluation, including physiological one. We are interested in the experience of individual, means to measure it trough expression. Physiology is an interesting indirect measure of the emotional state, we aim at use as an interpreted emotional measure of emotion, which could be used in the EAR model ([Villon and Lisetti, 2005b]). We will now examine the main three components which could be used within the EAR. (1) The system of stimulus responses connections, implements "processing reflexes". This system 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). The main brain structure hosting this function is the amygdala which will be deeply discussed in the section 4.3.2.1.

The existence of such system leads in the need to embed a associative data structure in the E.A.R., associating stimuli description and affective evaluation of theses stimuli. We can consider this system as a dictionary of stimuli and responses, resultant of phylogenesis and learning. (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. Among the whole PE, a "stimulus is identified when the input closely matches the features of a stored object" into associative memory which reactivates "internal state patterns on the basis of inputs from the subsystems involved in exteroceptive processing". This orients the design and inclusion into the EAR. of a calculation of distance process between the input and the stored representations.

Combined to internal state pattern activation subsystem, and to subsystems related to emotion generation, we propose to simplify this associative memory and its related systems for the EAR modeling as a [PE patterns]-[potential affective value] relationships system. We use relationships instead of connections due to the notion of "pointers toward interoceptive and exteroceptive representations", and not a dictionary as in stimulus response subsystem. We use PE patterns instead of stimulus, because the representa-

tions are a large concept of cognitive science and could range from a complete perceptual structure (i.e. a stimulus) to a simple feature (e.g. a change of pitch). We use potential affective value instead of response, because in the case of stimulus-response connection, the response was a direct output of the system (i.e. behavioral and/or conscious), here the affective state should follow other processes before to be an actual response (e.g. merging multiple affective evaluations related to multiples patterns detections, etc...).

Thus, the EAR structure hosting PE and affective values, will be made of the concept of this stimulus-response connection subsystem ("processing reflexes"), along with the PE patterns-potential affective value relationships system. We will next define how we combine it in the formalization of EAR in natural cognition (see section 5.2), into a general Long Term Associative/Affective Memory (LTAM). Finally, (3) the somatosensory buffer deals with our current and recent past internal state. Its inputs are one's current internal state (internal physiological feedback) which are, formally, queued. It is linked to the construction of the time-dynamic of emotion. It will be an affective type of short term memory, in the EAR, previously called STAB. So the emotional evaluation of PE will not be a direct stimulus-response reaction but will be modulated by current and recent past internal state.

We defined the general architectural requirements for the design of the EAR, simplified from the cognitive science modeling, but following natural cognition state-of-the-art. We will next describe the mechanisms of learning and internalization into memory, to precise the nature of the relationships of the EAR.

4.3.1.1.2 Learning affective relationship to P.E. As we pointed out in the section 3.5, and mainly defended by [Ledoux, 1995], emotional memory should be considered as a key to understand individual relationship with PE. Moreover, as stated by [Sander and Koenig, 2002], one of the functional principles of an appropriate architecture of emotion is to permit emotional learning (defined as the acquisition of new connections between environment and affective experience). As it is a central point for the design of the EAR, this section aims at define in detail this notion, from affective experience, to learning and finally synthesis into emotional memory. The focus of this section is not to model in detail the process of this learning, but we will simply describe it to then be able to describe precisely the result of such learning onto the future affective evaluations produced by an individual, and thus contributes to the justification of the notion of E.A.R.

Phylogenetic affective memory. A lot of preprogrammed affective response to 'evolutionary salient stimuli' ([Sander and Koenig, 2002]) has been shown. Theses correspond to previous learning of the species, embodied trough evolution processes. We can conceptualize that such preprogrammed response are stored into a phylogenetic affective memory (implemented as a

part of stimulus-response subsystem, involved in evolutionary-salient stimuli processing).

The types of stimulus-response range from the long term evolution (like fear of rapidly incoming object, which is present in all mammals) to more short term (like human face emotional expression, present in humans). Innate emotional responses of infants. Several stimuli elicit innate responses. In early childhood, infant almost already discriminates all main facial expressions.

Universally shared affective reaction at adulthood. Several stimulus response reaction still exists at adulthood. From emotion expression faces to injuries, affective evaluations are common among world human population (see the IAPS and IADS already discussed in section 3.3). Theses universally shared affective property of a specific PE could be the result of a common phylogenetic affective memory. Inter-individual difference in the affective responses of same configuration of PE (see section 3.3), logically raise two type of hypothesis trough the view of emotional memory: (1) theses evolutionary salient stimulus-responses had been modified within several individual, trough learning mechanism (see research on modification of unlearned fear of [Eifert et al., 1988]), or (2) theses configuration of PE are not evolutionary salient, meaning that theses stimuli were not related to specific affective responses at early childhood, and thus their affective properties had been learnt. In theses two situations, we should take into accounts the notion of learning.

Learning into ontogenetic affective memory. Learning, memory and emotion are intrinsically related ([Blanchard et al., 2001]). We propose to oppose the concept of ontogenetic affective memory to the phylogenetic one to accounts for the emotional learning results.

Emotional learning during early childhood. In the early childhood, cognitive construction of relationships with our environment is necessary to be able to generate appropriate actions. These relationships include the affective one, which will presents to understand how infants learn affective property of elements of their P.E. It has been found that most of the affective relationship infants create with their environment is done under the mechanism of 'social referencing', by looking at the facial expression of their parent. For example, in the study of [Sorce et al., 1985], infants are placed in front of an artificial cliff (a "visual cliff" made of a glass-covered hole, which seems to be a cliff). The cliff was situated between the infant and his mother. While the infants hesitate how to deal with such situation (a novel situation), it has been found that the facial expression of the mother oriented the choice to cross the cliff or not. Infants associate situation's valence according to caregiver's facial expression, and thus internalize the affective evaluation of their environment, by their caregiver's. As if this mechanism was primarily found for situations, it has also been found with object of the environment, like the study of [Klinnert, 1984]. Series of new toys were

presented to infants, using the same paradigm of the previous study. The infants presented reaction of attraction/avoidance as a function of mother's expressions. [Mumme and Fernald, 2003] showed that 12-month-olds can use social information presented on a television, to learn affective significance of new object. The retention of this learning was tested and proofed to be internalized.

Emotional learning after early childhood.As if the transition between during early childhood and after is not a direct transition, it is interesting to conceptualize it as two periods regarding the acquisition of affective properties for new stimuli. In the early childhood (1), child use the evolutionary-salient stimuli as reference to give new stimuli affective properties, by the mechanism of associative learning, thus infant use phylogenetic affective memory to build its new affective relationships with environment, i.e. to build its ontogenetic affective memory.

After the early childhood (2), children, and adults still builds new stimuli affective properties on the basis of the phylogenesis, but also have the possibility to use ontogenetic -previously- built affective relationships with environment. This is usually called high order conditioning (including mainly second-order conditioning). For instance, the experiment of [Levey and Martin, 1975], shows emotional learning could occur on the basis of non (necessary) evolutionary salient stimuli: pictures of paintings. They firstly asked each subjects to sort the stimuli from disliked, neutral to liked, then requested them to select the two pictures they liked the most, and the two they liked the less. This individual sorting reinforce the idea that the affective relationship with theses pictures is the result of ontogenetic emotional learning, due to the fact to ask each subject to perform this sorting. Then the four pictures were paired with the previously sorted neutral one, using the paradigm of evaluative conditioning (we will detail this notion in next paragraphs), to execute an emotional learning. The result of this experiment is that the pictures primary sorted as neutral by the subject were evaluated with a shift in direction to the affective evaluation of the pictures (most liked and most disliked) used for the emotional learning.

The interest with such ontogenetic learning on the basis on previous ontogenetic learning is that, depending of our experiences with the PE, high differentiation could occur, explaining inter-individual differences in the affective evaluation of PE. Thus this answers to the need of focus proposed in section 3.3., and precise the design of the LTAM proposed in section 4.3.2.1.

Mechanisms of emotional learning. The main mechanism responsible for the emotional/affective learning is the associative learning (we will not describe here non-associative learning as habituation, dishabituation and sensitization, to focus on the main form). This regroups two forms of learning, the first passive, and the second involving the action of the individual. An overview of formal models could be found in [Balkenius and Morén, 1998b]. The classical conditioning (Pavlov) refers to

the learning of a Conditional Response (CR) in presence of a Conditioned Stimulus (CS) after controlled and repeated presentation of this CS along with an Unconditional Stimulus (US), which originally elicits a Unconditional Response (UR) (evolutionary salient of previously learned). This had been mathematically formalized by [Rescorla and Wagner, 1972]. The second form, the operant conditioning (Thorndike, then Skinner requires the action of the subject, compared to the rather passive process of classical conditioning. In this case, the CR is actually the action of the subject, related to the affective value of the expected result of this action. For [Donahoe and Vegas, 2004], the most important mechanism in learning is not the CS-US (with UR), but the CS-UR. So any affective experience (elicited by any kind of US, i.e. any affective situation) could lead in the apparition of the UR to the CS ,according to specific timing between CS and UR (see also [Marcos and Redondo, 1999b] for a study about latency between CS and US and its effect on the strength of the conditioning).

The first interesting aspect is that behind theses responses, we can consider the existence of an affective value related to theses responses. Thus, can consider that there is an acquisition of an affective value, as a component of the CR (classical conditioning) or as a motivational source for the CR (we generalize the notion of UR as embedding there is an acquisition of affective value. The second interest is the common aspect of these two mechanisms: the fact that elements of the PE could acquire no endogenous (i.e. phylogenetic evolutionary stimulus-response) affective value, trough an associative mechanism. This is a strong argument to explain how the result of complex process involved into the generation of affective experience during an affective situation, which are only activated during the learning phase, could be acquired as an affective value associated to any elements of the PE which is present during the acquisition phase. For example, consider that you have repeated dispute with someone in as specific place. The complex process (social, goal-oriented, etc...) is an affective situation which elicit affective experience. The environment in this case could (under specific conditions, as reinforcement, which is out of the scope of this thesis) be associated to your current affective experience (the UR). Then, if you encounter a similar place, you could find that you don't like this place.

The more focused field of research on mechanism of transfer of affective value is the evaluative conditioning. It is a specific form of associative learning (a specific classical conditioning) which study the like and dislikes (i.e. the valence components of dimensional models of emotion) of stimuli ([Houwer et al., 2001], [Houwer et al., 2005]). It is particularly interesting in our context to understand such specific mechanism as the valence of PE is highly engaged into the selection of multimedia contents, like music delivery.

Finally, another interesting form of associative learning mechanism is the configural conditioning ([Pearce, 1994]), which aims at not considering a stimulus-response connection, but look in detail at the links which are done

between elements of the PE with affective values. In this case compound patterns form the stimulus, which is a more precise description of the underlying mechanism involved into associative learning, to understand how to model the relation between patterns inside configurations of the PE with affective states.

To simply summarize this section, we can consider that some patterns of our PE could be associated to affective potential, trough an associative mechanism, i.e. when an exogenous indicator paired with a specific configuration of PE allow us to learn the affective significance of this configuration of PE (i.e. social referencing, conditioning, etc...). In this case, the exogenous indicator could be any situation (from perception to action) which leads into an affective experience. This is what we will consider into the formalization of the EAR in natural cognition (see section 4.3.2.1).However it remains to precisely define the nature of the generated affective relationship, i.e. what is learned after an associative learning, to accurately model the contents of the LTAM.

4.3.1.1.3 Emotional memory: contents and internalization processes The aim of this section is to understand what is learned into the proposed notion of LTAM. To achieve this goal, we will describe (1) the nature of the internalized affective relationship to the PE, to then to be able to model the contents of the LTAM, and (2) the process of their internalization into the LTAM (emotional memory), to then be able to learn the LTAM of users from practical affective measures.

Nature of the internalized affective relationship to the PE. We described in the previous section the mechanism of associative learning by itself (i.e. the transfer of an affective experience to a configuration of the PE, creating a new affective relationship to this element of the PE). This means that the nature of the affective situation, and thus stimuli used to elicit the affective experience could be of any types, and that the elements of the PE on which the learning occurs could also be of any types. For instance, [Houwer et al., 2001] showed a similar transfer of valence could occur in unimodal domain (visual, gustatory, haptic, etc...) and cross-modal (e.g. visual-auditory).

That said, we should understand the nature of the affective relationships. We saw in the last section that the form of the LTAM is mainly made of relationship between representation of the PE and representation of affective state (included in the responses and related to internal state). How could we formalize such relationships? [Gewirtz and Davis, 2000] made a review of studies focusing of the nature of theses internalized relationships. They consider the higher-order conditioning, as a "window on the architecture of [emotional] memory" and conclude that emotional learning lead into the formation of emotional memory contents on the form PE representation;

emotional internal state, for ontogenetic learning based on previous ontogenetic learning. However, it seems that for first-order conditioning, i.e. when learning occurred on a phylogenetic basis, the relationship into the memory is of the form PE representation; phylogenetic stimulus representation; emotional response.

Then, we should (a) understand on what in the PE emotional learning could occurs, and (b) what is the nature of the related affective part of the relationship? We list here some different experimental results which will allow us to formalize the type of learned affective relationship with PE (see section 5.2.1). We start by enumerate forms of the PE which were be submitted to emotional learning (i.e. the target of learning). We take arbitrary studies, as the number of existing studies in experimental emotional learning is numerous. Static elements of PE were used, as paintings ([Houwer et al., 2001], as well as dynamic ones, like sounds ([Büchel et al., 1999]). Features, as color hue, were used ([Robinson, 2004]), as well as complex patterns, i.e. involving many features, as paintings. Moreover, absolute values or patterns, like a fixed sound tone ([Büchel et al., 1999]), or relative values or patterns, like music structure were used. Thus, affective properties are related to the PE with different forms, including several dimensions.

Then, we enumerate forms of the affective property which are memorized after an emotional learning. Discrete property, like the fear emotion ([LeDoux, 2000]), or dimensional, like the valence ([Houwer et al., 2001]) were acquired. Absolute property, like fear emotion, or relative property, like e.g. a shift of valence value regarding previous value ([Houwer et al., 2001]) were also demonstrated. Static value or pattern, like the valence value, was found as well as dynamic value or pattern, like the discrete emotion as fear. Indeed, instantaneous affective response, like valence value for a picture ([Houwer et al., 2001]), or delayed response (occurring after a specific pattern) like skin conductance responses ([Marcos and Redondo, 1999b]) was shown. Finally, we can consider the existence of direct response versus inter-items (ordinal) affective property (A have a more positive valence that B). Thus, PE could activate different forms of affective properties.

Process of affective relationship to PE internalization into the LTAM. Considering that an emotional/affective memory is not only a list of stimuli and responses, and that memory is not a simple repertory of added data (see the [PE patterns]-[potential affective value] relationships system in last section) but at the opposite a consistent memory structure : the result of the learning is considered as a consistent internalization.

One main important condition to make the memory consistent is the fact that the process of storing new incoming affective relationships to PE should be related to the previous existing content, in some way. For example, the fact that an individual creates a specific relationship linking an object A with the joy emotion, and then this individual creates a relationship of fear with another object B in some extent similar to A, the memory network of

the LTAM should be able to use these learning in next evaluation of object A and B. The notion of internalization is thus a synthesis of the learning into memory. What is this synthesis process? The fact that the learning lead into new relationships, we consider that conceptually the learning updates an affective memory consistently regarding previous content into this memory. We already see what is learned after an associative learning, we should now define how we combine new relationship with those already learned, or phylogenetically learned ?

Elements of responses could be found in the Sander and Koenig model: the LTAM links, with a pointer system, i.e PE representations with affective representations. Considering a new associative learning, we logically should compare the affective experience associated with the perceived PE with the representation presents in the LTAM. Thus involve a research method to extract similarities between what is currently experienced during the associative learning with what was previously stored. If nothing is similar, new representations are created (both for PE and affective experience) and new pointers are created. If some elements are already presents, only new pointers should be created to links elements already present in the LTAM. We will detail this mechanism in the next sections.

4.3.1.1.4 Applying the notion of compositionality to internalization of learning into memory, and emotion generation. As we denoted in the last section we can simplify one of component of the architecture of human emotional system as a [PE patterns]-[potential affective value] relationships system. In this case, we do not any more consider that an incoming stimulus (we will henceforth consider denote a stimulus as a configuration of the PE) elicit a response, but that some detected patterns inside this configuration of the PE elicit potential affective values.The problematic of this section is thus to known how the such patterns are related to potential affective values, within more global configuration of the PE.

To answer this problem, we should introduce and consider the notions of *compositionality* versus *emergence*. Compositionality is a term from linguistic and semantic, meaning that the interpretation of the totality is a function of the interpretation of the components of this totality. For [Teller, 1992], a property is emergent if it is not functional (with the mathematic definition). So, applying theses concepts to the E.A.R., it means that we can (if E.A.R is compositional) or cannot (if E.A.R. is emergent) compute and explain the affective value attributed by an individual to a configuration of the PE, on the basis of affective values attributed to each groups of percepts which compose this configuration of the PE. To schematize, we could imagine that the green color is associated to a specific affective state for an individual (e.g. a low arousal), and a specific shape is an emotionally significant pattern for this individual (e.g. associated to a high arousal) due to his everyday

life affective learning. The actual affective state of this individual will be a combination of the separate values, if the LTAM is compositional, or an independent value, if the LTAM is emergent.

The notion of compositionality should be understood at the internalization stage (Is affective property transferred to the CS, by emotional learning, retro-propagated onto each perceptual component?) and at the generation stage (Is the affective evaluation of the PE use combination of affective potentials or is just the reactivation of a stimulus response dictionary?).

A functional argument for the compositionality could be found in [Sander and Koenig, 2002] with what we simplify as the [PE patterns]-[potential affective value] relationships system. This tends to describe the emotion generation as a real time generation on the basis of pointers linking affective potentials with elements of the PE. This system of pointers could host complex relationships between partial PE representation and affective properties, not as simple as a dictionary of stimulus responses.

Responses regarding the opposition of theses two concepts could be found into formal models of associative learning. As introduced in the last section, configural associative learning is an interesting method to describe the nature of the stored affective relationship. As stated by [Pearce, 2002], two opposed models of the associative learning are the elemental and configural theories.

In the elemental theory, the acquisition phase provide "the opportunity for each element of the compound [i.e. the P.E. made of several identified patterns used as CS] to enter [individually] into an association with the representation of the [affective experience used for the associative learning]".

At the opposite, the configural theory is "based on the principle that a representation of the entire pattern of stimulation that constitutes the [P.E. used as CS] will be formed and will enter into a single association with the [affective experience used for the associative learning]". In the Rescorla-Wagner model ([Rescorla and Wagner, 1972]), the emotional response is made of a linear combination of stimulus-responses, meaning that the holistic affective property of the complete CS is retro propagated directly (linear) to each stimuli composing the CS. For ([Pearce, 1994], [Pearce, 2002]), with the configural conditioning model, the emotional response is made of a compound patterns which form the stimulus, which is entirely related to the response.

Theses two model does not seems to be easily combined into a clear one, without any exceptions, as concluded by [Pearce, 2002] : "An alternative and ingenious elaboration to the Rescorla-Wagner theory is the replaced-elements theory of Wagner and Brandon (2001) (...) compound. The advantage of this theory is that it is able to make predictions that are consistent with either the Rescorla-Wagner theory or configural theory. (...) Not all of the findings considered in this article can be explained by the replaced-elements theory of Wagner and Brandon (2001)". Thus as if a solution could be found by the cited model, it remains an open question. As this debate is still an ongoing one, we can conclude that both emergence and compositionality

exists, in the affective relationships to PE.

On one hand, we can consider that each elements (patterns, structures, features, etc...) of the PE can be associated to affective potentials in a specific relationship. The combination of these potentials could (but not necessary) correspond to the affective value attributed to give the actual response. In this case, we know the relation between patterns inside configurations of the PE with affective states, and thus could predict emotion generation regarding combination of affective relationships associated to elements of the PE. On the other hand, we can consider that the entirely main pattern describing the PE is related to a specific affective response. In this case, the compositionality mechanism is limited by the notion of irreducible entity. In section 3.4, we criticized the focus on such entity, as semantic ones, to the profit of more compositional affective relationships. Moreover, in section 4.3, we saw that on one hand we find stimulus response connections, and on the other hand more complex systems of pointer. The compositionality is thus substituted by the notion of emergence when we found an affective irreducibility of a configuration of the PE.

To combine the two notions, we propose to consider a threshold of irreducibility for the experimental modeling of the LTAM, i.e. when the inter-comparison of the totality of the measured affective relationships of an individual does not allow decomposing the affective relationships into combination of more simple affective relationships.

To conclude regarding compositionality, the internalization of affective properties and the emotional generation seem to be partly driven by compositionality of learned affective properties. However, we should consider a threshold of irreducibility, we will include in the experimental constitution of the EAR (see section 4.3.2.2).

4.3.1.1.5 Using memorized affective relationship to P.E to produce emotion. The emotional memory, denoted under the notion of LTAM (see section 5.1.1.1) as a reunion of stimulus-responses system (both innate and learned) and the [PE patterns]-[potential affective value] relationships system, it is the basis of the E.A.R. which serves to evaluate affectively the P.E. This section aims at providing elements regarding the natural cognition use of this LTAM to produce affective evaluation.

As already mentioned in section 3.1, generation of emotion involves several dimensions. However, it should be taken into accounts that (1) the E.A.R. use while generating emotion could be isolated (e.g. listening to music, alone, without explicit goal). It should also be taken into accounts that (2) the E.A.R. use is a process which is always recruited, even if modulated by other processes (e.g. action planning). It generates emotion in specific situation, as overlapped systems (following the thinking of Scherer, with its three levels used at the same time), or dually processing systems (following

the model of Sander and Koenig).

Firstly, evaluate the environment need to identify elements in our PE, which is done by perceptual and cognitive dynamical processing. We will not here discuss the kind of process that are engaged into perception, low-level cognitive process (as what is done by secondary auditory and visual area) and let-alone high-level cognitive process (as cross-modal and categorical process). However, we summarize this process as an overlapping *patterns, structures* and *features* extraction. The dynamic characteristic of this environment means that such extraction will produce also temporal structure according to short term memory abilities (see proposal of STPB in last section We detail the formalization of the PE in next section as which we were considered as perceived in the EAR formalization.

The stimuli generalization, i.e. the apparent transfer of a learned affective property to one stimulus to a similar stimulus ([Guttman and Kalish, 1956]), is precious information to the design of the use of the LTAM: the calculation of distance of similarities between the LTAM PE contents and the incoming stimuli to evaluate.

Combined to the recurrent emotional processing (see section 5.1.1.1), this leads into a dynamic overlapped evaluation of the PE. We will detail this in the formalization of the use of the LTAM (section 4.3.2.3).While perceived and/or processed by higher cognitive process, the PE is affectively/emotionally evaluated. As this evaluation is made on the basis of the LTAM, affective relationships to PE present in the LTAM are activated on the basis of the incoming detected patterns, structures and features (according to the notion of pointers of [Sander and Koenig, 2002]). According to the unsolved question of emergence vs. compositionality (see section 4.3.1.1.4)all possible activations could be realized. Finally, to fulfill the requirement of [Sander and Koenig, 2002], the use of the EAR needs a Short Term Affective Buffer, which deal with the dynamic and recurrent process of evaluation.

4.3.1.2 Norms for perceptible environment

4.3.1.2.1 Definitions. The proposed notion of perceptible environment (P.E.) is actually not all what an individual can perceive in its environment. We consider several criterions to define what in the environment of an individual could belong to the PE or not, according to the fact that we aim at (1) formalize precisely the contents of this environment (to be able to relate it to affective state of an individual), and (2) manipulate it trough computer (to be able to select, modify or generate it). The reduction of this environment to PE is denoted in last sections and illustrated into Figure 4.7. The P.E. is at the intersection of physical and perceptual/cognitive possibilities established by the psychophysics laws, and the possibilities of any generator (i.e. synthesizer) or analyzer of this environment.

Thus, only several elements of our environment can belong to the PE

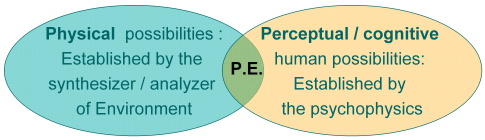

Figure 4.7: The P.E. is at the intersection of physical and perceptual/cognitive possibilities.

(see example with odors in section 1.2): those we are *both able to perceive and able to formalize.*

As the PE is at the intersection of both systems, it could be formalized closer to what is perceived by the individual who experience this environment, or more closer to what has been produced or analyzed by any device.

However, when it is possible, we should focus on the perceived contents, as we aim at relate it to the affective experience of an individual in presence of the contents of this PE. The ideal case of PE is when the formalization is both near to what is synthesized and what is perceived.

Keeping in mind that our aim is to link the PE with affective states, a special consideration should be given to artifact, like music, as it is based on a science of generation of PE which the aim is to elicit affective responses. The interest is that units of music formalization serve to the manipulation of the art elements by the artist, and to describe what is perceived by individuals. When this formalization is explicit (like in occidental written music) and potentially considered as a multimedia content (manipulated trough computer) like computer-based music we are near an ideal PE.

The level of representation of the PE contents is also important both from the analysis/synthesis side and perceptual one.

The analysis and/or synthesis of the PE possibilities of a system are related to the level of representation, and define the possibility of manipulation, from selection to modification and generation. A detailed formalization (like notes events in music) allow the synthesis of the PE while at the opposite a categorical formalization(like musical style which give us poor information regarding the contents of the PE) allow only to select multimedia objects, without possibilities of modulation.

Regarding the perceptual side, a detailed level of representation is closer to the perception. A higher one is closely related to higher cognitive analysis, as semantic categorization. As both could be engaged into the EAR, we should be able to get different levels of representation of the PE. However, the perceptual formalization had been shown to be mainly related to emotional learning (see section 4.3.1.1). Indeed, at the low level we can keep latent cognitive categorization (e.g. the notion of "tiger" is the reunion of perceptual traits, like set of black spots onto a yellow background, with a specific texture and a range of shapes). Also, there are few interests to focus on categorization level, reinforced by the fact categorization descriptors

could not be the basis of a detailed linking between emotion and PE formalization. Finally, structure extractions are a good compromise between low and high level perceptual/cognitive analysis.

4.3.1.2.2 Existing languages and systems. We will now describe several systems which could be used to formalize the PE, regarding the requirements explicated in the last sections.

The main points considered are the formal representation (enabling a precise description of contents) and an easy set up possibility (mainly by findings at working and embeddable multimedia analysis and synthesis systems).

What is perceived by human is still investigated. However years of research combined to computing advances lead into simulated perception systems using formalisms and encoding mechanisms similar to perception. As example, the IPEM Toolbox ([Leman et al., 2001]) is a neural coding based system. This toolbox simulates the neural coding of auditory cells and then extracts musical high level features from this, with respect with the cognitive processing. It is highly interesting for the EAR design, as it allow a computed representation of human perception to cognition process, and thus allow having a multilevel compatible PE representation. In the visual domain, there are numerous works, as the one of [Guyader et al., 2002] which consist of the embedding of human perceptual principles as criterion for visual classification. At a more perceptual level, SpikeNet ([Thorpe et al., 2000]) is a powerful simulation of the perception, applied to image retrieval.

Classifications systems based on phenomenological description are rather high level formalization. They could give interesting data as they are related to what we consciously perceive, but still requires automatisms of extraction. Main formalisms are the Music Genome Project (http://www.pandora.com/) in which experts manually annotates a wide sets of songs with 400 variables as selection criterion, and the more low level Temporal Semiotic Units ("Unités Sémiotique Temporelles " [Delalande et al., 1996]) based on previous sound object formalism of [Schaeffer, 1977], allowing a description in terms of e.g. ascending, breaking sound, etc...

However, as if these perceptual-based systems are close to what an individual perceive, they only allow the analysis of the PE contents, and thus the selection of contents. The human artistic practice also led into interesting formalization of the PE. Music midi score based formalization (see figure 4.8), along with analysis and rendering engines, allow to extracts melodic line from a recorded performance and outputs a midi score ([Marolt, 2004]), or generate musical outputs from a score. MusicXML [Good, 2002]) is a music representation grammar which is based on the extensible, flexible and platform independent XML language. Finally, MPEG-7 is a powerful specification and set of tools based on XML language, which allow from perceptual

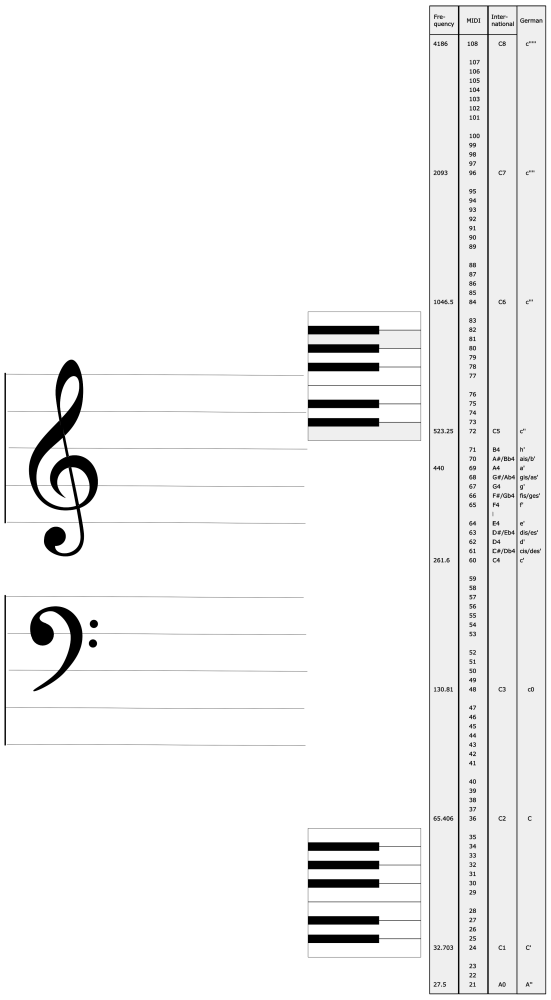

Figure 4.8: Different norms for the notes in midi and musical representation.

to semantic analysis of multimedia content (video, sound). For example, in the audio domain: "MPEG-7 low level descriptors are useful for semantically low level search and retrieval applications as music identification, music similarity or genre recognition" (Motion Picture Expert Group page, http://www.chiariglione.org/mpeg/). Indeed, 'Description coding' scheme possibility open to embedding of any indexing of multimedia, on the basis of MPEG 7. Despite the MPEG-7 is not a synthesis language, it is the more unified form of analysis.

All these representations of the PE are interesting candidates to be potentially included into the EAR, as the information we have regarding the perceptual part of the emotional memory is perceptive/cognitive representation (see section 4.3.1.1.1, 4.3.1.1.3).To be able to embed all this representation into the formalization of the PE as an unified, we should made an appropriate embedding system. The main direction we currently takes are to make parallel representations of the same environment, or translate non-XML representation into XML one, and then merge the different underlying XSD schema into a global one. For first steps of the model we will develop a simple XML based language, used as a test framework of environment representation (see section 4.5).

4.3.2 Formalization of the E.A.R. in natural cognition: synthesis from the literature and hypothesis

According to [Villon, 2003], the approach taken by the EAR modeling, is to (1) make an accurate model from cognitive science theories, and (2) fill this model with data leaned from each user (user-modeling). This section aims at establish a formal description of the E.A.R. in natural cognition according to the literature synthesis (see previous sections) regarding affective learning (EAR initialization and updates) and emotion generation (EAR use). This drives the E.A.R.Multimedia model design.

We proposed previously that the affective experiences each individuals feel and/or express while experiencing the 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 E.A.R. The E.A.R. is the human functional (simplified) subsystem we model to accounts for our ability to associate affective experience with media, and thus learn and use theses associations. The contents of the E.A.R. will be different at the inter-individual level, due to the respect of inter-individual level (see section 3.3). Moreover, the contents of the EAR of an individual will change over time according to his new affective experience with the PE.

The main hypothesis of the E.A.R. is that is possible to model the individual's dynamic relationship between affective state and perceptible environ-

ment from a model of previous affective experiences' resultant: the E.A.R. of individuals ([Villon, 2002] [Villon, 2003]). We are aware of the fact that the EAR could be filtered by other process (section 3 and figure 3.1), but we emphasis on the fact that such EAR is engaged in every emotional evaluation, and thus could be modelled as an entity.

As we seen in previous section, producing affective experience to media could be considered as building in real time an affective experience using associations of affective perceptive/cognitive properties of environment with affective properties. Thus, the E.A.R. model will describe the association of, on one hand, some perceptive/cognitive properties of the environment with, on the other hand, affective properties. The E.A.R. is defined as a set of data structures and processes hosting an affective associative memory resulting of affective learning with P.E. It represents a synthesized and stored result of affective experience with P.E. into memory. The main hypothesis of the E.A.R model is that we use our E.A.R. each time we produce an affective experience which is elicited by the perceptible environment.

4.3.2.1 E.A.R. contents

The content of the EAR are the Long Term Affective/Associative Memory (LTAM); the Short Time Perceptual Buffer (STAB), and the Short Time Affective Buffer (STAB). These three structures comes from the reduction of, and are compatible with, the Sander and Koenig model (see section 4.3.1.1.1).LTAM. The LTAM is the emotional memory (in the Ledoux acception, see 4.3.1.1.3).

As proposed into the section 4.3.1.1.3, the LTAM is made of the fusion of the stimulus-responses subsystem of Sander and Koenig, with the [PE patterns]-[potential affective value] relationships system, reduced from the same model.

Why it is needed to make such fusion ? The main difference between the stimulus-response and the [PE patterns]-[potential affective value] relationships system is that the responses in the first system are directly expressed as responses ("processing reflexes"), while the potential affective values are concurrent and multiples, and need to be pre-processed before to become actual responses. The second main difference is that the stimulus-responses connections (amygdala) are made of evolutionary salient stimuli, or learned stimuli. For learned stimuli-responses, they are created on the direct basis of evolutionary stimuli (first-order conditioning), and as stated in section 4.3.1.1.3, on the form PE representation; phylogenetic stimulus representation; emotional response. At the opposite, the [PE patterns]-[potential affective value] relationships system could host relationships learned from previous learning (i.e. high-order conditioning). Making such fusion implies that starting from the two forms of relationship established in the section 4.3.1.1.3, we make a shortcut to the form of emotional memory for ontoge-

netic learned stimuli, on the basis of the evolutionary salient stimuli, as a direct linking between the general PE representation associated to emotional response. This shortcut is made because we don't know, from a practical measure point of view, and considering that we have no tracks of the past experiences of the individual, how to differentiate the relationship resulting on a first-order conditioning, or on a high-order one.

Thus, combining theses two systems into the LTAM, we can describe the whole LTAM as a [PE patterns]-[potential affective value] relationships system within which patterns could be entire stimuli, and the potential affective value could be actual responses.

We thus include the stimulus response system within the LTAM: in the case of the PE pattern is an entire stimulus, and the potential affective value an actual (direct) response, we are in the case of a stimulus response. Moreover, we keep the notion of affective potential, or internal emotional state (see 4.3.1.1.3) as a means to accounts for mixture of potential leading into a response. This is a compromise between (1) the inability to differentiate phylogenetic responses, what an individual learned from phylogenetic response, or learned from previous ontogenetic learning, (2) the requirement of accuracy according to the human architecture of our approach (we merge two systems but keep their functionality), (3) the need of practically measuring the contents of the LTAM of an individual. Thus, we host in the same manner phylogenetic and ontogenetic contents, as we can't precisely separate the phylogenesis and the ontogenesis of an individual while we measure the EAR.

Finally, emphasing on the form of the LTAM as a [PE patterns]-[potential affective value] relationships system, we emphasis the fact that the nature of the E.A.R is not only a map data structure containing stimuli on one hand, and affective response on the other hand, like the stimulus-response system. This relates to the aim of the EAR to model which is to model how we can generate dynamic emotion for new complex stimuli, made of subtle variations (e.g. in music).

Regarding phylogenetical and ontogenetical involvements into PE affective evaluation, such LTAM form is coherent with a common involvement of phylogenetical and learned affective relationships (which mainly makes the inter-individual differences in emotional evaluation of same pieces of music).

Thus, we define the Long Term Affective Memory as a data structure which contains relationship on the form of perceptive and affective pairs (the pointers), with different degrees of complexity within each part of the pairs. Each perceptive-affective pair has a strength value, due to reinforcement (see section 4.3.1.1.2).

According to the type of memorized associations which were experimentally demonstrated (see section 4.3.1.1.3),we propose to formalize here what could be stored into the E.A.R.

All theses opposed categories could be mixed. So an affective pair could

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

be on the form multimodal complex dynamic pattern subpart of PE; delayed direct discrete affective property The notion of complex pattern combined to the relative category constitutes the notion of structure, which thus could be embedded in the subpart of the PE (we do not explicitly place it in the table as it is a combination of properties).

Once we defined the atom of the E.A.R. to PE contained into the LTAM as such perceptive-affective pairs made of any subpart of the P.E. with any affective property, we should consider three main notions from review of last sections. (1) The emotional memory is the result of a synthetic process, a convolution of incoming data with previous data stored (see section 4.3.1.1.3). (2) It was shown that affective learned response could occur on low-level feature like color, as well on complex pattern (as a painting) (see section 4.3.1.1.3). (3) Emotional generation could be partly driven by compositionality (see section 4.3.1.1.4).

Taking together theses statements, we can hypothesis 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 (e.g. convolution) according to the previous contents (following the natural cognition principle of section 4.3.1.1.3).This means that the perceptive-affective pairs should be hosted in a structure presenting a high interdependency of the items it contains.

We propose that such perceptive-affective pairs are organized into the

LTAM as a network, to support the notion of pointers proposed by Sander and Koenig, and following the notion of memory usually considered as a network, not a directory.

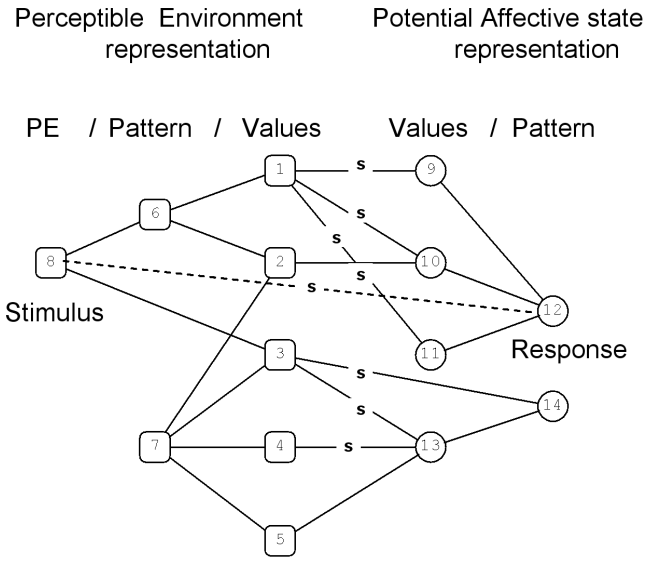

Figure 4.9: The proposed representation of the Long Term Affective Memory.

The proposed formalization of the LTAM is thus 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:

$$G_n(t) = ((PV_n, PE_n), (AV_n, AE_n), RE_n) \tag{4.1}$$

PVn=1, 2,3,4,5,6,7,8 (Perceptual Vertices) and AVn=9,10,11,12,13,14 (Affective Vertices) are respectively the sets of vertices standing for the representation of the PE, and the potential affective state representation. 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 necessary, it could be a stimulus-response. PEn=(1,6) ; (2,6) ; ... ; (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 will be at the middle of the graph (e.g. (1,9)), while a pair of type whole PE, affective response will link extremity of the graph (e.g. (8,12)). Such extreme pair hosts thus also the stimulus-responses. The links could be made between 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) corresponding to the probability of occurrence, and the possibility of extinction. This value of strength is the result of the EAR initialisation and updates.

The LTAM could thus hold the fact, for example, that a feature as a specific color embedded into specific patterns could exhibit several different 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. This LTAM support both the notion of compositionality and emergence (see section 4.3.1.1.4)of emotional memory, by defining the REn.

We are aware of the fact that such model is less precise than existing computational model of emotional learning (e.g. [Balkenius and Morén, 1998a]). However, the general requirements of a functional architecture of emotion are fulfilled (see 4.3.1.1.1), and this modeling is a compromise between cognitive modeling accuracy, need of implementation, and descriptive status of the model.

Moreover, it could embed several levels (from whole PE, patterns to values) and types (absolute, relative values, etc...) of representation, both at perceptual and affective one, which is mainly interesting as we don't know precisely what an individual extract in its PE, and to which level of affective property it create perceptive-affective pairs.

However, the affective part of an affective pair is not always directly a response. When it is no the case, we have to combine the different potential affective state outputs, which is done, among other tasks, by the STAB. We will now see the short term memory components, required by the reduction of the Sander and Koenig model.

STAB. The Short Term Affective Buffer is what was the Short term memory in [Villon, 2003]. The STAB keep tracks of the affective evaluation of the PE. It is divided into two buffers: the buffer of 'felt' and the buffer of 'prediction', which are separated by the 'now'. The STAB has two roles: merge different proposed response from the LTAM, and add the current proposed response from the LTAM, to the ongoing affective experiences of an

individual. For example, the current ongoing pattern of affective valence of an individual is augmenting. Then two specific patterns are identified in the PE which are related respectively to (a) a descending valence during 3 seconds, and (b) a neutral valence during the moment of presentation of the second PE pattern. The STAB will merge the outputs of the LTAM, and then merge this result to the present buffer of prediction. Then, information of the buffer of prediction will be moved to the buffer of felt, according to time. The STAB also extracts pattern and dynamics (see table 1) from experienced experiences, from the buffer of felt.
STPB. The Short Term Perceptual Buffer is what the perceptual and simple cognitive processes extract from the PE. It deals with any modal or multimodal contents of the environment. It is a dynamic element of the EAR as it should be able to extract temporal structure in the incoming PE. This stands for the exteroceptive processing in Sander and Koenig model (see section 4.3.1.1.1).

4.3.2.2 E.A.R initialization and updates

We provide here a simple overview of the mechanism of EAR intitialization and update according to the emotional learning synthesis of previous sections, and without taking care of several other factor (e.g. sensory preconditioning, habituation, dishabituation and sensitization which are not necessary for the formalization in artificial cognition).

We explain here how the phylogenesis and ontongenesis build the E.A.R. which is used to evaluate affectively the perceptible environment. Given that the E.A.R. is used each time we produce an affective experience which is directly elicited by the perceptible environment, this presupposes that such E.A.R. is built before we produce any affective experience, elicited by experiencing the P.E.

The Figure 4.10 illustrates the initialization and update phases of the LTAM. The first initialization of the E.A.R. is done by the phylogenesis (1). This is implemented as a set of predefined stimuli-responses. Then, the E.A.R. is updated according to individual experiences, during ontogenesis process, with cultural influences and personal experiences.

The process of update is done with three components: an affective situation generates an affective experience (2) which is analyzed by the STAB. The individual is experiencing some elements of the PE, which are processed by the STPB. Then, an association is made between the outputs of STAB, and STPB, and thus between PE formalization, and affective experience formalization (3). Under specific condition, this new association is encoded and synthesized into memory. The update of the LTAM of the E.A.R. (4) is made by a convolution of the incoming associations with the previous content of the E.A.R. (synthesis). During this process, the content of the P.E. and its affective properties of the incoming association will be analyzed and compared

to the current E.A.R. affective properties, in order to (a) encode the new association consistently with the current E.A.R, and (b) update the existing perceptive-affective pairs. Update process justifies the notion of embodied relationship, as it is an actual ontogenetic process, dealing consistently with previous learning to embed new ones.

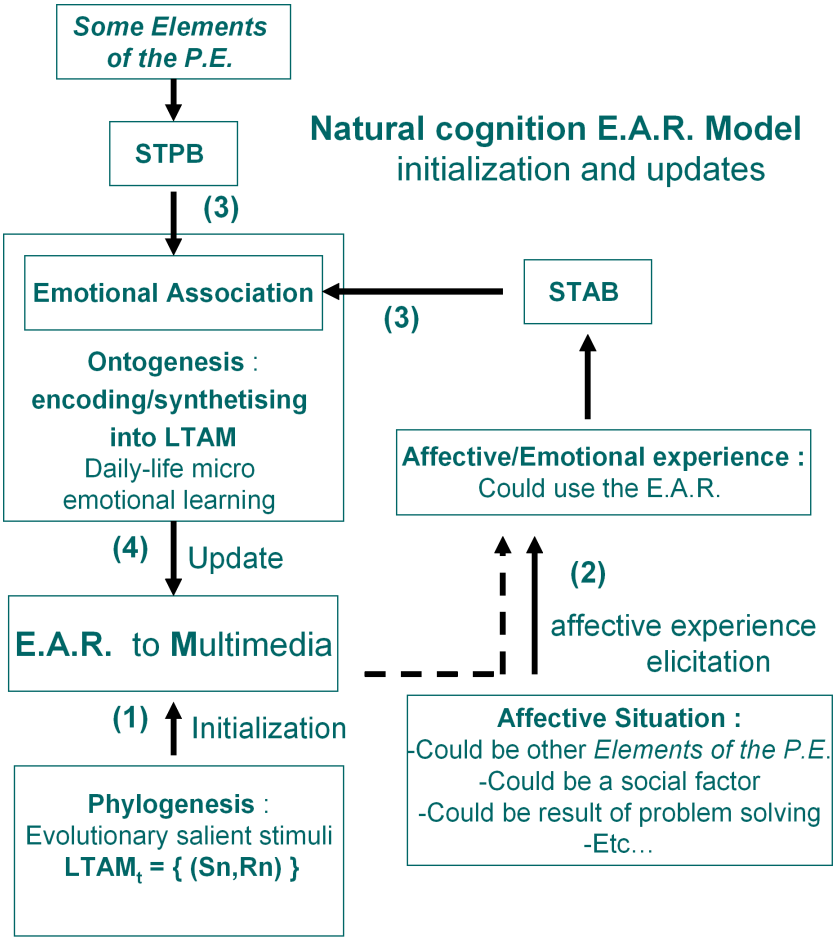

Figure 4.10: Natural cognition E.A.R. Model. : initialization and updates

The affective experience elicitor could be elements of the PE or other emotional situation (see 3.1). In the case of an elicitor considered as elements of the PE, it is the EAR which is engaged to produce the affective experience (we will detail this in the next section). In this case, such as when we listen to new music, this is the novel contents of the PE which are used in the update process, on the basis of already experienced elements of the PE.

Example of learning. Let's consider the following situation : an individual is listening to a song (s)he know well, and really appreciate. Actually, the song is not exactly the well-know song, but a new version with a melodic

line with an unusual timbre (e.g. an electric guitar instead of an oboe). This new song represents the P.E. Some elements of this P.E. (the well known song) constitute the affective situation which elicits an affective experience. If appropriate reinforcement (like listening to this song several times), and according to the affective properties of the other elements of the P.E. (the new timbre), E.A.R. will be updated. (1) if the new timbre (alone) is not already related to any affective property (does not sound like any known sound by this individual), or associated to near-neutral affective properties, the update will associate affective properties of the song to this timbre. (2) If the new timbre (alone) is already related to any affective property, or associated to near-neutral affective properties, the update will associate affective properties of the song to this timbre, and also associate affective properties of the timbre to the song.

4.3.2.3 E.A.R use

The E.A.R. utilization to produce an affective evaluation of the environment is based on long term previous affective experiences with elements of the PE, and previous recent affective state of the individual. The affective evaluation produced by the E.A.R. is not related to individual's goals or social current situation. However, the evaluation of the P.E. produced by the E.A.R. could be modulated by the individual's goals and social current context (see section 3.1). It is less filtered into the situations where the individual are rather passive or engaged into the process of selecting/modifying/generating the P.E. (e.g. an artistic creative process).

The affective situation elicits an affective experience using E.A.R only if the affective situation involves elements of the P.E. as elicitors. The use will be realized by the extraction of elements in the PE, using the STPB which continuously formalize the incoming PE (see table in the EAR contents). Then, theses elements are compared to all the stored elements of the LTAM.

A distance calculation is performed between STPB output and PVn of the LTAM, using a threshold (related to the degree of stimulus generalization, see section 4.3.1.1.5). If direct relationships are found among REn, i.e. if REn contains elements of the PVn which are similar to the outputs of the STPB, affective potential (the AVn contained into REn) are activated. Then, all theses activations are sent to the STAB, which merge the LTAM outputs and the buffer of prediction of the STPB (see STAB description in 4.3.2.1), to make the profile of the evolution of the affective state (which is relative to the present one). The STPB performs a recurrent continuous emotional processing (see section 5.1.1.1) which leads into a dynamic overlapped evaluation of the PE.

Example of use. The sound of the opening of a projection screen is rated as low arousal, and neutral valence by an individual, according to its personal and cultural background (opening a projection screen is for this

person a neutral situation). However, the sound of a female screaming is rated as high arousal, and negative valence, according to the phylogenesis and its personal and cultural background (through films, reportage, etc. . .). This individual is confronted to a P.E. of screen projection opening. During the first 2 seconds, it perceives sounds corresponding to a perceptual pattern of a female screaming (not conscious cognitive categorization), producing the associated affective reaction, then a pattern of projection screen is recognized and the neutral affective state is elicited. Note that the pattern identification is unconscious. At this stage the subject used its E.A.R. to produce an emotion.

Then, or with overlapping, other processes responsible for emotion generation can be realized, like those engaged into the Scherer's deliberative and cognitive level.

We described the primary use of the E.A.R. which is the affective evaluation the P.E. To summarize, the output produced by the evaluation of the P.E. on the basis of the E.A.R. is automatic and passive (see section 3.2). It is also is engaged into active process as the creative one. Then, this E.A.R. output could be modulated by other components, depending of the situation (see Figure 3.1, section 3.1).

We do not take position regarding the conceptual (an interdependence of different systems of emotional generation, like the Scherer's sensory, schematic and conceptual parallel levels of processing) or effective (an actual E.A.R. output is produced, i.e. exists as a primary internal output in the brain, then is modulated by other brain structures and/or processes) nature of the modulation of the E.A.R., as it is out of the scope of this thesis.

4.3.3 Formalization of the E.A.R. for artificial cognition toward implementation

After having presented the EAR in natural cognition, we will now present the artificial cognition side of the same model, i.e. how to use it as a simulation of the evaluation of the environment, and thus applications regarding selection and design of the PE contents.

4.3.3.1 E.A.R. learning algorithm

The E.A.R. learning algorithm is based on, but is different compared to the natural cognition "E.A.R initialization and updates". Learning should be understand in the sense of machine learning, from computer point of view : how to build the E.A.R. of a subject ?

Actually, the main differences lie in the update. In natural cognition, a new experience modifies the E.A.R. Here we want to extract 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 onto a

subject should be considered as using an already present base into the E.A.R of the subject and not as a new incoming experiences.

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 emulate the role of the STAB, which could be done by computation in the valence*arousal space. Then we should also emulate the role of the STPB. This is already done by the fact that the perceptive/cognitive representation already 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. This could be achieved using highly emotional stimuli, but respecting the notion of inter-individual difference, and so not selecting only evolutionary relevant stimuli. At the opposite we should try to find high pleasure and displeasure for the individual (which could be not emotional at all for an other person). The best is to find a high variability of emotional response for same stimuli (inter-individual differences) between subjects, but a low variability of emotional response for the same stimuli, for the same subject, using repeated measure on the subject (intra-individual consistency). Such consistency could be assessed using the standard deviation of a group of repeated measurements over the same individual. The less the standard deviation is, the more the consistency is.

Once we get consistent aesthesis result from an individual, we should get an actual (consistency) and useful (detailled relationships REn to then be able to manipulate the PE) LTAM. The idea is to be able to start from a measure, and then get what serve to this subject to produce 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 (see next section).

The algorithm is on the following form (commentaries are placed after "//"). Not all the rules are provided herebut give an overview of the learning mechanism :

The rules to decrease and increase the weights are still at the design phase. They should come from the number of times 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 stops 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 [Villon, 2003].

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, but converting it as something embodied, to be able to then make simulation

```
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
```

ALG 1: Form of the EAR learning algorithm.

(see next section) and PE manipulation.

4.3.3.2 E.A.R. recognition algorithm

How to use the E.A.R to simulate the affective response the user would have been produced in presence of specific media? The recognition algorithm is similar to the method presented into the use in natural cognition (see section 5.2.3).

4.3.3.3 Measure and computer based use of the E.A.R.

As we saw previously, the EAR of a user is valid until an important update. Thus, we should be aware of the implication it has on the robustness in time of a measured EAR. A LTAM could vary if a subject has new experience with its environment. Thus if we have the LTAM of a user, we should then use it quite quickly or be sure that this user didn't made modification of its LTAM from the moment we measure it, and the moment we want to use it, as a user model. For example we measure today the LTAM of a user and find that the color red with a specific sound is associated with a high valence. Tomorrow, the same user experiences a very low valence, due to a specific event, in presence of the red color with the same sound. The emotional learning theory tells us that the red color will get a new value (here a low valence), under specific conditions (reinforcement, etc...). The day after, if we want to use the LTAM of this subject, and especially using the red color associated to high valence, it will produce an unexpected emotion in the user.

So, there are two potential solutions. (1) We measure precisely the new emotional experiences of the user with the multimedia environment, from the moment we firstly measured the EAR of this subject. This is quite impossible, as it requires (a) to be able to record all the multimedia contents associated with emotional experience, from the moment we firstly measured the EAR, (b) to be able to formalize this environment and (c) to be able to infer the storage of emotional properties of the perceived environment on the basis of the associated emotional measure - like conditioning theory, and reinforcement problematic.. (2) We control the actual expected user emotion and update the LTAM accordingly if changes are found. This is a simple and accurate solution. We could also measure specific elements of the EAR of the subject at each new session, with a rapid protocol. In this case, for example with images, it is possible to expose subject to subliminal elements of the PE while measuring (subliminal) emotion, with rapid reactions (e.g. physiological signals).

4.4 Computational Architecture

4.4.1 Overview

The currently implemented architecture of the EAR model is presented in this section at the computational and implemented level, according to the specification of the section 4.3 and the possible uses in HCI presented in the section 4.1.5.

The goal of this formalization and implementation is to provide a set of processes and components to build the Long Term Affective Memory of an individual, using a computer.

The main components of the formalized computational architecture are :

- ≻ Multimedia Contents
  - **PerceptibleEnvironment** : the main handler of MMItem. It builds, register MMItems.
  - **MMItem** : handle both multimedia contents and contents description
  - **Possibles** : registers all possibles values of the PerceptibleEnvironment
  - **Formalizer** : converts any external langage into a VariableAndValues representation (e.g. a midi representation)
  - **Renderer** : converts any VariableAndValues representation into an external langage to be rendered by a device.
- ≻ Affective and Emotionnal classification of media contents
  - **EmotionRepresentation** : stands for the abstract representation of affective state and emotion
  - **Aesthesis** : A set of processes and datastructures to measure affective state and emotion felt by individual for each MMItem
  - **AesthesisResults** : a datastructure hosting the results of the aesthesis
  - **LongTermAffectiveMemory** : the above-mentionned structure which can be update on the basis of incoming AesthesisResults
  - **ShortTermPerceptualBuffer** : handle percepts dynamic for LTAM
  - **ShortTermAffectiveBuffer** : handle affective information dynamic for LTAM[1]

[1] The short term memory components are not fully formalized nor implemented. There will not be described in detail here as several method were simplified and used directly within the LongTermAffectiveMemory (i.e. stimulus decomposition)

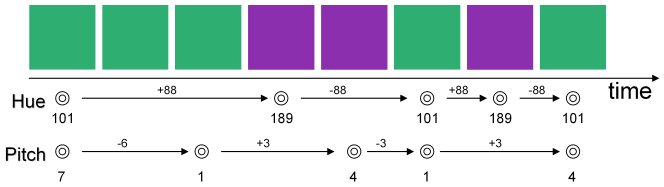

Figure 4.11: An example of Perceptible Environment representation, using Hue (viual) and Pitch(audio)

- Common components for multimedia formalization and affective classification
  - **VariableAndValues** : an element for content description
  - **Event** : a sub element for content description, including timing information.
  - **Value** : a sub element for content description. It can be embeded into an Event component.

The components are mainly organized according to three groups : perceptible environment handling, aesthesis (i.e. measure of affective information associated to multimedia contents) and the long term affective memory.

4.4.2 Perceptible Environment : multimedia content handling

We built an xml norm and a set of associated methods and object to handle the perceptible environment. As specified in the section 4.3 and in the table 4.3 the aim is to be able to handle different type of contents, and then be able to associate it to emotion representation.

We designed a set of components to be able to handle content at the analysis and synthesis levels. Different possibilites of importation were achieved : with associated multimedia file which can be played, or with active renderer which can render the content of the media of the description. To be able to import any type of content, a mechanism of renderer interface to implement was achieved. Any new kind of media content could be wrote by following the specfication of the renderer.

The perceptible environment is formalized as a list of absolute and relative values of percepts which comes from an analysis or a synthesis of the multimedia contents. For instance the figure 4.11 present an example of perceptible environment made of Hue and Pitch values.

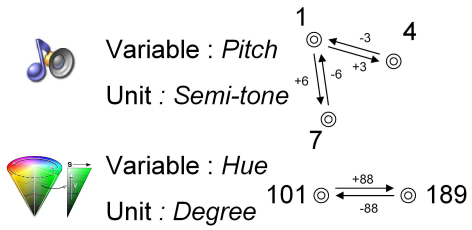

Figure 4.12: An example of constraints of the perceptible environment. The possibles percepts of the multimedia contents are defined as a set of related absolute and relative values, for each variable.

4.4.3 Possibles

The **Possibles** component registers all possibles values of the PerceptibleEnvironment. It is organized as a set of related absolute and relative values organized into variables. For instance, the Possible component associated to the perceptible environment in the figure 4.11 is schematically presented in the figure 4.12. In this example, we consider that the perceptible environment is made of two variables (pitch and hue) and that those variables have a set of constraints regarding their absolute and relative values (represented by a graph of absolute and relative values in the figure). Formally, the component is associated to an xml representation, as presented in the figure 4.13 The xml specification defines a set of variable and associated absolute values.

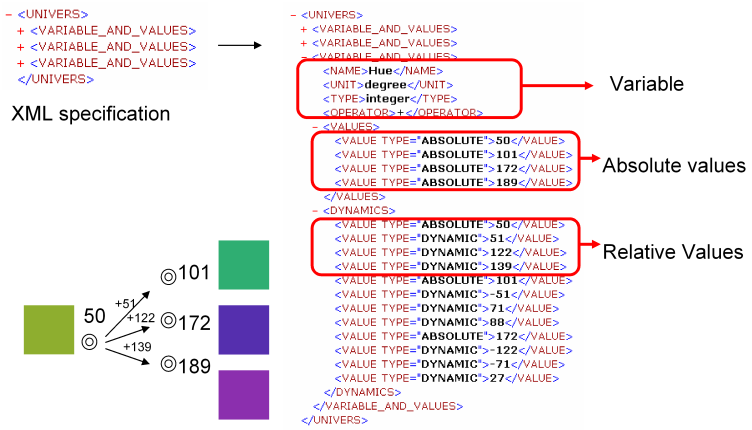

Figure 4.13: An example of xml representation of the Possibles components.

For each value, a set of relative values allowed are specified.

```
<MMItem Id='1'>
  <FileName>0008.jpg</FileName>
  <MultimediaType>visual</MultimediaType>
  <TimeUnit>second</TimeUnit>
  <Duration>5.0</Duration>
  <TimeStart>0.0</TimeStart>
  <TimeEnd>5.0</TimeEnd>

  <VARIABLE_AND_VALUES>...</VARIABLE_AND_VALUES>
  <VARIABLE_AND_VALUES>...</VARIABLE_AND_VALUES>
</MMItem>
```

Figure 4.14: XML representation of a Multimedia Item (MMItem)

4.4.4 Multimedia Item : MMItem

As discussed in the table 4.3, affective association could be performed by human on a variety of elements of the perceptible environment. Being able to manipulate media contents in a systematic way according to affective information thus need to have a computer based representation of the contents. This representation should allow any kind of representation specified in table 4.3. While it exists several representations according to different modalities (e.g. musicXML for music, see section 4.3) we needed a simple and multimedia norm to handle any kind of contents, regardless the modality of the content. A multimedia item can be made of any media content. It is formalized by a set of tags. The form of the header is displayed in the figure 4.14. The tags allow to identify the MMItem. The four last tags (TimeUnit, Duration, TimeStart and TimeEnd) are optional : they can be omitted or empty. The header is then followed by a set of VariableAndValues groups of tag.

4.4.5 VariableAndValues

Each VariableAndValues is made of a set of tag, as denoted in the figure 4.15. The VariableAndValues serves both for the possible representation and for the MMItem representation. In the case of Possibles components, VariableAndValues contains only Values which could be either absolute or relative. In the case of MMItem component, VariableAndValues contains also absolute and relatives values, but contains Event components. The Event component embed Values with timing information. A list of events can thus represent a pattern. An optional parameter is the type of the group of Events denoted in <EVENTS TYPE='...'>. If the type is 'analysis' it means that the system will have no mean to use the primitives described as a synthesis for creating a new MMItem (it could be the case for instance with MPEG-7 for audio when a vector of value is extracted from the audio but

```
<VARIABLE_AND_VALUES>
  <NAME>Notes</NAME>
  <MULTIMEDIA_TYPE>audio</MULTIMEDIA_TYPE>
  <UNIT>semi-tone</UNIT>
  <TYPE>integer</TYPE>
  <OPERATOR/>
  <EVENTS TYPE='synthesis'>
   <DEVICE>MIDI</DEVICE>
   <TimeUnit>tick</TimeUnit>
   <EVENT>
    <VALUE TYPE='ABSOLUTE'>30</VALUE>
    <TimeStart>0</TimeStart>
    <TimeEnd>5</TimeEnd>
   </EVENT>
   <EVENT>
    <VALUE TYPE='ABSOLUTE'>38</VALUE>
    <TimeStart>5</TimeStart>
    <TimeEnd>10</TimeEnd>
   </EVENT>
  </EVENTS>
  <EVENTS TYPE='analysis'>
  <DEVICE>MIDI</DEVICE>
   <EVENT>
    <VALUE TYPE='RELATIVE'>8</VALUE>
    <TimeStart>50</TimeStart>
    <TimeEnd>5</TimeEnd>
   </EVENT>
  </EVENTS>
 </VARIABLE_AND_VALUES>
```

Figure 4.15: Example of XML representation of Variable And Values within a MMItem

there are no system to generate an audio from the MPEG-7 vector of such values). At the opposite if the type is 'synthesis' it means that the system is able to then reuse the values of the formalization to create new MMItems (i.e. the values in VariableAndValues are extracted and sent to the Possible component).

The VariableAndValues are then used in the LongTermAffectiveMemory component to be associated to EmotionRepresentation components.

4.4.6 EmotionRepresentation

The EmotionRepresentation component stands for the abstract representation of affective state and emotion. It is based on both a dimensional representation (i.e. a coordinate in the space made of valence and arousal, and even dominance, see section 3.4) and a discrete representation (e.g. joy, fear, etc...). Moreover a qualtitative affect representation is possible by partionning the space of valence and arousal in 5 regions (i.e. high/low arousal, high/low valence, neutral). The different representation can be converted (e.g. converting a valence and arousal value into a discrete emotion). Moreover dynamic representation of emotion (e.g. changing from a coordinate to other) can be represented within the component.

4.4.7 Aesthesis

An aesthesis session let a user associate **MMItems** with **EmotionRepresentations**. This is performed using any interface involved in the measure

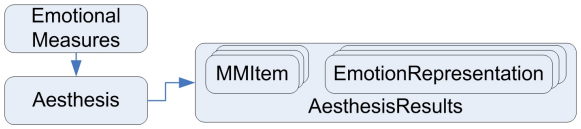

Figure 4.16: Emotional evaluation of MMItem by an individual : Aesthesis.

of emotion with multimedia contents. It thus allow to measure affective evaluation of multimedia contents by an individual. The aesthesis does not specify the method to measure the affective state. It could be an indirect measure (e.g. using physiological measure, see section 5) or a direct measure i.e. by directly asking user).

However we designed a specific software to let user classify MMItems according to the felt affective state, described in other sections (see section 6.3 and 5.3) which is an example of possible measure. Once completed, an **AesthesisResult** serves as a basis for the initialization and update of the LongTermAffectiveMemory component.

4.4.8 Long Term Affective Memory

This component follow the guidelines of the section 4.3. Therefore, its main datastructure is a graph made of a perceptual graph and and affective graph. The perceptual graph is made of VariableAndValues, Event, Value components. The affective graph is made of EmotionRepresentation component. The figure 4.17 presents schematically the relationship between the actual ltam of the user and the measured ltam (i.e. the model of the ltam we build in computer from the aesthesis sessions). The upper part schematize the LTAM of the user. This LTAM may evolve after a specific amount of time and if new affective experience occurs (as specified in section 4.3). In parallel the bottom part of the figure present the successive aesthesis which allow to create a ltam for the considered perceptible environment. The first aesthesis intialize the ltam model. Then a second aesthesis (i.e. with different perceptual content) allow to update the ltam model for a same ltam of the user. Later, another aesthesis may update the ltam model to follow the content of the user's ltam. The intialization and update of LTAM are practically explained in the section 4.5.

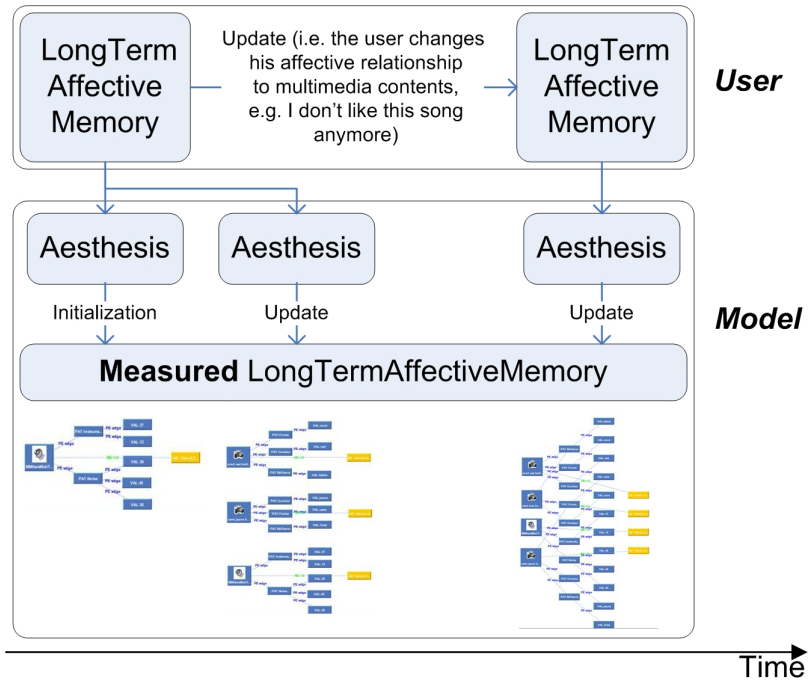

Figure 4.17: A scenario of ltam update. The LTAM of the user might evolve in time, and the model of LTAM could be updated with an unchanged LTAM of the user (i.e. the user still have the same affective relationship to multimedia contents after a time period) or after an update from the user (i.e. the user changed his relationship to some contents).

4.5 Implementation and testing

We present in this section the implementation following the specification of the EAR model as Java objects. We especially applied the algorithm presented in the specification to build the LTAM. The implementation uses Java 1.6.0 (including javax.* packages for graphical interface and external device connections) for core implementation, org.w3c.dom.* and org.xml.sax.* packages for xml parsing and Java Media Framework (JMF) packages for multimedia support. It is also based on jGrapht (org.jgrapht.*) for graph mathematical structure handling and jGraph (org.jgraph.*) for graphic adapters. Each component presented previously had been implemented as a java class.

We discuss in this section how each object could be used and how we tested those components with external devices. A software-oriented description is provided in the form of a Java Application Programming Interface (API, to allow researchers and developpers to test the proposed approach) in section 6.

4.5.1 MMitem creation

Several methods were implemented to build the MMItems. In any case, a MMItem should contain both media contents to be played or rendered and a description of this contents in terms of VariableAndValues component.

4.5.1.1 Loading stimuli-like files

We can load media files expected to be listened, viewed, etc... by the user to build the MMItem. Different methods are possibles :

- **Loading and xml file standing for the formalization and a media file.** In this case the xml file contains only the formalization of the content of the associated media file. The media file contains already the contents which will be experienced by any user. This is presented in the figure 4.18.

- **Loading and xml file standing for formalization associated to a renderer**. In this case the xml file contains formalization of some contents and a special tag "DEVICE". The name in the tag will be then used to retrieve the associated renderer in the software (e.g. a MIDI -Musical Instrument Digital Interface- renderer). The renderer can convert the content of the formalization into media contents which will be experienced by the user (e.g. sound)

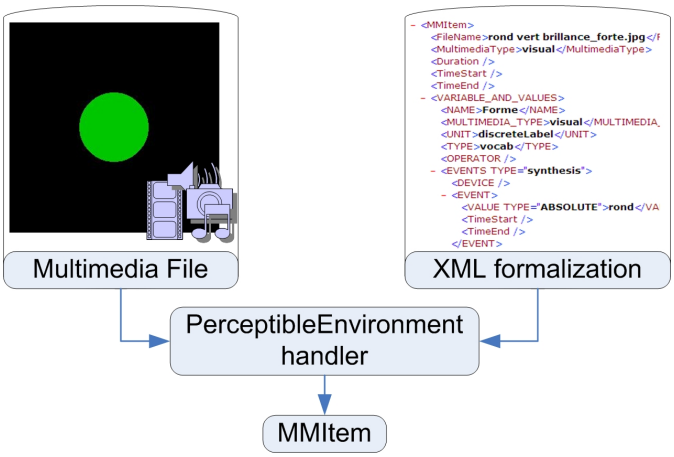

Figure 4.18: Creating a MMItem from a Multimedia file and a formalization file in xml langage.

4.5.1.2 Loading primitives to automatically build random stimuli

In this case, we load a file which will fill the contents of the Possibles component. Then MMItem can only be created randomly from the set of variables and values stored in the Possible component.

Additional developpement of this approach could be to embed rules of MMItem generation, e.g. music composition rules, rules for transition between colors, etc... but this is out of the scope of this thesis. The figure 4.19 presents the construction of MMItems from the Possible component. In the figure, the source of the Possible is a XML file specifiying the different VariablesAndValues. However, the Possibles component could be built using already loaded MMItem, i.e. by extracting all the existing values in the loaded MMItems.

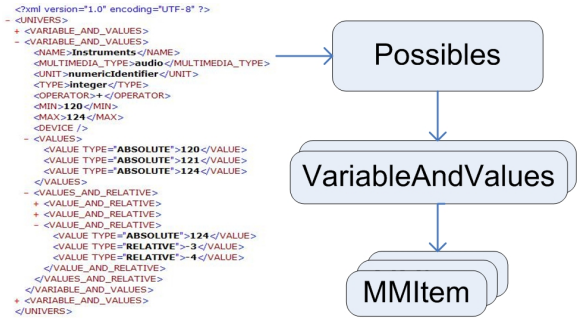

Figure 4.19: MMItem can be built from a xml specification file. The file first build the **Possibles** component. The possibles component then uses the creted VariablesAndValues to build MMItems.

4.5.2 Formalizers and Renderers

To allow more interesting kind of multimedia contents that the examples presented with music and colors, a flexible system of Formalizer and Renderer interface which can easily be extended to other external device have been designed. For demonstration purposes, the only protocol implemented for external device is the MIDI one.

Formalizer and Renderer allow to command any external devices, both in input and output by converting the VariableAndValues format into any external device format. The interest is to (1) be able to control any external device for playing purposes. For instance tests had been made with an external piano which had the ability to play with an external control (Yamaha Disklavier). The second interest (2) is to be able to sense any use of the external device to then formalize the content and then be able to build a "Possibles" object from it (as the Possibles object register all the possibles actions to render an environment).

The system of formalizer and renderer is presented in the figure 4.20.

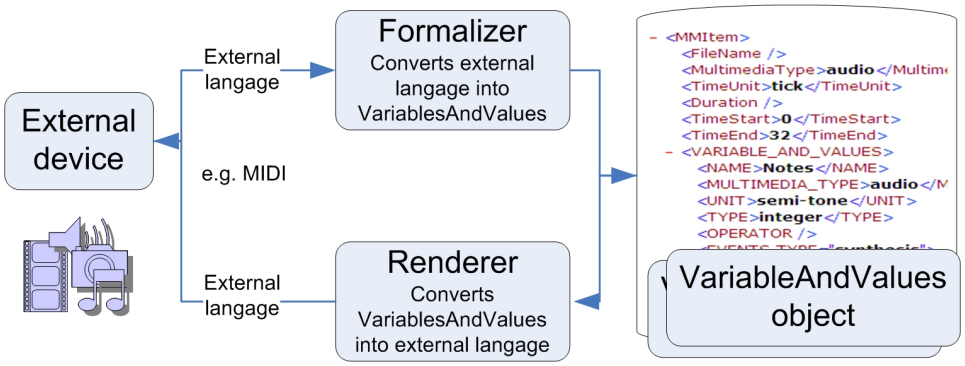

Figure 4.20: The mechanism to formalize and render VariableAndValues from and to any external device.

4.5.3 Testing LTAM initialization

We tested the LTAM initalization by writing a set of classes to allow any user to create MMItem, make them evaluated by a user and then build the ltam. We list the steps to build the ltam of a user. First (figure 4.21 step is the creation of a mmItem. In the presented example we used two kind of techniques above-mentionned. We load couples of associated files made of one multimedia file (*.jpg, *.avi,*.wav) and one formalization file (*.xml) following the variableAndValue format[2]

[2]Note that any kind of formalization could be associated to the file. The only condition is to write an appropriate formalizer which will be able to convert the associated formalization (e.g. audio MPEG-7 descriptors) into a VariableAndValues component. Moreover

We also load unique formalization file, containing the tag

```
<RENDERER>MIDI</RENDERER>
```

to tell the parser that this file is not associated to a media file but should be rendered by an implemented 'MIDI' renderer.

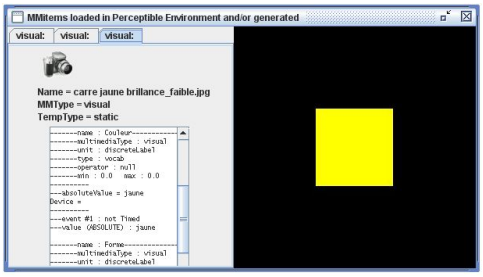

Figure 4.21: Loading MMitem to prepare an aesthesis.

Once a mmItem is loaded, our implementation show a panel containing the content of the VariableAndValues set associated to the loaded mmItem (left part on the figure 4.21. The user can clic on the icon situated in the upper left part of the interface (the icon displayed is associated to the multimedia type of the mmItem, i.e. visual, audio, video) and watch or listen the mmItem on the right part of the panel.

Figure 4.22: Selection of MMItems.

Among all the loaded mmItems, we can select a group which will be used for the next aesthesis (see figure 4.22. The aesthesis is then performed by any implementation which can output AesthesisResults components. We implemented a software interface (d'n'dMultimedia) allowing user to place mmItems into a space made of valence and arousal dimensions of affective state. We thus get a coordinate for each mmItems evaluated by the user. This coordinate is used to create an EmotionRepresentation component, in which we are able to convert theses coordinates into other type of emotion representation (e.g. discrete).

In the example we asked a user to evaluate four mmItems (3 visual, and one audio). Finally, the ltam is created on the basis of aesthesisResults and by applying the algorithm proposed in section 4.3. The LTAM is displayed

in this case no renderers are needed as the file could be directly played.

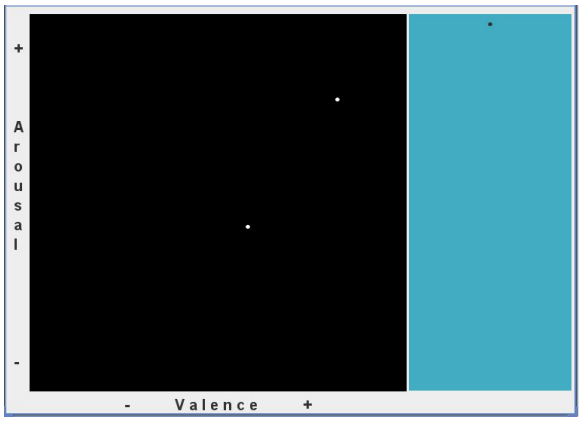

Figure 4.23: Aesthesis : measure of the affective information expressed by an individual about the selected MMitems

in the form of a undirected graph with perceptual vertices on the left (blue) and affective vertices on the right (yellow). The left perceptual vertices represents the stimuli, i.e. the mmItems, the middle perceptual vertices the patterns and the right perceptual vertices the values.

4.5.4 Testing LTAM update

Once created, a LTAM can be udpated following the algorithm presented in the section 4.3. Any incoming aesthesisResults can be sent to the current LTAM associated to an individual. The mechanism is similar to the intialization but the differences lies in the research of already exisitng vertices. In this case we do not create new vertices (i.e. if the user already experienced the green color and the green color is present into a new aesthesisResults) but associated the current results to the existing node.

This approach is presented in the figures 4.25 and 4.26 where an udpate occured between the two states of the LTAM.

4.5.5 Testing with an external device

We tested the possibility to build a LTAM from an external device and the associated aesthesis. We thus use a Yamaha Disklavier, as the only implemented renderer and formalizer is the midi standard. The approach is presented in the figure 4.27. The first step from developper point of view, is to write implementations for the Formalizer and the Renderer components (provided in the form of interfaces in the java API). We did this step by implementing the following classes : FormalizerMidi and RendererMidi. The first class has the ability to record any internal or external synthesiszer and translate the midi norms instuctions into VariablesAndValues components. The RendererMidi class can render any VariablesAndValues component into

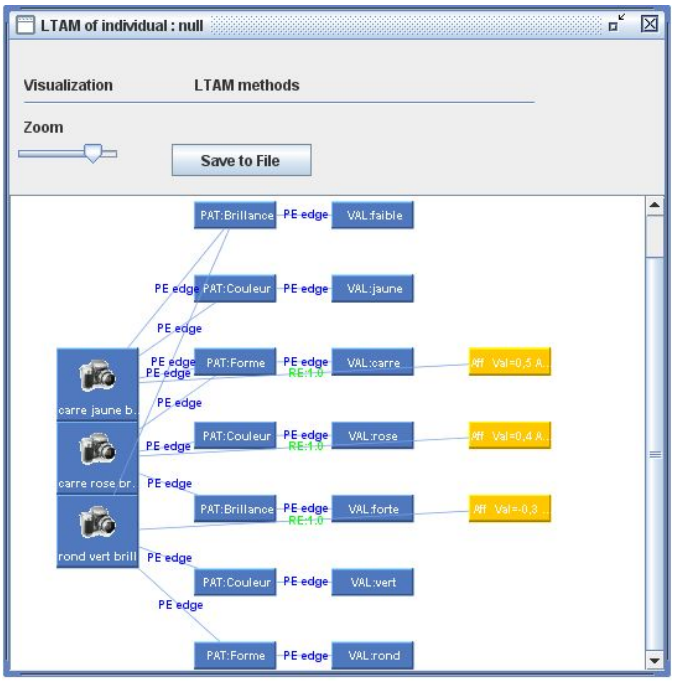

Figure 4.24: the output of the LTAM from the aesthesis results.

either internal midi device (i.e. it plays the sound in the computer) or external midi device (i.e. it send information to an external synthesizer e.g. a midi piano).

We were thus able to connect a Yamaha Disklavier (midi piano which has the ability to sense any musicians action into midi instructions and to output a musical performance to the piano. Using EAR formalizers and renderers (with midi norm) a LTAM could be built from the play of a musician associated to the measure of its affective information.

The first step of the figure 4.27 is the performace using the piano. While someone plays, the midi formalizer receive midi instructions associated to the action of the musician on the piano. The FormalizerMidi can then output VariablesAndValues components translated from the midi instructions.

Then MMItem are created within the software by combining the VariableAndValues components. Those MMitems are sent as the basis for an aesthesis.

During the aesthesis session, emotion measure (i.e. EmotionRepresentations performed directly or using indirect measure as physiological signal input) are performed to classfiy the MMItems. Such aesthesis could be performed at the same time someone use the external device allowing the second and third scenario presented in the figure 4.3. Given any real-time emotion measure, aesthesis can performed on the fly for each piece of music

Figure 4.25: An example of LTAM.

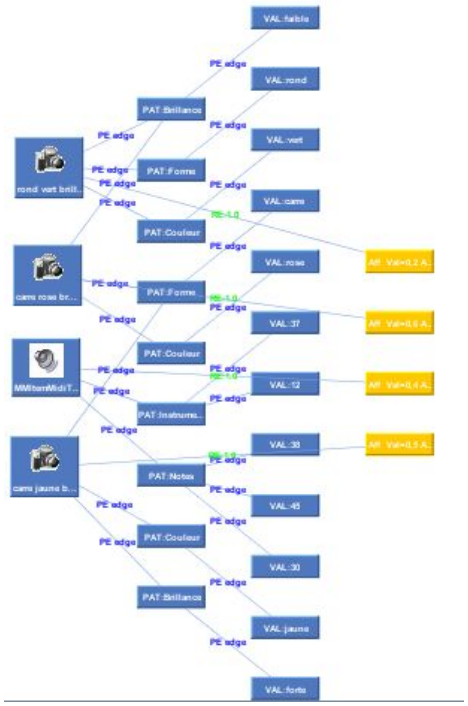

Figure 4.26: The same LTAM after an update using incoming aesthesis results.

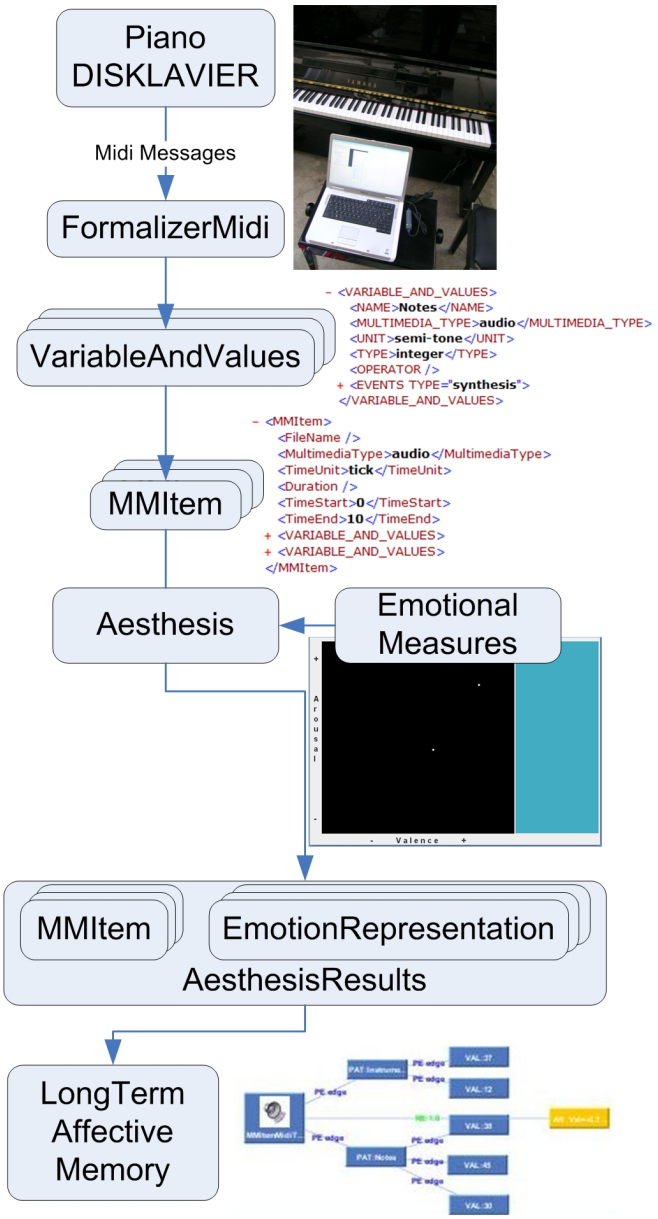

Figure 4.27: The steps to build the ltam from an external device (e.g. a midi piano)

performed.

Once we get the aesthesis, we produce the aesthesisResults component. This result is used to initialize or update the LTAM. The LTAM could then be used as a basis for any kind of interaction with the user.

4.6 Conclusion

In this chapter we have shown how several fields of experimental researches focuses on the *human ability to associate affective experience to media*, along with the fact that computing could take benefits of theses researches in several applications involving affects and emotion.

The inter-individual difference among individuals reacting to same media had been poorly investigated and is still a barrier to the development of personalized applications like multimedia content delivery and HCI design. We proposed to consider the notion of Embodied Affective Relationship as a mean to take into account cultural and personal individual past affective experience responsible for the individualisation of affective experience to perceptible environment. We shown a formalized model of this notion, along with an implementation, to simulate the EAR of an individual with a controlled environment and then use it into multimedia application.

The implementation gives opportunity to embeds results of several field of research into a software platform which, we think, could be an interesting (experimental) plug-in for several existing and new application, like interactive art based on emotion. We did not performed user studies with this model : the aim was to build a computational aproach from natural cognition knowledge to handle multimedia contents using affective information. The number of features involved in emotion is a huge field of research and was not the purpose of this chapter. Any perceptual features (description of multimedia contents) used to be associated to emotion representation could be added depending on the domain (e.g. music, visuals, odors).

The mechanism of association between perceptual features and emotion representation is based mainly on psychology and neurosciences and could now be tested practically with this approach. The goal of this approach was reached : we provided a practical framework to associate perceptual features describing multimedia contents and emotion-related representation in a systematic way.

As presented in the chapter 6 a software specification is formulated to enable programmers to extends the approach on the presented basis. The mechanism of retrieval is implemented and could let developper to build different type of retrieval techniques. The architecture related to evaluation is also implemented (confusion matrix evaluation). The type of features could be tested, e.g. in music research, while novel forms of interaction could be designed, e.g. in interactive art, home design, etc...

Improvements in the EAR architecture could be done by adding other fundamental research findings into the model and the software (e.g. average rules related to emotion and musical structure), and working on the connection with existing systems (emotion recognition, multimedia control, etc...).

The next chapter is dedicated to the second related user modeling problem, i.e. the indirect measure of emotion from physiological contents.

Chapter 5

Physiological Indirect Measure of the Subjective Experience of Emotion (P.P.E.M. Model)

5.1 Introduction

Emotion is generally considered as a mind-body phenomenon. It exists several observables of emotion, depending on the level of organisation considered (social, psychological, cerebral, cellular, molecular). For example, emotional phenomenon exists at *social level* with empathy, or behavioral aspects like facial expression and body movements ; at *psychological level* with consiousness of emotion ; at *cerebral level* with measured activity of limbic system while experiencing emotion, or physiological aspects like chills ; at *cellular level* with the excitation of hypothalamus neurons related to fight or flight process ; and at *molecular level* with study of neurotransmitters situated in limbic system.

In this chapter we present a model to link two opposite kind of observables of emotion :

- ≻ an expression of consiouss aspect of this phenomenon : psychological level, accessible with a *1 st person approach*
- ≻ some peripheral expression of the Autonomic Nervous System (ANS) activity, modulated by the emotion trought the control from subcortical structures : physiological level, accessible with a *3 rd person approach*

The aim of this chapter is to manipulate theses two measures of emotionals situations, confronted by the concepts of 1 st and 3rd person approach, using measure from the psychophysiology research field.

- ≻ In the case of the 1 st person approach, the interpretation of the subject's emotional experience is done by the subject himself. It corre-

spond here to a psychological approach, as we explicitly request the subject to express the affective state (s)he's feeling.

- ≻ In the case of the 3 rd person approach, its an external observer (the experimenter) who produce the meaning associated with the measured signals. In this case, such signals are the physiological one.

Including affect sensing in Human Computer Interaction (HCI, e.g. in teaching application), and Computer Mediated Communication (CMC, like mailing, artistic collaborative systems), are dependant on the possibility of extracting emotion in a continuous way, independent from the explicit user request of his(her) affective state.

Actually, an interruption in HCI, with the purpose to explicitly ask user about his (her) current feeling could modify true feeling of user. Such interruption should be avoided by a suitable third person approach, i.e. a continuous extraction of actual affective state of user, without explicit asking for consciously experienced emotion.

As stated in the section 3.4, several knowledge from the physiological domain are not necessarily used. Therefore we aim at following the next objectives to adopt a cognitive science oriented approach rather a machine learning approach as defined in the figure 1.2.

- ≻ Objective 1 : **Take into account physiological knowledge** by extracting and selecting only emotion-relevant features from the ANS signals.

- ≻ Objective 2 : **Take into account existing representations of emotion** by combining different emotional representations (e.g. discrete dimensional

- ≻ Objective 3 : **Adopt a 1st / 3rd person methodology** by considering the subjective experience of emotion (1st subject) as the output of interpretation (3rd observer)

- ≻ Objective 4 : **Make a unique descriptive model** by combining user-dependent and user-independent data, and by combining both known/average psychophysiological mappings and subjective modulation

Motivated by the possibility to extract user emotion, the aim of this chapter is (1) to show experimentally how the psychological and the physiological cues are related, and (2) propose a method and a system to infer psychological meanning from measured physiological cues.

This chapter is organized as following. We present an overview of the methodology used, and an overview of the proposed model built upon the

litterature related to the Autonomic Nervous System (ANS), and the possibilities to extract information from it, especially emotional one presented in the section 2.

Then we present an experiment along with the hardware system and software algorithms for heart rate and skin conductance extraction and analysis, including design and software implementation. The data analysis lead in the creation of subject's PPEMs.

Finally, an experimental software is provided to extract emotion cues near to real-time.

5.2 Methodological overview

Several physiological peripheral activities have been found to be related to emotional processing of situations where the subjects are. We mean by situation the exteroceptive and interoceptive state of subject, also the present and historic of environment exposure, events elicitation, etc ... The physiological parameters studied here are the heart beat rate and the skin conductance.

5.2.1 Framework

The **proposed framework** consists on extracting in real time heart beat rate and skin conductance, and, using the appropriate PPEM, extracts current emotion of the user (see figure 5.2.1).

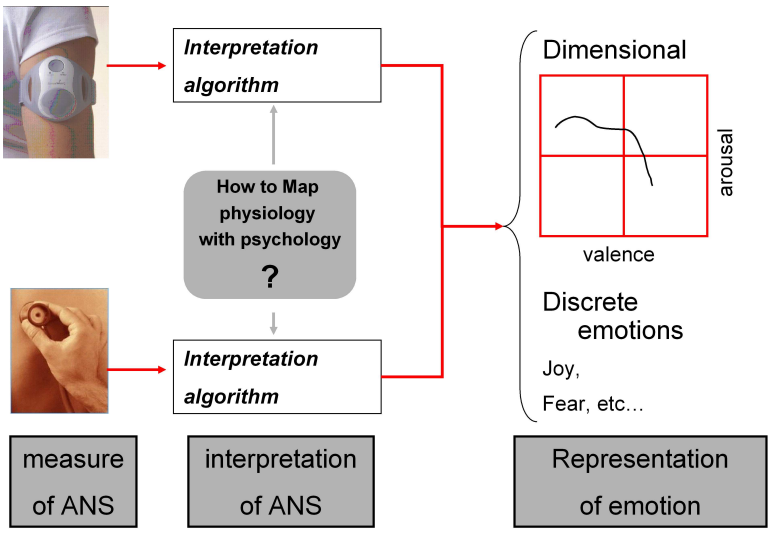

Figure 5.1: Proposed Framework to extract affective state, in continuous and discrete emotion representation.

5.2.2 Proposed methodology

The **proposed methodology** consists on two steps, illustrated in figure 5.2.2. The first is the experimental one, based on the 1 st and 3 rd person approaches. We propose an emotional situation to the subject, and measure its emotional evaluation with both psychological method (e.g. explicit emotion expression) and physiological measures (using cues related to emotion). The relations between theses two types of data are analyzed to extract a semantic of physiological measure (which we call Psycho Physiological Emotional Map - PPEM, see section below). The second step is the applicative one, using a 3 rd person approaches, with a reference to the results of the previous 1 st person approach. We continuously extract physiological parameters from the user, and try to extract in real-time the affective state his feeling, according to the semantic interpretation of the measures, done with the previously assessed PPEM.

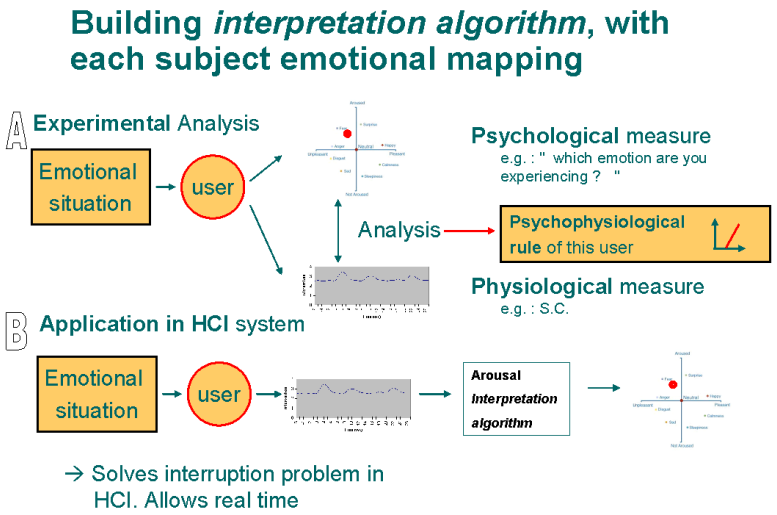

Figure 5.2: Proposed Methodology to build interpretation algorithms.

5.2.3 Psycho Physiological Emotional Map (PPEM) definition

The PPEMs are the mode of representation of the psychological link to physiological features. The above-mentionned methodology assumes the hypothesis that the *static* measures (post presentation of stimuli stimuli, i.e. an emotionnal resultant) and *dynamic* (while the subject his experiencing the stimuli) are closely related. Actually the idea is that an expressed static evaluation of an emotional stimulus (i.e. with discrete emotion belonging questionnaire, or with continuous representation of emotion scale) is pro-

duced from the analysis of the felt affective dynamics during the stimulus experience. Thus, we should be able to find some rules whithin the dynamics of affective experiences of each stimulus, linking the dynamic measures (here the physiological measure), with the static measures (here the psychological ones) of stimuli affective evaluation.

Indeed, at the opposite of [Kim et al., 2004] who set up a "user-independant emotion recognition based on physiological signal" , the aim of our approach is to tailor the interpretation of physiological measure, refering to psychological measure, on the same person.

As mentionned in section 5.2, the idea is to be able to create a mapping of physiological onto psychological measures, what we call the Psycho Physiological Emotional Map (**PPEM**). Because we want such mapping to be suitable for, and so tailored to, several user, we decided to model the PPEMs as a sum of common processes to the population, modulated for each individual, and which could be modulated within an individual due to specific reasons.

We define the link exhibited by a subject between the physiological and psychological measures of a given emotional situation as a PPEM. Let be $PPEM_i$ the PPEM associated to the subject i. Let be S a group of specific physiological patterns, represented as sets of features values derived from the physiological signal.
Each set of features values, is denoted by $S^f_{n_f}(j)$,$j = 1, \ldots, s$ where s is the number of sets. $S^f_{n_f}$, $f = 1, \ldots, F_j$ is a set of F_j features (denoted by f) values (ranged from $n_f = 1, \ldots, N_{f,j}$ for each feature f) computed from the physiological signal.
For example, let's considering the pattern $S^f_{n_f}(1)$, defined by a succession of 3 SCR amplitude values represented in $S^1_{n_1}(1)$, with $N_{1,1} = 3$ and the succession of 10 energy values in MF Bands of Heart rate PSD represented in $S^2_{n_2}(1)$, with $N_{2,1} = 10$).

Let be (x, y) a coordinate in valence*arousal space. Let be (x_j, yj) the coordinate of a point j in the valence*arousal space, and (a_j, bj) a value to add to current coordinate of the (x, y) point.

The *single subject* form of $PPEM_i$ is a set of psycho physiological associations. The psychological part is denoted by a coordinate (x_j, y_j) (see equation 5.1), or by a dynamic (a_j, b_j) (see equation 5.2) :

$$PPEM_i = \{(x_j, y_j), S(j)\} \quad \text{with} \quad j = 1, \ldots, N \quad N\text{number of PPEM element} \tag{5.1}$$

$$PPEM_i = \{(a_j, b_j), S(j)\} \quad \text{with} \quad j = 1, \ldots, N \quad N\text{number of PPEM element} \tag{5.2}$$

Once created, a PPEM is used as following by a recognition system. Let be V a set of feature values, in the form of $S(j)$. Let be $PPEM_i(V)$ the

function associated to the $PPEM_i$, which returns a specific coordinate (x, y) or dynamic (a, b), see equations 5.3 and 5.4 :

$$(x, y) = PPEM_i(V) \tag{5.3}$$

$$(a, b) = PPEM_i(V) \tag{5.4}$$

Let be t a threshold. The $PPEM_i(V)$ is defined by (see equation 5.5) :

$$PPEM_i(V) = \begin{cases} (x_1, y_1) & \text{if } |V - S(1)| < t \\ (x_2, y_2) & \text{if } |V - S(2)| < t \\ \dots & \text{if } \dots \\ (a_3, b_3) & \text{if } |V - S(3)| < t \\ \dots & \text{if } \dots \end{cases} \tag{5.5}$$

The PPEM allows to retrieve psychophysiological relationship qualitatively. However, each user should have a unique PPEM, which make the comparaison between user difficult, and to relie on previous findings.

We defined the *single subject* form of $PPEM_i$. Let's define now the *parametric* form of PPEM : $PPEM'_i$, which should return the same result as the $PPEM_i$, but with a different internal process. In this form, the psychological output is based on the PPEM of a virtual subject ($PPEM_{\text{average}}$), which represent the prevous findings in the litterature, i.e. the pshychophysiological links of the average population. To exhibit the inter-individual differences, as a modulation of $PPEM_{\text{average}}$ output, we introduce $dx_{j,i}$ and $dy_{j,i}$, which represent the subject i modulation of the average population results, for the pairs $((x_j, y_j), S(j))$. To exhibit the intra-individual differences, as showed with "Day-dependance" phenomenon ([Picard et al., 2001]), we introduce $dx_{j,i,c}$ and $dy_{j,i,c}$, which correspond to the subject i modulation due to specific conditions c, as day, moment of the day, etc... Let be $j = 1, \dots, N$ with N the number of PPEM elements. The $PPEM'_i$ is (see equation 5.6:

$$PPEM'_i = \{((x_j + dx_{j,i} + dx_{j,i,c}, y_j + dy_{j,i} + dy_{j,i,c}), S(j))\} \tag{5.6}$$

We can consider that $dx_{j,i}$ and $dy_{j,i}$ correspond to a personality of the subject (see section 3.4), while $dx_{j,i,c}$ and $dy_{j,i,c}$ are more related to mood and body state, and so day changes. So, the litteral form of $PPEM'_i$, is : averagepopulation + $user_i$ delta + mood of $user_i$ delta.

The Psycho Physiological Emotional Map (**PPEM**, see figure 5.3) aim at mapping physiological emotional measures with associated psychological emotional measures in a emotional given situation, for a specific user (but using both user-dependent and user-independent data). The PPEM are not implemented with a specific machine learning technique. PPEM are made of a dictionary. A comparison of several implementation techniques for psychophysiology could be found in [Changchun et al., 2005]. The proposed

model does not aim at optimizing the emotion prediction using optimized machine learning techniques. At the opposite our approach aim at refining the process of emotion prediction by questionning methodologies and dat used from psychophysiology.

The PPEM consider two phase: a learning phase followed by a use phase. The learning is done one-time, before the interaction with the system. This learning could be done again, if we want to precisely model short term changes as the day-dependant phenomenon. This learning extracts significant statistical rules between psychological (the valence and arousal positions) and physiological features. Once found, the rules are stored into the PPEM of the current user.

In a learning phase, we provide a set of emotional situations to the user (1), which elicit affective experiences (2). We perform psychological (3) and physiological (4) measures associated to the affective experience. The psychological measure can be converted into different representations (discrete and dimensional). A set of features extraction is performed from the physiological measure. Then, the user model called PPEM (single subject form) is built from the association of psycho-physiological measure (5), for the user i. Then, from a PPEMaverage (user-independent data : synthesis of existing findings in terms of psycho-physiological maps), we build the modulations from this average for this subject. Finally, by combining these modulations with the PPEMaverage, we build the PPEM'i (parametric form combining user-independent and user-dependent data) which will be used to recognize emotion.

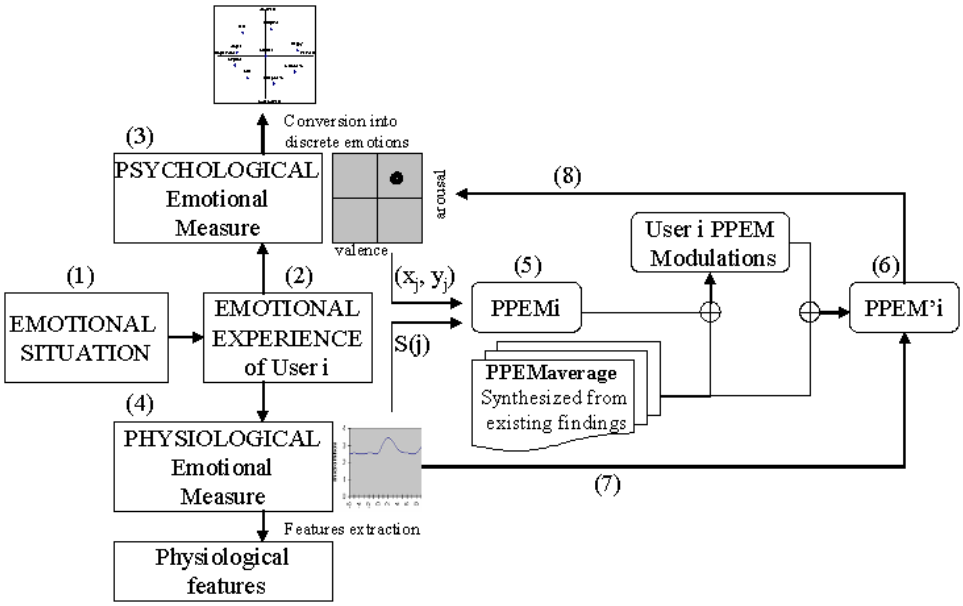

Figure 5.3: The Psycho Physiological Emotion Map

In a use phase, we continuously measure physiological signal from the user and extract features related to emotion (7). By comparing the current features values, with the contents of the PPEM'i, we estimate the emotion representation actually felt by the user (8). The use phase may be made of a distance calculation between the current set of features measured on the user and each sets of features (Sj) associated to psychological coordinates in the PPEM of the current user. However any machine learning techniaue could be considered to perform this task.

5.2.4 Representing previous litterature results as $\text{PPEM}_{\text{average}}$

We provide here a synthesis of found results at the population average level, in terms of $\text{PPEM}_{\text{average}}$. We built the map using bibliography of the section 3.4.3.2, and the pointers provided by synthesis of [Lisetti and Nasoz, 2004] and [Peter and Herbon, 2006]. We considered relationship between heart related features and skin conductance related features with psychological dimensions of valence and arousal, or in discrete emotions representation (which could correspond to a specific region in the valence arousal space, see 3.4).

$\text{PPEM}_{\text{average}}$ is made of *static* psychological representation(mainly discrete emotion in the reviewed litterature), or *dynamic* psychological representation (as change in arousal) related to a physiological signal pattern (either *static* or *dynamic* representation). As proposed into the PPEM formalization (5.2.3), the notion of pattern could be e.g. a SC value, or a SC derivative, or a derivative of the ampltude of successive SCRs.

5.2.4.1 $\text{PPEM}_{\text{average}}$ from valence and arousal representation based studies

Main relationship are statistical linear relationship without detailled information about the nature of the linear relationship (e.g. valence = 5*SCRamplitude). Thus we model it using the dynamic form of the PPEM, and with values to specify empirically.

Table 5.1 summarize the bibliographical findings as well as our proposed modeling into PPEM. The PPEM is in the form : $PPEM_i = \{(a_j, b_j), S(j)\}$.

	Psy. emotion rep.		$\text{PPEM}_{\text{average}}$ rep.		
Phy. features HR-related	**Valence**	**Arousal**	a_j	b_j	S_j
HRAverage	positively		$> 0^*$		$HR+' > 0$
SCR amplitude		positively		$> 0^*$	$SCR+' > 0$
SCAverage		positively		$> 0^*$	$SC+' > 0$

Table 5.1: $\text{PPEM}_{\text{average}}$ using a dimensional emotion representation. * means that the value should be set empirically.

Let be $PPEM_{\text{average}}$ the generic PPEM standing for the average population from the litterature. As a source, we can consider the synthesis of

[Lisetti and Nasoz, 2004] and the comparative synthesis of [Peter and Herbon, 2006]

5.2.4.2 $PPEM_{average}$ from discrete emotions representation based studies

We considered C_j^c as the converted coordinate of the (x_j, y_j) psychological part of the element j. Discrete emotions classes used are constituted by 7 classes instances : {Sad, Happy, Calmness, Surprise, Neutral, Fear, Sleepiness, Disgust, Anger}. In this case $PPEM_i = \{((x_j, y_j), C_j^c), S(j)\}$, with c ranged from 1 to 7. Table 5.2 presents the PPEM associated.

Phy. features	**Psy. emotion rep. and $PPEM_{average}$ rep.**					
C_j^c	Sad	Happy	Surprise	Fear	Disgust	Anger
c	1	2	4	6	8	9
HRAverage	$HR+' > 0$	$HR+' > 0$	$HR+' > 0$	$HR+' > 0$	$HR+' > 0$	$HR+' > 0$
SCAverage	$SC+' < 0$			$SC+' > 0$	$SC+' > 0$	$SC+' > 0$

Table 5.2: $PPEM_{average}$ using a discrete emotion representation. * means that the value should be set empirically.

5.3 Experiment : Materials and methods

5.3.1 Variables

According to the main hypothesis of this study, which consist on the possibility to tailor the interpretation of physiological signals in term of psychological (self-report, 1rd person) meaning, the variable are divided in three categories. The independant variable are the valence and arousal individual evaluation, constituting the *psychological* variables. The dependant variables are the heart beats and the skin conductance, constituting the *physiological* one. Moreover, a questionnaire regarding the general emotional state (personality) and regarding the past week was implemented, using a french translation of the Positive Affects and Negative Affects Scale (PANAS) ([Gaudreau et al., 2006]).

5.3.2 Subjects

40 subjects, 21 men and 19 women involved in different socio-professional activity, from 19 year to 53 years old (average 32) and without known cardiac troubles, participated to the experiment. Subjects were paid to participate.

5.3.3 Procedure

After a brief explanation of the four steps of the experiment, each subject was invited to read and sign a consent form compliant with national ethical

comitees (CCPPRB and CNIL). Then, physiological sensors were installed on subject (skin conductance sensor and cardiac activity, see section 5.3.6). Subject was requested to sit in the experiment room, in front of a screen of 17 inches from which (s)he watch and listen to stimuli, and in presence of headphones.

Subject could interact with the system with a PC mouse, and keyboard when requested. The experiment duration was approximatically of 1 hour and 15 minutes for each subject.

The logic of the experiment is a three steps exposure of the same multimedia items, combining physiological and psychological measure of emotion. The experiment is made of four steps.

1. phase : slideshow of multimedia items (the stimuli) and recording of physiological measure (heart rate and skin conductance).This step constitute the 3 rd person measure of subjects' emotion.

2. phase : *static* classification of the same multimedia items in the emotional space of expression made of valence*arousal dimensions

3. phase : *dynamic* measure of the valences, during a slideshow of a selection of multimedia items (only the dynamic ones),

4. phase : Questionnaire including the PANAS scale questionnaire using the general and last week time period (see 5.3.1)

The experiment is finished after the four steps are completed.

5.3.4 Software engineering

The software *d'n'dMultimedia* we used in this experiment was a multimedia extension of the d'n'dSound software previously designed ([Villon, 2003], p.8 and Annex 6 p.47). It is available.

For synchronisation purpose, a button had to be press on the skin conductance system, at the end a countdown done by the d'n'dMultimedia software. The subject was warned before about this. After this the subject was requested to close its eyes and to perform 3-4 long breaths. After, the steps of the experiment were completed by the subject. A slider measure was also implemented, with a syncronization to the slideshow of stimuli.

5.3.5 Choice and duration of the stimuli

The set of stimuli was choosen (1) to be suitable to elicit physiological responses and (2) to be emotionnally varied (both in intensity and type).

5.3.5.1 Emotional variety

A set of 61 stimuli was selected to be varied both regarding the type of media (audio, visual et video) and the intented emotional caracteristic, to try to cover the most extended range of emotion. 31 images from the International Affective Picture System (IAPS, [Lang et al., 2005]) were selected. The IAPS is a set of 944 images which comes with a normative emotional rating of each images, in term of valence, arousal, and dominance. The emotional ratings are based on several studies previously conducted where subjects were requested to rates theses images using the Self Assessment Manikin [Lang, 1980]. As if such study doesn't take into accounts inter-individual differences in the affective ratings by averaging the responses ([Villon, 2003]), it is however a useful and widely used indicator of the average individuals emotional evaluation. Thus, we used this database with the objective to propose a varied set of emotion, whereas our analysis (see 5.2 and 5.3.7) doesn't use this normative evaluation, but instead use the evaluation produced by subjects. Several selection algorithms were tested, with the aim to select a representative subset of images within the set. A java class (available on demand was designed to handle the IAPS database (retrieve IAPS images file by Valence, Arousal or Dominance, etc...), and perform selections.

Four selection algorithms were designed :

- RANDOM performs a full random over the set.
- DISTRIB extracts a subset of points according to the distribution in terms of density and spatial location. This algorithm divide the set using a grid, count the points in the grid and then select a proportional amount of point in this grid, according to the desired amount of points to extract from the whole set.
- SPATIAL_LOC_CENTER extracts a subset of point according to the distribution in terms of spatial location only. This algorithm divide the set using a grid (based on the boundaries of the whole set) whose dimensions are proportional to the desired amount of points to extract from the whole set. Then, the central point of each grid area is extracted.
- SPATIAL_LOC_RAND acts as the last algorithm, but extracts a random point inside each grid aera, instead of extracting the central one.

We get the more varied sets using the SPATIAL_LOC_RAND, which respect the best the distribution in terms of location only of the points (see figure 5.4).

25 sounds extracts were prepared. The sounds extracts are characterized by their musical or non-musical nature. The musical extracts were varied

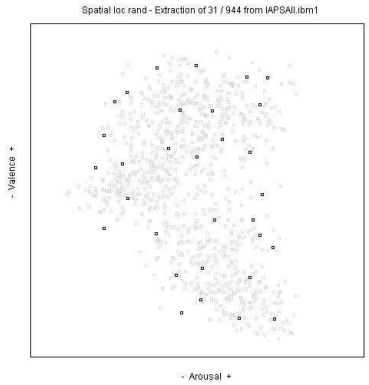

Figure 5.4: The selection from IAPS set. Each point represents an image of the set whose coordinates are the normative rating regarding valence and arousal. Gray points represents the whole set, and black one the selection we made.

in style (reggae, rock, folk, classical, zeuhl, acousmatic), location (indian, african, brazilian, european), contents (instrumental, vocal), and from traditional or classical repertory. Also, the duration was ranged from 4 to 46 seconds, with an average of 29.48, as well as the average beat per minutes. The non-musical sounds extracts consists of a soda can-opening, a scream, and very high (dentist-like) and low pitch tones.

5 videos ranged from 8 to 149 seconds (average 68.4) were used. Two of them were extracted from the recommendation of [Rottenberg et al., 2006] (i.e. *Harry meet Sally* and *Shinning* scenes), and the three other were extracted by us for their intented emotional caracteristics. We choose a scene from *Amelie Poulain*, where someone drink then spit out the contents of his glass (intented to elicit disgust), a scene from *Latcho Drom*, where an indian girl dance (intented to elicit aesthetical emotion), and a scene from *La Vita e Bella* in which a women discover that a buglary had just occurs in her house.

Thus, the chosen stimuli were *intented to be* varied in the valence and arousal ranges. Actually we can't *apriori* know how the subject will *precisely* emotionally react to these stimuli, due to the fact emotional evaluation engage individual personal and cultural background ([Villon, 2003], see section 3.4.2). So, on one hand we have an average estimated responses for the images stimuli, and 2 videos. On the other hand, we have our set of stimuli, which are varied in type and thus considered as potentially varied in terms of emotions, for individuals.

5.3.5.2 Multimedia items duration and physiology

Physiological responses. The multimedia items were choosen to be superior to 30ms for SCRs elicitation (see 3.4.3.4), and enough long seconds for Heart Rate changes. So the multimedia items were chosen to be superior to this duration, for audio and video. The presentation duration of images during the slideshow (Image Duration Display, IDD) was set to 6 seconds.
Emotional consistency.The length of stimuli was chosen to be varied (from 3 seconds to 2 minutes 30) long in order to get more robust physiological recordings to duration.

5.3.5.3 Inter stimulus interval

Given that SCRs post stimulus presentation reaction is rapid, and given that typical duration of SCRs is around 1-3 seconds (3.4.3.4), the inter-stimlus-interval (ISI) had to be set superior to 3 seconds in the slideshow presentation (physiological recordings, see 5.3.6), to be largely sufficient to allow to a new SCR to appear, without superimposition with previous SCRs. Regarding HRV, the ISI was intented to be sufficient to let the heart return to a normal variability during the ISI. As no precise values was found in the literature, and due to the wish to not have a whole duration of the experiment exceding 1H15 minutes, the ISI was set at 5 seconds.

5.3.6 Physiological recordings : 1st phase

The physiological recordings were realized during the slideshow of multimedia items. The multimedia items were presented with the ISI duration of 5 seconds and an IDD 6 seconds. After a countdown, a synchronisation with the external physiological system for skin conductance measurement was done by pressing a timestamp button, to get a common time reference with the slideshow timing of multimedia items presentation.

5.3.6.1 Skin Conductance

Hardware device. The bodymedia armband, along with the InnerView Research Software 4.1 from Bodymedia (`http://www.bodymedia.com/`) was used to record skin conductance (see figure 5.5). Despite that this system is designed to work on the arm, we found that the data are more precise and amplified using the armband onto the hand palmar, a measurement site recommended since [Fowles et al., 1981]. Thus, subject were asked to wear the armband on the left hand, using a bipolar placement (on the forefinger and the middle finger), on medial phalanx. The timestamp button of the armband was pressed when the countdown of d'n'dMultimedia ended, to start the measure.

Figure 5.5: Bodymedia Armband : hardware device to measure skin conductance.

Software.The SC sampling rate was set to 32 sample.s^{-1} using the InnerView Research Software. After the experiment, data were retrieved trough the software, in the form of an Excel file.

5.3.6.2 Heart Rate measure using a phonocardiogram

Both hardware and software were made for this acquisition system. The software part is embedded into the java package : **com.earmultimedia.sensor**

Hardware device. The hardware was intended to be easily accessible to end user. Several techniques exist to measure heart activity (i.e. electric Electrocardiogram -ECG-, optic or phonologic). The chosen technique is the phonocardiograph one, easy to set up with a stethoscope, and with a low cost, measured trough a microphone. The microphone is plug into the mic-line of the computer. Such method give less detailed information about heart activity than an ECG, but allow heart precise rate extraction (see for instance [Sava and Durand, 1997]).

Heart Activity Acquisition : Volume measure. The volume measure was done in Java 1.5.0, with the Java Sound API (see [Bomers and Pfisterer, 2005]). The audio signal was sampled at 44100 Hz, 16 bits (precision of 65536 points). The volume extraction was tested on fixed window of 0.03 seconds (i.e. 1323 samples), compared to a precise floating time-window of 0.005 seconds (i.e. 222 samples). The fixed time window, which is better for real time computing, worked enough precisely to keep the shape and the timing of the peaks (see 5.3.6.2). Let be the volume sample duration $sd = 0.03$ and $Fs = \frac{1}{0.03} = 33.3Hz$ the sample frequency of volume.

A smoothing was applied to volume. According to [Smith, 1997], chap 15 : "the amount of noise reduction is equal to the square-root of the number of points in the moving average". The buffer length (l, in samples) of smoothing was calculated according to a noise reduction factor of 2.5, which was empirically set up ($l = 2.5^2 = 6.25$, i.e. 6 volume samples). The smoothing buffer corresponds to 0.18 seconds ($l * sd = 6 * 0.03 = 0.18$ seconds). The smoothing returns the mean of the buffer. (see 5.3.6.2).

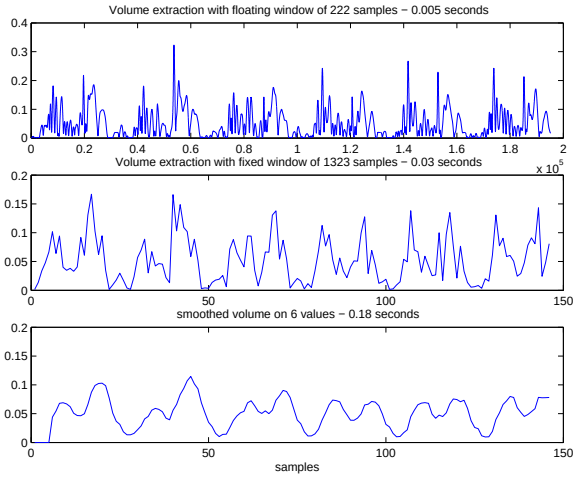

Figure 5.6: Volume measure from phonocardiogram.

Heart Beat detection. The heart rate extraction was designed in matlab and then implemented in Java 1.5.0, in a class called HeartRateSensorPhonocardiographSoundcard. As the processing applied on the stethoscope audio signal should be suitable for real-time processing, the matlab algorithm was designed to fit this purpose.

The peak extraction aims at extracting peaks into the volume signal, which corresponds to the heart beat. It exists differents methods, like in [Sava and Durand, 1997] for phonocardiogram acquisition and peak detection, [Jennings et al., 1981] for the main Digital Signal Processing of heart rate, [Pan and Tompkins, 1985] which propose a detection method of R wave in ECG. As there are no common technique to detect peaks corresponding to heart beat in a volume buffer, we propose our adaptive method based on peak detection, local maximum, comparison of amplitude and triggering according to the timing of each new detected heart beat.

The proposed method utilize a queue buffer. The size of the buffer is inferior to the minimum Inter heart Beats Interval (IBI) to be sure we get only one heart peak in the floating window. The minimum IBI was set to $0.3s$. (i.e. a maximum of 200 beats per minutes (bpm)). Thus, the size of the buffer is : $(s = \frac{min(IBI)}{sd} = \frac{0.3}{0.03} = 10$ samples)

Let V be the buffer of volume value and $V(n)$ the value situated at the half length of the buffer. Let VB be the buffer of volume value of the detected beats. Indead, let IBI be the set of detected Inter heart Beats Intervals (IBI) defined by $IBI(i) = t_i - t_{i-1}$, for each $i-th$ beat, occurring at time t_i, and $BPM(i) = \frac{60}{IBI(i)}$. Let be $IBI(lastbeat)$ the IBI corresonding to the last detected heart beat.

For any value $V(x) \in V$, the *left derivative of* $V(x)$, denoted by $V-'(x)$,

is defined by :

$$V_{-}^{'}(x) = V(x) - V(x-1) \tag{5.7}$$

and the *right derivative of* $V(x)$, denoted by $V_{+}^{'}(x)$, is defined by:

$$V_{+}^{'}(x) = V(x+1) - V(x) \tag{5.8}$$

A heart beat is detected if the $V(n)$ correspond to the following conditions (see 5.7) :

- $V(n)$ is a **positive peak** (i.e. $V_{-}^{'}(n) > 0$ and $V_{+}^{'}(n) < 0$)
- $V(n)$ is a **local maximum**(i.e. $V(n) = \max V$)
- $V(n)$ has the amplitude of an heart beat, compared to last three detected beats (i.e. $V(n) > \left(\frac{60}{100} * \frac{\sum_{i=1}^{3} VB_i}{3}\right)$)
- the IBI **acceleration** isn't too high (25 bpm) for an heart beat, as heart rate changes are slow (i.e. $\frac{60}{t_n - t_{lastbeat}} - \frac{60}{\left(\frac{\sum_{i=lastbeat-3}^{lastbeat} IBI_i}{3}\right)} < 25$)

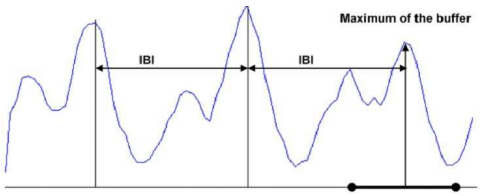

Figure 5.7: Heart Beat detection. Conditions to realize to detect a beat from the phonocardiogram.

The 60 % was chosen because the within-subject R-wave amplitude is quite constant in normal subjects (see [Aittomaki and Salmenpera, 1997] and [He et al., 1995])and so this value was used to discard non heart beat from detected peaks in the signal. The threshold of 25 bpm was chosen as cardiac rhythm changes are slow for normal subjects.

The IBI were stored in IBI buffer, then recorded to a text file containing the timestamp of the begining of the experiment (for synchronisation purpose).

5.3.6.3 Heart Rate using a Polar T31 transmitter

Both hardware and software were made for this experiment. The software part is made of two classes, embedded into the java package : **com.earmultimedia.sensor** (see chapter 6)

Hardware device. Among several existing techniques to measure heart activity (i.e. electric Electrocardiogram -ECG-, optic or phonologic), the

chosen technique is the ECG one. The hardware was made from a chip of Polar, the HFUi receiver, working with the T31 transmitter of Polar (placed on the chest). For each found Q-R-S complex signal (which contain the R-wave, the main peak used for heart beat timing), the T31 send a pulse to the HFUi receiver. The HFUi receiver is connected to the serial port of any PC, with an appropriate wiring (see figure 5.3.6.3). As the current delivered by the HFUi was not sufficient to elicit event detection into the serial port, a system with a transistor was designed. The HFUi is continuoulsy set at 9v. The poor current is amplified and the serial port detect the incoming voltage, between pin 1 and 5 which represent the CD signal (carrier detect for modems). When an heart beat is detected by the T31 transmitter, the HFUi receiver down the digital pulse to 0 volt, for $100ms$, which is detected by the serial port as a CD event.

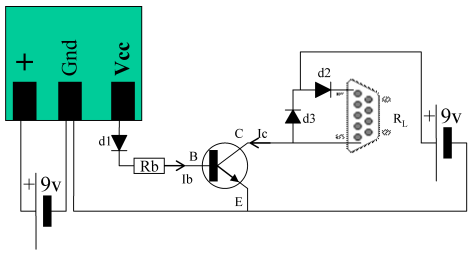

Figure 5.8: Our proposed system to connect the Polar's HFUi receiver to the PC serial port.

Heart Beat detection. The heart rate detection was implemented in Java 1.5.0. The processing applied on the serial port should be suitable for real-time processing.

An event detection was implemented using the Sun API for javax.comm package. The class HeartRateSensorPolarSerial.java. After searching for serial ports, the serial port named "COM1" is searched. Once done, and EventListener is created and added to this serial port. When an event occurs, we check if this event is the Carrier Detect signal, i.e. if some voltage changes occur between the pin 1 (CD) and pin 5 (signal ground). When the event is detected, we check that the IBI isn't inferior to the minimal duration of an heart period (300 ms, i.e. 200 $beats.min^{-1}$). We should only consider beats inferior to this period as artifacts of the polar/serial system. Then, the class deals with Inter-Beat Interval calculations, graphic plotting (following the structure of the class HeartRateSensorPhonocardiographSoundcard.java, previously designed to detect heart beat from a phonocardiogram).

The IBI were stored in IBI buffer, then recorded to a text file containing the timestamp of the begining of the experiment (for synchronisation purpose).

5.3.7 Psychological recordings : Affective experience

5.3.7.1 2nd Phase : static measure into valence and arousal space

The first psychological recording was done through the d'n'dMultimedia software, in drag and drop mode. Subjects were requested to manipulate the PC mouse to express the emotion felt for the presented multimedia items.

This interface (see figure 5.3.7.1) is provided to allow the subject to express its emotional evaluation of the multimedias items, represented as black dots in the right part of the interface. The subject can right-click on the dot to start the visualization or the listening of the item, and right-click again to stop the image visualization, or to stop the audio or the video before its regular end (user can stop the experience of a multimedia item when he already know where he will place the stimulus). The subject were requested to place each item on the left panel, with a drag'n'drop using the mouse left button, according to the emotional evaluation produced by the subject. The two dimensions are :

- **valence**(negative : you don't emotionally like the experience given by this multimedia item; positive : you emotionaly like the experience
- **arousal**(negative : weak emotional arousing ; positive : strong arousing, emotion really present at the consciousness)

The subject can move the items as many times (s)he wants. Subject should not only place items according to the proposed dimension, but also compare the items between-them in order to refine their position.

5.3.7.2 3rd phase : dynamic measure of valence

In this phase, subjects were aksed to evaluate the valence of dynamic multimedia items (sounds and vidéos). This was done using the up and down arrows of a PC keyboard, while the whole duration of the listenning and visualization of the items. As showed into the figure 5.3.7.2, the interface we implemented is constituted on a left panel, where videos are displayed, and a right panel made of a cursor. Subject were told by a written introduction that the more they place the cursor up, the more the item is giving subject a pleasant emotion as this moment ; the more they place the cursor down, the more the item is giving subject a pleasant emotion at this moment. Also they were told that places the cursor at the middle of the cursor means thta the item is neither pleasant or unpleasant at this moment. Finally, they were also told that the fact to move up and down respectively means that item is more and more or less and less pleasant.

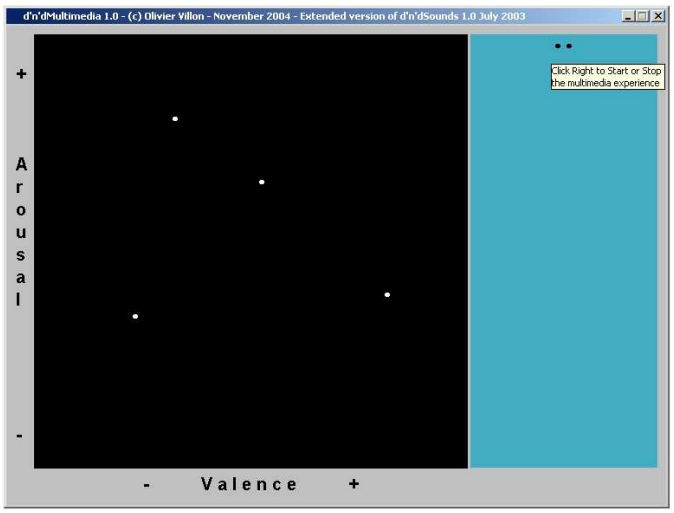

Figure 5.9: Psychological recordings of subject's emotion for multimedia items (represented as dots) through the d'n'dMultimedia software in drag and drop mode.

Figure 5.10: Psychological recordings of subject's emotion for dynamic multimedia items using a slider.

5.3.7.3 4th phase : questionnaire

A questionnaire made of two parts was designed. The first part serves to evaluates the experiment itself, by asking subject regarding the emotional caracteristic of stimuli ("Were you under the impression that you felt emotions ?", "Were you under the impression that the felt emotions were varied ?") and regarding the usability of the interface provided for the experiment ("Were you under the impression that you expressed yourself sufficiently in the «valence*arousal » interface ?"). Theses question were asked, using both a likert scale ranged from 1 to 5, and a text box.

The second part of the questionnaire consist on two PANAS Scale questionnaire. The PANAS scale ([Watson et al., 1988], [Watson and Clark, 1999] for the main version, and [Gaudreau et al., 2006] for the french one, which requires) is a tool to measure individuals affects. It was tested recently [Crawford and Henry, 2004] and found to be a robust means to measure affective state at personality level(life period), or at mood level (last week, month period).

This questionnaire is made of 20 adjectives which are associated to Likert scales ranged from 1, "very slightly or not at all" to 5,"extremely", accompanied by time instructions. We implemented the french translation (see fig. 5.11), with the translated time instructions ("Indicate to what extent you generally feel this way, that is , how you feel on the average", for personality level and "Indicate to what extent you feel this way during the past few weeks", for mood level).

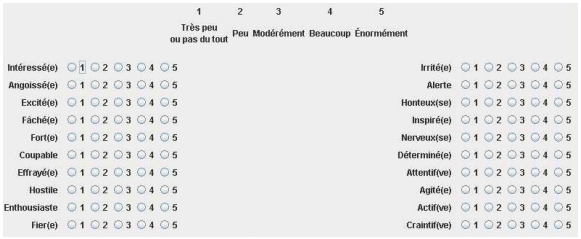

Figure 5.11: The interface designed to test the Affective state of the subjects at personality and mood level.

These questionnaire were designed to understand any inter-individual differences on this basis.

5.4 Data preprocessing : Physiological and Psychological features

5.4.1 Data preprocessing : Physiological features extraction

In the next section, we will use the term mmItem and stimuli to denote the stimuli used in this experiment, and t_{event}, as the moment associated to a particular event.

5.4.1.1 SCRs extraction

As mentionned above, the interesting signal in the raw SC signal are the SCRs. We will distinguish the SC, the Skin Conductance Level (SCL), which represents the tonic component of SC, and the SCRs which constitute the phasic component of SC. It exists several techniques to extract theses SCRs from SC.

- The band-pass filtering is a simple and quite efficiency technique. The SCRs are found in the bands 0.05 to 1 Hz, as in [Zink et al., 2004], p. 516).
- The substraction of the SCL (obtained with a low-pass filter set to 0.05Hz) to the SC signal, as in [Villon, 2002].
- The study of the derivative of the SC (dSC) as in [Nagai et al., 2004], or the first forward difference, followed by a triggering of the SC peaks above a threshold (see the paragraph "skin conductance" in [Healey et al., 1999]).
- A more precise and complex technique, is proposed by [Lim et al., 1997]. 'typical' SCRs are modeled with their characteristic form, then the model is searched in the SC measured signal.
- The simulation of sudomotor nerve signal (driver) activity and SCRs reconstruction, by [Alexander et al., 2005] is another precise approach, eliminating easily the notion of overlapping detection

As if the [Lim et al., 1997] method and [Alexander et al., 2005] are precise, we focused on a simple algorithm, suitable for real-time. We extended the method of [Healey et al., 1999]. The skin conductance was smoothed using a averaging window of 3 samples.

Let be SC the buffer of raw SC signal, and $SC(t)$ the SC value at time t. Let dSC be the derivative of the SC buffer, computed with the right derivative (see 5.8) : $dSC(t) = SC'_+(t)$. The peaks are first computed on the whole dSC buffer, with $dSC'_-(t) > 0$ and $dSC'_+(t) < 0$. A trigger is then applied to extracted derivative peak value situated above a threshold,

as denoted in (1) in the 5.12. The threshold was set empirically at 0.0001 μ-siemens, to separate SCR candidates from low increase of SC; which does not correspond to SCR. Once theses derivative peaks extracted, they are potentially considered belonging to an SCR (SC(t_dSCpeak) as they are situated at the middle of a SC high rise.

The first left negative peak (2, i.e. $SC(t_{\text{SCRstart}})$) of the SC is searched (i.e. $SC'_{-}(t) < 0$ and $SC'_{+}(t) > 0$) as it corresponds to the onset of the SCR. Then the first right positive peak (3, i.e. $SC(t_{\text{SCRpeak}})$) of the SC, which corresponds to the amplitude peak of the SCR is searched. If theses two elements are found, we are in presence of an SCR candidate.

Then, to ensure that we are in presence of an SCR, we applied to each SCR candidate an amplitude minimum threshold of 0.005 μ-siemens ([Dawson et al., 2001]) and a rise-time (i.e. $t_{\text{SCRpeak}} - t_{\text{SCRstart}}$) maximum threshold of 4 seconds (as rise time is usually between 1 and 3 seconds). Once a SCR candidate has been validated, we skip all other derivative peaks situated within the range of the last detected SCR (i.e $[SC(t_{\text{SCRstart}}), SC(t_{\text{SCRpeak}})]$) for the SCR search. Actually, it could exist several derivative peaks within this interval as SCR rise is not straight forward, and thus each of theses peaks should not lead to detection of new SCR candidates.

An example of extracted SCRs during the experience of an MMItem is provided in fig. 5.13.

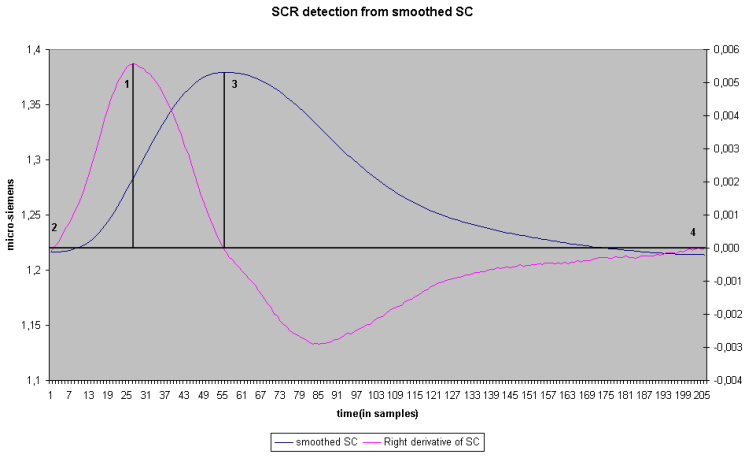

Figure 5.12: SCR extraction using derivative and peak detection above a threshold

Finally, each extracted SCR is modeled with a set of features

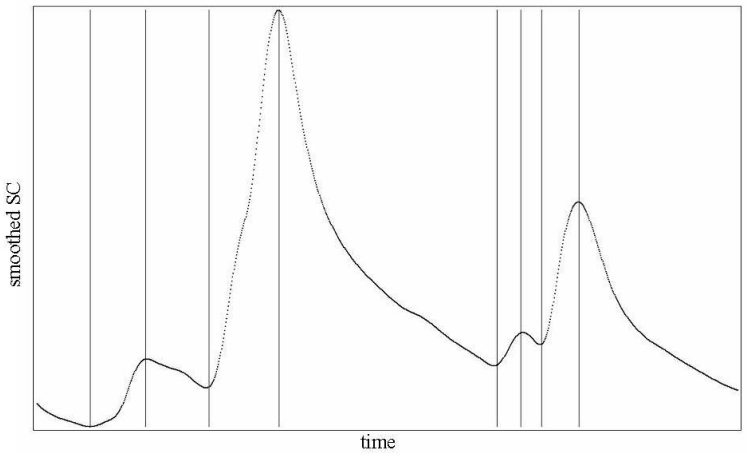

Figure 5.13: Example of extracted SCRs during the experience of an MMItem. The vertical lines delimit the start and peak points of each SCR. Then, each SCR is modeled with a set of features, see table 5.3

Table 5.3: Each SCR is modeled with this set of features.

Features for SCR	Extraction method	Definition
timeStart	$t_{\text{SCRstart}} - t_{\text{mmItemStart}}$	latency between the onset of the first SCR and the related stimulus onset(onset latency)
riseTime	$t_{\text{SCRpeak}} - t_{\text{SCRstart}}$	latency between the rise and the peak of the SCR
amplitude	$SC(t_{\text{SCRpeak}}) - SC(t_{\text{SCRstart}})$	the amplitude of the response, relative to the SC level
duration	3	
intensity	4	area of the SCR
relativeLatency	5	to the previous SCR, if this SCR is not the first, according to the stimulus onset

5.4.1.2 raw SC, SCL and SCRs related features

Several features were considered and calculated from the raw SC, SCL and SCRs signal, to then be used in statistical analysis. The SCL was extracted using a simple smoothing, applying a mean to a floating window of 90 samples (i.e. around 3 seconds of signal which correspond to the average most long duration of SCRs). For each stimulus (i.e. during the stimulus presentation), the raw SC, SCL and SCRs features are (see 5.4):

Table 5.4: SC-related features calculated for each mmItem.

	SC-related Features by MMItem	Description
SC (raw)	SCAverage	
	SCMaxAmplitude	
SCL	SCLOnsetOffsetDiff	$SCL(t_{\text{mmItemOffset}}) - SCL(t_{\text{mmItemOnset}})$ difference of SCL between the onset and the offset of the stimulus
	OnsetOffsetDiffAv	mean of SCL 2 seconds before the stimulus, during the stimulus, and 3 seconds after the stimulus
SCR	SCRsRelativeNbr	$SCR_{\text{nbr}}/mmItem_{\text{duration}}$ number of SCRs relative to stimulus duration
	timeStart	
	meanRiseTime	
	meanAmplitude	
	meanDuration	
	meanRelativeLatency	if it exist several SCRs for the same mmItem

The table 5.5 presents an example of such SC related features, for a subject.

Table 5.5: Example of SC related features for each mmItem.

mmItem	nberofSCR	SCRsRelativeNbr	time start	meanAmplitude	meanRiseTime	meanDuration
parisCombo.wav	4.0	0,128	2,937	0,062	1,563	7,086
scream1.wav	1.0	0,213	2,969	0,089	2,469	3,531
latchoDrom1.avi	14.0	0,171	2,563	0,098	1,634	3,739
6830.jpg	1.0	0,166	3,156	0,018	2,125	5,781
thegirlformipanema.wav	4.0	0,091	2,750	0,025	1,109	2,352
HTexier.wav	2.0	0,052	4,094	0,052	1,844	3,750
harrysallyVidCinepack_Snduncompcut.avi	23.0	0,154	2,531	0,115	1,489	4,849
ChiDodge1972.wav	5.0	0,173	2,125	0,122	1,812	5,144
GainsbourgReggaeVieilleCanaille.wav	4.0	0,118	3,844	0,014	1,641	2,024
breakcoreManouevre.wav	3.0	0,104	2,187	0,144	1,948	5,490
7490.jpg	1.0	0,167	3,062	0,091	2,000	6,000
adb9.wav	3.0	0,137	6,406	0,087	1,292	4,646
supertrampThelogicalsong.wav	2.0	0,055	6,094	0,018	1,406	1,531
8475.jpg	1.0	0,166	2,656	0,163	1,657	7,313
killbill.wav	3.0	0,090	2,812	0,032	1,844	2,250
7220.jpg	1.0	0,166	3,250	0,012	1,219	1,250
satie.wav	8.0	0,176	2,281	0,225	1,640	4,262
DaveBrubeckQuartet.wav	4.0	0,114	2,688	0,112	2,344	6,430
TheDoorsTheEnd.wav	7.0	0,151	4,157	0,126	1,482	5,826
1441.jpg	1.0	0,166	1,344	0,063	3,063	3,313
JimiHendrix.wav	6.0	0,212	7,906	0,181	1,656	4,276
2152.jpg	1.0	0,166	2,563	0,016	1,219	1,250

5.4.1.3 HRV extraction in time-frequency domain

5.4.1.3.1 Technique. The set of detected Inter heart Beats Intervals (IBI), also called R-R intervals, when measured with an ECG, was stored as an interval tachogram (IT), defined by :

$$IBI(i) = t_i - t_{i-1} \text{ for each } i-th \text{ beat , occurring at time } t_i \tag{5.9}$$

It was stored into a timestamped file, using java classes.

As mentionned We will focus of the Heart Rate Variability (HRV), an interesting source of information.

Actually, the IBI fluctuates due to several reasons. The first is the breathing, due to the fact that cardiac and respiratory systems are related. Healthy individuals exhibits periodic variations in IBI, which is known as the respiratory sinus arrhythmia (RSA). For each respiration cycle, inspiration shorten the IBI (the heart is accelerating), while during expiration, the IBI is increasing (the heart is decelerating). According to [Meste et al., 2005], which give precise method to RSA measurements, "precise patterns concerning the respiratory frequency can be extracted from the heart period series". Practically, as breathing is a slow phenomenon compared to heart rate, it could be simply extracted with a high-pass filtering.

Among other reason, the second reason is the resultant of emotional processes. As mentionned below, the interesting signal in the raw IBI signal is the HRV in the frequency domain, which is related to emotion.

Starting from the IT, HRV could be measured in time and frequency domain, with several methods [Electrophysiology, 1996] and [Clifford, 2002] and [Carvalho et al., 2003] who presents a tool to compute HRV in matlab, with different techniques (STFT, Wavelet, etc...) Different techniques for frequency domain measure of HRV are :

- ≻ **FFT** : It assumes that IT is stationnary, which is not the case. Morevover, as heart beat are irregulary sampled data, a resampling of IT from beat number to time, or other technique, is needed to be able to get the actual meanning of FFT frequency axis, like in [Barbieri et al., 2003]. An adaptation of the FFT approach is proposed in [Castiglioni et al., 2002] to try to encompass the fact that heart beat are irregulary sampled and non stationary data.

- ≻ **STFT** : It is designed for non stationary signals, but still difficult to have instantaneous spectrum; due to the rather
non stationnary nature of the signal. It needs also to find a technique to get the actual meanning of FFT frequency axis. A specific beat-to-beat approach, quite complex but suitable for online HRV measure, is presented in [Castiglioni et al., 2002]

- **Lomb periodogram** : designed to work without resampling. The problem is that the accuracy of this approach is critized (see [Castiglioni and Rienzo, 1996]). [Moody, 1993] and [Mateo and Laguna, 2000] used this techniques.

- **Wavelet** : Not really studied for the moment. Seems to be difficult to set up with short-term computation

- **SPWVD** : This is a method used mainly for accurate instantaneous HRV. It is used in [Kettunen and Keltikangas-Jarvinen, 2001]

- **TRS** : The trigonometric regressive spectral analysis is proposed by [Rüdiger et al., 1999], as another alternative of FFT, which could be suitable for real-time computing

- **ARMAsel** : It is a method proposed by [Broersen, 2000], for the " spectral representation of irregularly sampled data", which is the main problem with heart beat sampling. It is based partly on AutoRegressive (AR) techniques, also used for heart rate (see [Carvalho et al., 2003]). ARMAsel is a method used in [Kim et al., 2004] to estimates the power spectral density of heart rate online.

The method we chose is based on Short Time Fast Fourier Transform (STFT) applied on a floating window of 32 beats, with an overlap of 31 beats, as the new HRV measure is applied every new incoming IBI value. 5.4.1.3.1 presents the method we detail below :

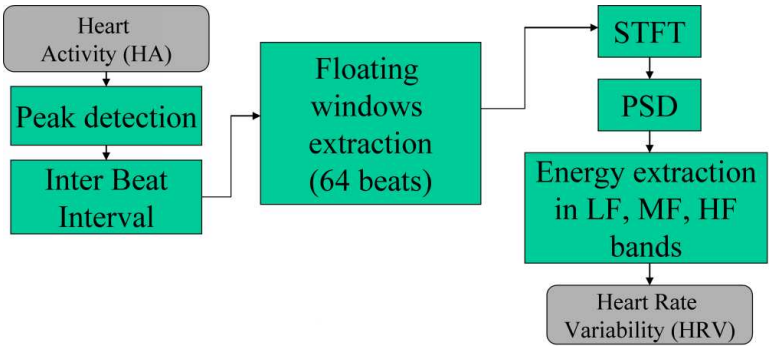

Figure 5.14: Measurement of Heart Rate Variablity in frequency domain.

For each new recorded beat, an N-points FFT is computed on the Inter heart Beat Interval (IBI) values, on a time-window of N beats. Let be $FFT = fft(IBI, N)$ the set of fft values from the IBI. To avoid resampling, which introduces smoothing, the scale of the frequency axis of the FFT is found by averaging the IBI on the N beats time-window. Let be Fs the sampling duration of the signal (in seconds), defined by (5.10):

$$Fs = \frac{\sum_{i=1}^{N} IBI_i}{N} \tag{5.10}$$

Then, one point frequency of FFT ($f(i)$, in Hz), where i is the fft's output index, is given by $f(i) = i * Fs/N$

A Power Spectral Density (PSD) is then computed on each FFT, i.e. on each time-window. It represents the amount of power per unit of frequency, as a function of frequency ([Castiglioni et al., 2005]). It is a useful tool to compute the distribution of IBI variance with frequency. For each $FFT(i)$, the PSD (in s^2/Hz) is computed with (see 5.11):

$$PSD(i) = |FFT(i)|^2 \tag{5.11}$$

Finally, for each time-window, the energy is computed in three bands (Hz): LF [0.01,0.08[, MF [0.08,0.15[and HF[0.15,0.5[which had been found to be related to emotion ([McCraty et al., 1995], see section 3.4.3.3 above-mentioned). Let be E_{LF}, E_{MF} and E_{HF} the energy in these three bands. Let l and h be the low and high limits of a band. to compute the energy in one band, let's consider $PSD(m, n)$ the set of PSD values like $f(m) \geq l$ and $f(m-1) < l$, and like $f(n) < h$ and $f(n+1) \leq h$. Energy is computed as a sum of PSD values within the band. For example, the Low Frequency Energy (E_{LF}) is (see 5.12):

Knowing m, n like $f(m) \geq 0.01, f(m-1) < 0.01$ and $f(n) < 0.08, f(n+1) \leq 0.08$

$$E_{\text{LF}} = \sum_{i=m}^{n} PSD(i) \tag{5.12}$$

Thus, the energy of the PSD is computed for each time window of N beats, in the three bands (see 5.4.1.3.1, for an example with an IBI set of 178 values, with $N = 32$). We denote (HRV(i)) the set of energy values in the three bands.

5.4.1.3.2 Extraction for each stimulus. We extracted HRV using two methods. We first computed HRV with $N = 8$ over the whole set of ibi. The values of N was choosen to ensure that the duration of this set of beats fits every mmItem (even the shortest), for all subjects. Then, for each IBI time window associated to a HRV (i), starting at ($\text{HRV}_{\text{start}}(i)$) and ending at ($\text{HRV}_{\text{end}}(i)$), we made the average of each bands included in $[\text{mmItem}_{\text{start}}, \text{mmItem}_{\text{end}}]$.

5.4.1.4 raw HR, and HRV related features

The calculated features from the HRV signal were (see table 5.6) :

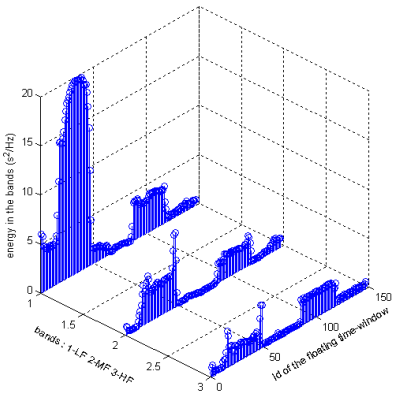

Figure 5.15: Power Spectral Density in LF, MF and HF bands. Each value correspond to a time-window of 32 beats, denoted by the floating time-window id.

Table 5.6: HR-related features calculated for each item.

	HR-related Features by MMItem	Description
HR (raw)	HRAverage HRMax HRMin	
HRV, in each Freq. bands : (with i =LF, MF or HF)	meanE_{i} minE_{i} maxE_{i} meanDerivativeE_{i}	meanDerivative$E_{\text{i}} = \frac{\sum_{n=1}^{N-1} E_{\text{i}+1} - E_{\text{i}}}{N}$, with N the number of E_{i} belonging to the mmItem

5.4.2 Data preprocessing : Psychological features extraction

5.4.2.1 Position in the valence*arousal space processing

The d'n'dMultimedia software output file (*nameOfSubject*_x=Valence_y=Arousal.xml) contains the result of the second step of the experiment. It consist of the raw result of drag'n'drop ((x, y) coordinates of each item in the valence*arousal ranged from $[-1, 1]$ and $[-1, 1]$ respectively) and several pre-processing variables computed directly into the software. Each item is denoted by an *id*, a *name of file*, a *multimedia type* and a *duration*. The calculated variables are :

- ≻ *inter-item distance* between each item position (for valence, arousal and both valence and arousal, which is the Euclidean distance
- ≻ *coordinate of central virtual point*. It is the center of the cloud of points (computed by averaging in each dimension), which might allow a scale change, considering this point as the center of the interface.
- ≻ *number of experiences* the subject had of this item.

We substracted the central virtual point to the coordinate of points, and rescaled theses new coordinates according to min and max of the points set (see figure 5.16). Using raw coordinates permits to respect the position chosen by subject to express their affective state, while using preprocessed coordinates permits to scale all responses into a normalized space.

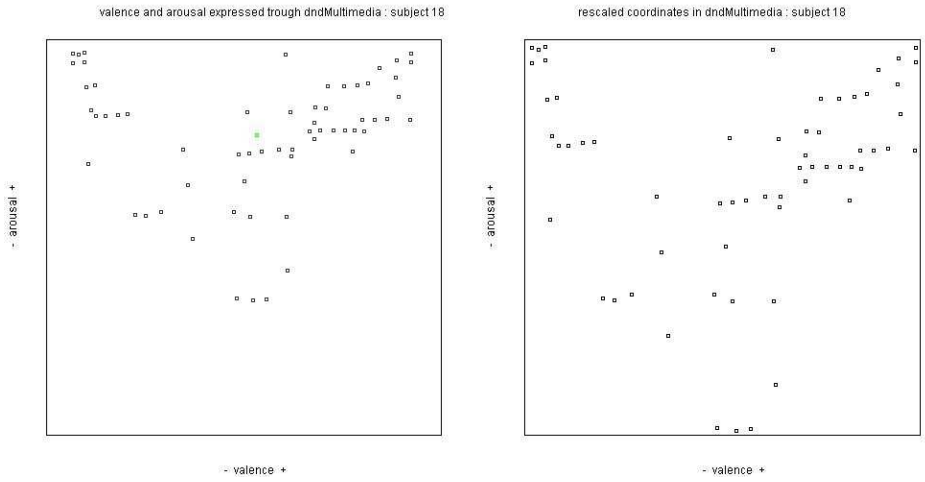

Figure 5.16: Substracted central virtual point (left figure, green dot) and rescaling applied to valence and arousal points. The left figure is the actual response of the subject, and the right figure is the rescaled response.

5.4.2.2 Dimensional and discrete emotion representation conversion and Clustering

5.4.2.2.1 Discrete emotion classes. For each subject, the (x,y) coordinate were estimated as discrete emotion. We designed a java class (EmotionModel.java) implementing a dictionnary of conversion from the discrete-dimensional space of Russel ([Russell, 1980]). We implemented a simple euclidian distance calculation to estimate the discrete emotion belonging of a point in the valence*arousal space. Figure 5.17 is an example of such discrete emotion estimation, from the coordinates of points set expressed by a subject who participated to the study. Each point belong to a discrete emotion, according to its position in (valence,arousal) coordinate, in a normalized space ranged from [-1,1]. The dictionary is thus :

- {(Sad,(-0.3875,-0.5059))
- (Happy,(0.6625,0.1176))
- (Calmness,(0.6,-0.3647))
- (Surprise,(0.1125,0.6))
- (Neutral,(0,0))
- (Fear,(-0.4875,0.5176))
- (Sleepiness,(0.1625,-0.6353))
- (Disgust,(-0.65,-0.2118))
- (Anger,(-0.6125,0.0706))}

To estimate the discrete emotions associated to coordinates, we used the raw coordinates in the valence*arousal space provided to subjects or the rescaled coordinates according to the central virtual point of the cloud of the point set.

5.4.2.2.2 Clustering. Despite convert the coordinate into discrete emotion classes, we also used two clustering techniques. Firslty, we considered five classes regarding the spatial location in the affective space : {(Neutral(N), (0,0); Low Valence and Low Arousal (LvLa), (-0.5,-0.5)); (Low Valence and High Arousal (LvHa),(-0.5,0.5)); (High Valence and Low Arousal (HvHa),(0.5,-0.5)) ; (High Valence and High Arousal (HvHa),(0.5,0.5))}. Then, we used the rescaled coordinates from central virtual point (to normalized space) and computed euclidian distances using the coordinates associated to classes. The figure 5.18 show an example of such qualitative affect segmentation based on spatial location.

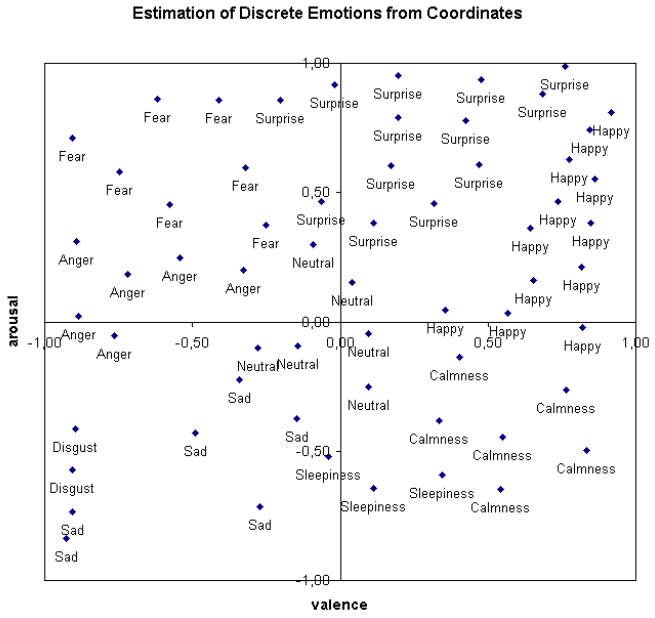

Figure 5.17: Estimation of Discrete Emotion from the valence and arousal coordinates expressed by a subject. Each dot corresponds to the evaluation of a mmItem.

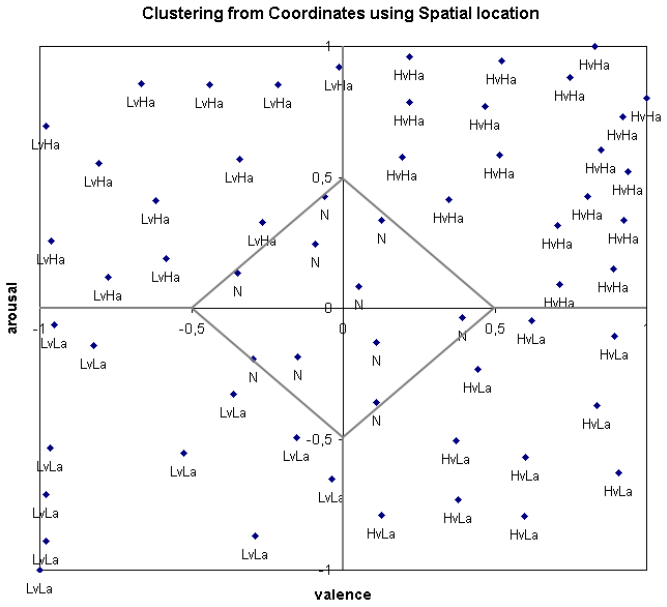

Figure 5.18: Clustering of expressed coordinates by subjects in the valence*arousal space using spatial location. The grey lines constitute the limits of each region.

We also considered two classes by dimension : Low Valence (Lv) versus High Valence (Hv) and Low Arousal (La) versus High Arousal (Ha). Then, we used the rescaled coordinates from central virtual point (to normalized space).

The third clustering technique was kmeans. Kmeans is interesting for this purpose as it can accounts for spatial grouping subjects performed when it evaluates the stimuli. Thus, it could help to find an accurates classes for subjects who used a relative strategy to classify the stimuli in the valence*arousal space. Figure 5.19 plots an examples of such clustering. Detailled plots for each subject could be found in appendix **??**.

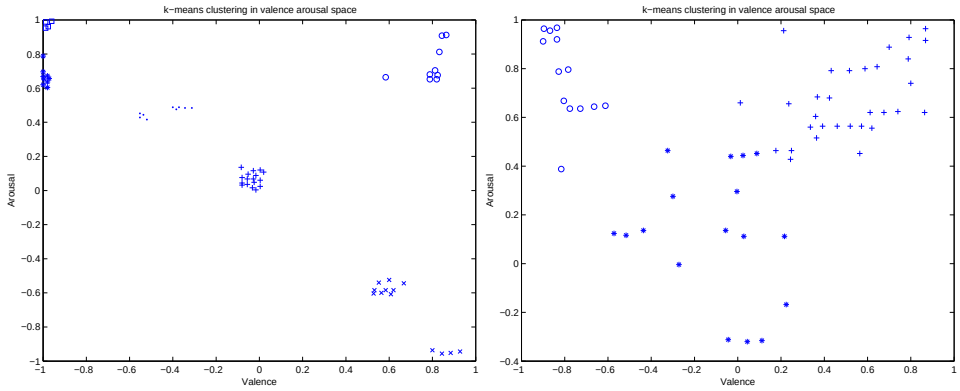

Figure 5.19: Clustering of coordinates into the valence arousal space using k-means.

5.4.2.3 Valence continous measurement

Slider continuous measurements were preprocessed in order to accounts for dynamic (see example of measure in fig. 5.21). For this purpose, the Internal Affective State library was used ([Villon, 2002], see fig. 5.20). This library provides simple features designed to accounts for what the subject intent to express while using a slider as expressive interface. Among theses features, we applied the dynamic of x DX (right derivative) and the quantity of experienced dynamic Q (in additive mode). To get one value associated to each stimuli, the mean of DX was computed. Assuming n, the number of samples, X slider data, and Q the quantity of experienced dynamic (see equation (5.13)).

$$Q = \sum_{i=1}^{n-1} |X(i+1) - X(i)| \tag{5.13}$$

For normalization purposes, we divide Q by the number of samples to obtain $meanQ$.

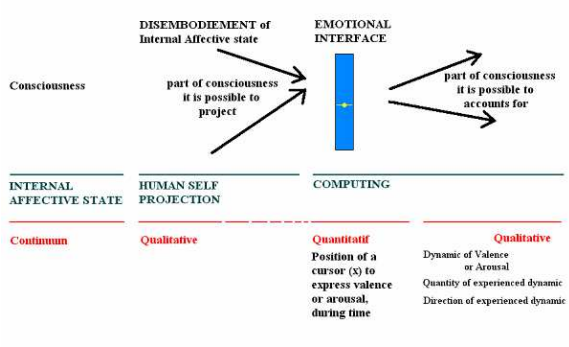

Figure 5.20: Assessment of internal affective state (conscious affective feeling). The bloc dynamic for eyesweb generates outputs that aims at being close to the subject dynamic of affective experience (figure from [Villon, 2002]). We use it to process the valence expression while watching video and listening sounds.

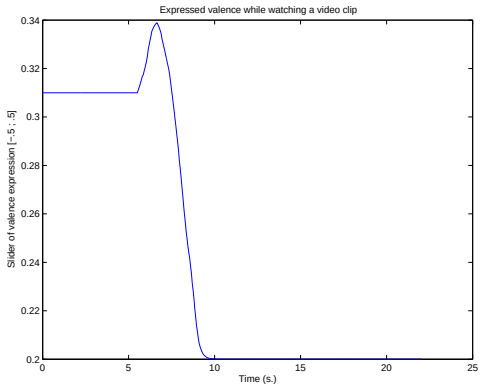

Figure 5.21: Example of valence recording while watching a video clip.

	Panas Features by subject	Description
Personality level (General)	PA_{pers}	average of scale values of Positive Affect adjectives
	NA_{pers}	average of scale values of Negative Affect adjectives
Mood level (Last week)	PA_{mood}	average of scale values of Positive Affect adjectives
	NA_{mood}	average of scale values of Negative Affect adjectives
Mood level modulation of personality level	$\Delta PA \in (pers, mood)$ $\Delta NA \in (pers, mood)$	$= PA_{mood} - PA_{pers}$ $= NA_{mood} - NA_{pers}$

Table 5.7: Panas Features by subject

5.4.2.4 Panas : personality and mood level

We calculated the positive (PA) and negative (NA) indexes of each subjects, for the emotion measure at general and week level. This is the standard features calculated by averaging the 10 positives adjectives and the 10 negatives adjectives of the PANAS. Moreover, we computed a time-domain additional feature, to denotes the changes from the personality level to the mood level (see table 5.7).

We clustered subjects according to PA_{pers} and NA_{pers}. We made 2 classes : (1)high PA,low NA (i.e. positive) (2)low PA,high NA(i.e. negative).

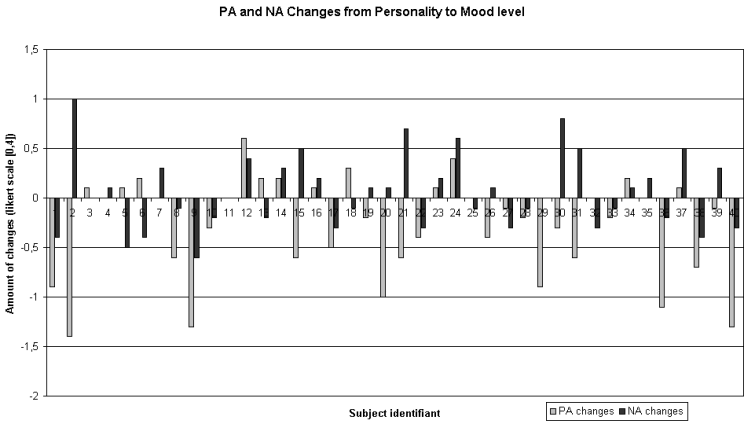

Figure 5.22: Changes from the personality level to the mood level.

5.5 Statistical Results of Experiment

The raw SC signal was exported from InnerView Research software as an excel file, sampled at 0.031Hz. Data situated before the timestamp were removed. The heart beats'IBI were recorded into a text file containing the timestamp of the begining of the experiment (for synchronisation purpose).

A log of each multimedia items experienced by the subject during the slideshow, with the start and end time of each item (the 0 time corresponds to the timestamp, i.e. the end of the countdown)), and with the duration was produced as an xml file by the interface. The position coordinates for each multimedia items in the valence*arousal space were also stored into an xml file, along with some data pre-processing, performed on-the-fly by the software (see section 5.4.2.1). Data produced by the questionnaire were also analyzed.

The approach to analyze data is made of four steps

- ≻ Establish correlations between Psychological and Physiological evaluations of mmItems, for each subject (intra-indiviudal or within-subject statistical analysis)
- ≻ For each subject, get all found correlations and establish the rules (user-modeling)
- ≻ compare this rule to the population average rules

5.5.1 Experiment evaluation

The analysis of the questionnaire regarding the experiment evaluation (4th phase, part 1), lead in a positive evaluation of stimuli (sensation to feel emotion= 73.5%, StdDev= 24.1 ; variation of emotions= 70.5%, StdDev= 21.23). Moreover, the evaluation of the usability gave a result of = 69.5%, StdDev= 22.6.

5.5.2 Summary of Features statistical analysis

Statistical analysis was done using all psychological features versus the physiological ones. The analysis was processed mainly at intra-individual level. By intra-individual level, we mean that we statistically process the psychological and physiological measures of each stimuli (see fig. 5.24.1), for each subject, as pairs (fig. 5.24.2).

The aim of this analysis is to find robust *intra-individual rules* describing precisely the psychophysiological statistical rules of each subject, and allowing the estimation of emotion of individuals according to physiological dynamics. Physiological responses (3 rd person approach to emotion) produced by the subjects and the psychological responses expressed by the same subjects (1 st person approach to emotion) were statistically compared.

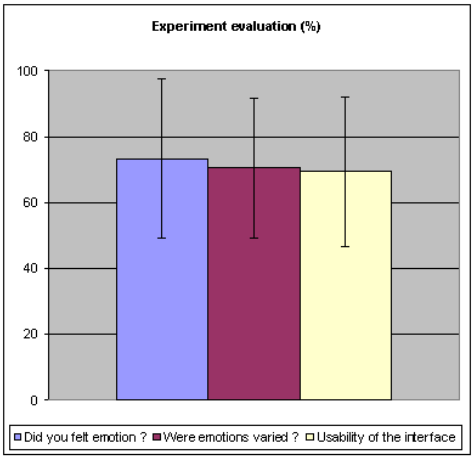

Figure 5.23: Experiment evaluation regarding the stimuli and the usability of interface.

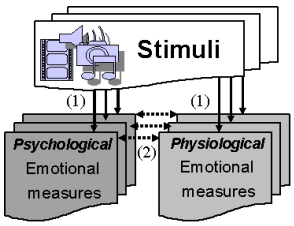

Figure 5.24: Intra-individual approach to data analysis. The psychological and physiological measured expressions of stimuli (1) are analyzed as pairs, for each subject.

Table 5.8 and the flowchart in figure 5.25 summarize the statistical analysis. The statistical analysis of Psychological features (i.e. based on valence and arousal ratings of subjects) versus the Physiological features extracted from Heart Activity and Skin Conductance was performed at intra-individual level.

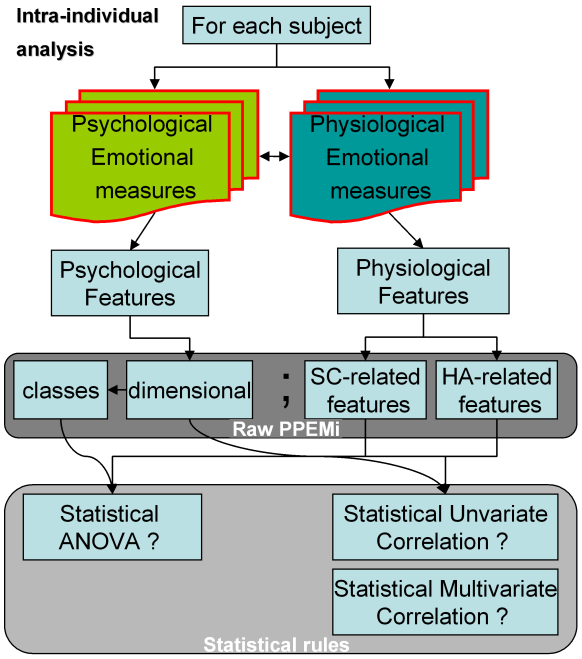

Figure 5.25: Flowchart for statistical analysis of the experiment and PPEM building

Table 5.8: Statistical Analysis (Legend : Regression (reg.) / multiple (mult.) / correlation (corr.))

	Physiology		Psychology		
			Static		Dynamic
	SCRs features	HRV features	Valence	Arousal	Valence features
SCRs features		*Corr.*	Reg. Mult.	Reg. Mult.	Corr. / Reg. Mult.
HRV features			Reg. Mult.	Reg. Mult.	Corr. / Reg. Mult.
Valence stat.				*Reg.*	*Reg.*
Arousal stat.					*Reg.*
Valence dyn					

5.5.3 Correlations

5.5.3.0.1 Valence and arousal versus physiological features. First, for each subject, correlations were calculated between valence, arousal, valence rescaled, and arousal rescaled, and with all the raw SC, SCL and

SCRs related features. Unexpectedly, few significant correlations were found. The correlation between SCRs amplitude and Arousal was not systematically found (no inter-individual similarities regarding the psychophysiological statistial rules). The rescaled coordinates according to the center virtual point gave similar results as the non rescaled one.

This analysis lead in various relations between SC and HA, and the 1 st person categorisation in the valence*arousal space. Figure 5.26 presents the amount of correlations found ($p<0.05$ and $p<0.01$) for each subject, between the whole set physiological features and valence or arousal. Figure presents the number of subjects for which we found a significant linear correlation, for each feature, splitted into heart rate features and skin conductance features.

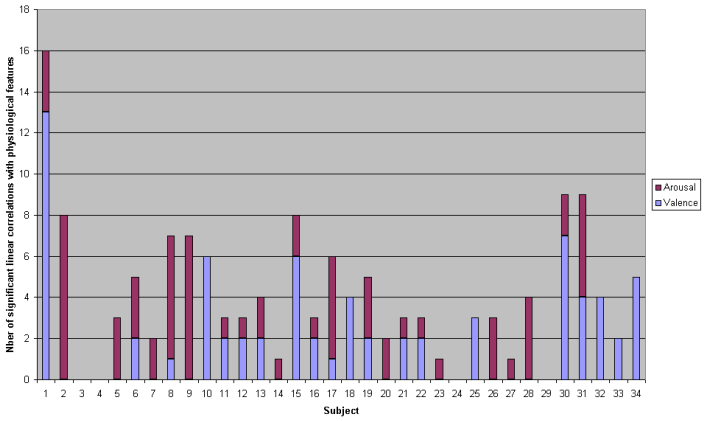

Figure 5.26: Amount of linear significant correlations ($p<0.05$ and $p<0.01$) between the whole set of physiological features and valence or arousal.

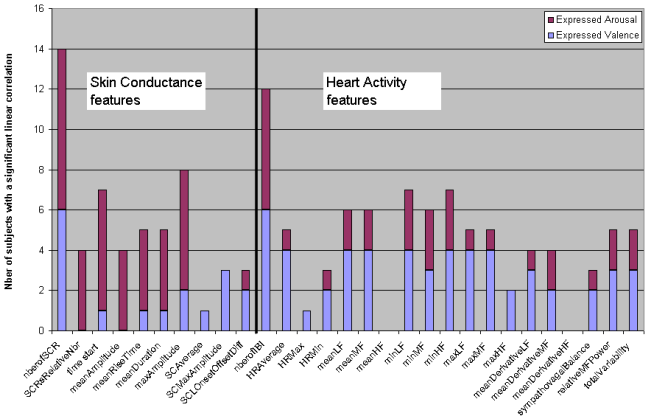

Figure 5.27: Number of subjects for which we found a significant linear correlation, for each feature.

Next figure plots an example of significant correlations (r^2 with $p < 0.05$, blue, and $p < 0.01$, red) between physiological features and psychological values, along with the linear model associated.

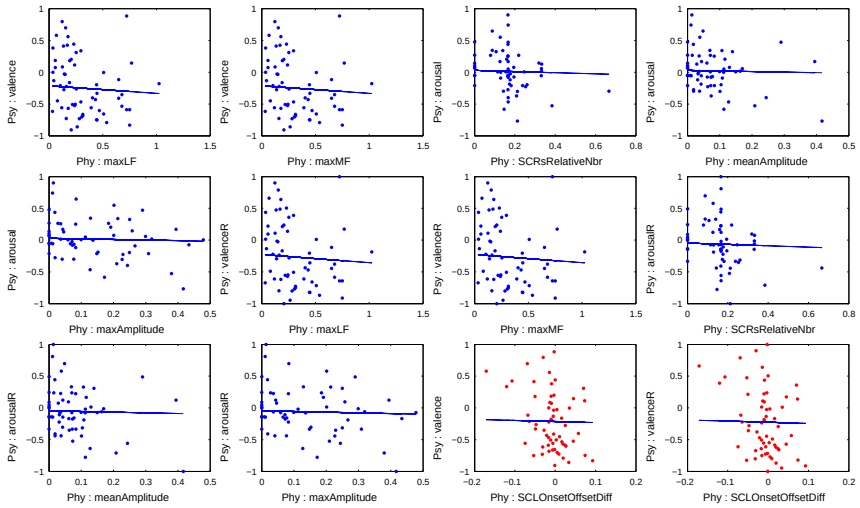

Figure 5.28: Linear correlation between physiological features and psychological values for subject 12.

Each subjects presented a different range of significant correlation, either at $p < 0.01$ or $p < 0.05$, and thus linear relation. For each subject, a mean of 5.35 significant relations over a combination of 100 possible relations were found (min = 0, max =16, average=, stdDev=4.34). The maximum percentage of variance in a psychological variable related to the variation in a physiological variable was, 16 % of variance explained ($r^2 = .161$, max=, min=, stdDev=) for arousal versus meanDerivativeMF. The maximum of the variance explained by a linear relationship was not enough robust to build PPEM_i using linear model to predict the coordinates in the valence arousal space from the physiological features.

5.5.3.0.2 Valence dynamic measure versus physiological features. Regression were performed between features extracted from the valence measurement performed using the slider (see section 5.4.2.3), and with physiological features. Only dynamic stimuli were considered, as slider measurement was requested during the experience of stimuli (video and audio). Several correlations were found for each subject, which indicates relationships between the real-time expression of emotion and physiological parameters, at intra-individual level. Figure **??** presents a summary of the results, plotting the number of significant correlation for each subject between physiological features and dynamic valence features. Figure 5.30 plots an example of cloud of linear relationhsip between (with $r^2 = 0.81, p < 0.01$).

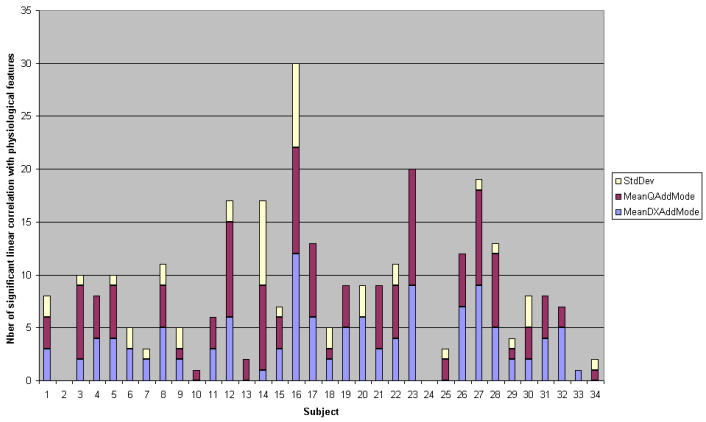

Figure 5.29: Significant correlation between physiological features and dynamic valence features measured using a slider.

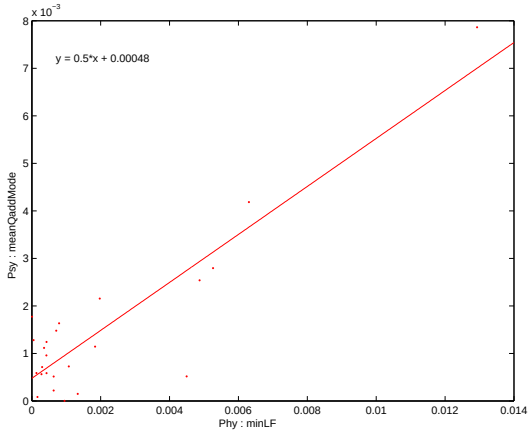

Figure 5.30: Significant correlation between a dynamic measure of valence feature and physiological feature for one subject.

5.5.4 Multiple correlation : linear multiple relation between numeric variables

Multiple correlation coefficients (R) were calculated using the whole set or subset of physiological features as independant variables and either valence or arousal as dependant variables. Table 5.9 presents the squared coefficients R^2, and p values associated (empty cells means that non significant multiple correlation coefficient were found).

5.5.5 ANOVAs : relation between discrete emotion representation and physiological variables

Discrete emotions were analyzed with the set of physiological features with a one-way ANOVA (testing the discrete emotion class effect over the set of physiological features).

5.5.5.1 Intra-individual analysis

Four group of emotions classes were tested : the discrete emotion from Russel circumplex, from rescaled or raw coordinates, the clustering according to the spatial location in the affective space (see section 5.4.2.2), and the clustering using k-means. Tables 5.10, 5.11,5.13,5.13 contains the significance of One-way ANOVAs tests using theses classes, for only first subjects. We display the significance of the result and not the F value. All tests are made with F(1,63), with $p<0.05$ or $p<0.01$.

5.5.5.2 Mixed intra-individual analysis

Then, we performed ANOVAs by mixing all the intra-individual pairs made of emotion classes and physiological features.Thus, each physiological values associated to the emotion class (e.g. 'Happy', or 'LvHa') for each subjects were combined.

This approach respect intra-individual associations between physiological and psychological features and allow to exhibit general trend in the studied population. Three group of emotions classes were tested : the discrete emotion from Russel circumplex, from rescaled or raw coordinates, and the clustering according to the spatial location in the affective space. We do not adopted the same approach for the classes found using k-means, as classes are different for each user and thus not comparable. This lead into interesting several shared statistics (see table 5.14).

Figure 5.31 plots an example of the effect of nearest discrete emotion class and the number of SCRs relative to the duration of the stimulus. Happy class elicits more SCRs compared to the neutral class, while sadness and calmness produce less SCRs.

subject	Rsq V HA	p	MaxErr	Rsq A HA	p	MaxErr	Rsq V SC	p	MaxErr	Rsq A SC	p	MaxErr
1							0.36	**	1.02			
5				0.79	*	1.35				0.45	**	1.40
6	2.87	*	0.72	26.44	**	1.17						
8										0.33	*	0.99
13							0.30	*	1.05	0.38	**	0.97
15										0.36	**	0.99
16				1.61	*	1.18						
17	0.63	*	1.22									
19										0.30	*	1.06
24				0.63	*	1.08						
31	1.02	*	1.07									
37	1.35	*	1.12									
39												
40	3.93	**	0.95	7.53	**	0.61						

Table 5.9: Squared multiple correlation coefficients (R) calculated from physiological features versus valence and arousal.Only significant R-Square are shown, ***** mean that correlation is significant at the 0.05 level ($p < 0.05$) and ****** mean that correlation is significant at the 0.01 level ($p < 0.01$).

ANOVAs-d emo	1	2	3	4	5	6	7	8	9	10	11	12
nberofSCR		**			*				**			
SCRsRelativeNbr							*					
time start					**						*	
meanAmplitude							**		**			
meanRiseTime			*				*		*			
meanDuration									*			
maxAmplitude		*					*		**			
SCAverage				*								
SCMaxAmplitude									*			
SCLOnsetOffsetDiff		**		**								**
nberofIBI	*				*							
HRAverage	*											
HRMax												
HRMin												
meanLF												
meanMF												
meanHF												
minLF												
minMF												
minHF												
maxLF		*										
maxMF		*										
maxHF					*							
meanDerivativeLF			*									
meanDerivativeMF			*									
meanDerivativeHF												
sympathovagalBal								*				
relativeMFPower	*											
totalVariability												

Table 5.10: One-way ANOVAs to test relation between physiological features and nearest discrete emotion class (calculated from raw locations).

ANOVAs-d emo R	1	2	3	4	5	6	7	8	9	10	11	12
nberofSCR		*			*				**			
SCRsRelativeNbr							*					
time start		**			*							
meanAmplitude							*	**	**			
meanRiseTime							*		*			
meanDuration				*					*			
maxAmplitude									**			
SCAverage												
SCMaxAmplitude									*			
SCLOnsetOffsetDiff												**
nberofIBI	*											
HRAverage	*											
HRMax												
HRMin												
meanLF												
meanMF								*				
meanHF												
minLF												
minMF												
minHF			*									
maxLF		**										
maxMF		**										
maxHF					*							
meanDerivativeLF						**	**					
meanDerivativeMF						**	**					
meanDerivativeHF												
sympathovagalBal												
relativeMFPower	*							*				
totalVariability												

Table 5.11: One-way ANOVAs to test relation between physiological features and nearest discrete emotion class (calculated from rescaled locations).

ANOVAs-qualAff	1	2	3	4	5	6	7	8	9	10	11	12
nberofSCR					*				**			
SCRsRelativeNbr									*			
time start												
meanAmplitude									**			
meanRiseTime									**			
meanDuration									*			
maxAmplitude									**			
SCAverage												
SCMaxAmplitude									*			
SCLOnsetOffsetDiff												**
nberofIBI	*					*						
HRAverage												
HRMax												
HRMin			*									
meanLF		*						**				
meanMF		*						**				
meanHF								**				
minLF												
minMF												
minHF												
maxLF		**										
maxMF		**										
maxHF												
meanDerivativeLF			*					*				
meanDerivativeMF			*					*				
meanDerivativeHF								*				
sympathovagalBal												
relativeMFPower												
totalVariability		*						**				

Table 5.12: One-way ANOVAs to test relation between physiological features and qualitative affect classes.

ANOVAs-kmeans	1	2	3	4	5	6	7	8	9	10	11	12
nberofSCR					**				**			
SCRsRelativeNbr									*			
time start												
meanAmplitude									**			
meanRiseTime					*	*			**			
meanDuration						*	*		*			
maxAmplitude					*				**			
SCAverage												
SCMaxAmplitude									*			
SCLOnsetOffsetDiff												
nberofIBI	*					*						
HRAverage												
HRMax												
HRMin												
meanLF		*				*						
meanMF		*				*						
meanHF						*						
minLF												
minMF												
minHF	*							**				
maxLF	*	**										
maxMF	*	**										
maxHF												
meanDerivativeLF												
meanDerivativeMF												
meanDerivativeHF												
sympathovagalBal												
relativeMFPower												
totalVariability		*				*						

Table 5.13: One-way ANOVAs to test relation between physiological features and k-means based classes.

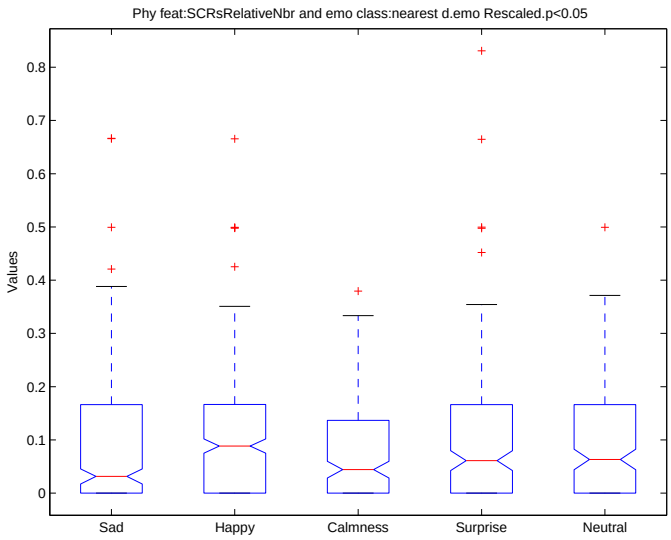

Figure 5.31: Nearest discrete emotion classe effect on SCRs relative number.

ANOVAs mixed data	nearestdemo	nearestdemo R	qualAff
nberofSCR	**	**	**
SCRsRelativeNbr		*	
time start			*
meanAmplitude			
meanRiseTime			
meanDuration			
maxAmplitude			
SCAverage	**	**	
SCMaxAmplitude	**	**	
SCLOnsetOffsetDiff			
nberofIBI	**	**	**
HRAverage	**	**	**
HRMax	**	**	**
HRMin	**	**	**
meanLF	**	**	**
meanMF	**	**	**
meanHF	**	**	**
minLF			
minMF			
minHF	*	**	**
maxLF	**	**	*
maxMF	**	**	*
maxHF	**	**	
meanDerivativeLF			
meanDerivativeMF			
meanDerivativeHF			
sympathovagalBalance	**	**	**
relativeMFPower	*		
totalVariability	**	**	**

Table 5.14: Significant One-Way ANOVAs considering the discrete emotion classes effect on Physiological features, using psycho-physiological pairs of all subjects. * mean that the effect is significant at the 0.05 level ($p < 0.05$) and ** mean that correlation is significant at the 0.01 level ($p < 0.01$).

5.6 Affective State and Emotion prediction from physiological signals

5.6.1 Estimating possibilities of prediction

In this section, we focus on two different approaches toward emotion recognition : (1) the *user-dependency* of psycho-physiological data collected (i.e. do we choose to keep track of the specificity of individuals' responses or do we ignore such specificity), and (2) the *degree of subjectivity* of the stimuli used to elicit emotions (i.e. stimuli with high level of agreement in terms of what emotional experience they elicit among a population can be chosen versus stimuli without such an agreement).

5.6.1.1 Results : Empirically Comparing inter-individual differences of psychological and physiological responses to stimuli

When adopting a set of stimuli with a social agreement, it is a mean to control the certainty of the emotion elicited (as subjective expression of subject might be misleading), and usually meant to ensure a rather uniform physiological response among a population. We can then consider the following problematic : Is the fact that individuals agree or disagree about the subjective emotion elicited by a stimulus (i.e. several individuals disagree about the pleasure/displeasure and/or the arousal elicited by a stimulus (fig. 5.32 (1) σ_Ψ) is an indicator of the uniformity of the measured physiological responses accross a population (fig. 5.32 (2) σ_Φ) ? We investigate the influence of social agreement of stimuli on physiological response similarities by the following hypothesis 1:

Hypothesis 1 *The fact that individuals have a social agreement about the subjective emotion elicited by a stimulus (σ_Ψ) is not an indicator of the uniformity of the measured physiological responses across a population (σ_Φ)*

For each stimulus, and for each subjects, the set of 28 physiological features was used from heart rate and skin conductance.

To estimate the hypothesis 1, we first computed the standard deviation for each stimulus of the set of psychological coordinates expressed by subjects and the standard deviation of the set of features associated to physiological responses measured on each subject. Figure 5.33 plots the physiological inter-individual differences as a function of the psychological inter-individual differences (valence on left part and arousal on right one). The x axis values correspond to standard deviation of the valence (left), or arousal (right) computed on the 40 subjects, for each of the 61 stimuli rated. On the y axis are the values of standard deviation of each physiological features values for

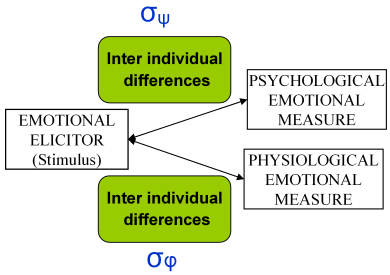

Figure 5.32: Comparison of inter-individual differences in the psychological and physiological evaluation of stimuli.

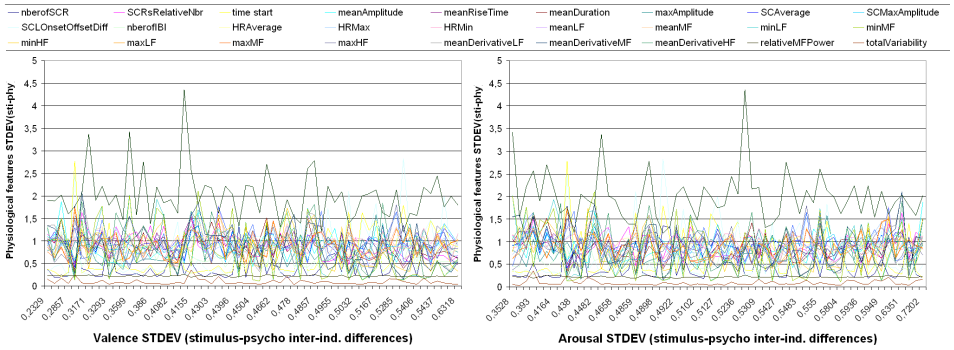

Figure 5.33: Relation between psychological and physiological standard deviations.

the associated stimuli. Because no correlation was found to be significant between thoses variables, the hypothesis 1 is confirmed.

This means that adopting a social agreement approach for the stimuli *does not guarantee* a more uniform physiological response across the population nor a more robust emotion recognition *outside the context of the experiment.*

5.6.1.2 Empirically estimating the effect of the Social Agreement versus Subjective Rating approaches on emotion recognition

To analyse the modeling posibilities of psychophysiological representations, we first averaged psychological and physiological measures of affective states elicited by the stimuli. Then, we selected psychophysiological representations of stimuli with a psychological index of dispersion around the mean of different amounts. For each amount, we selected psychophysiological representations as belonging to the data series on which we test correlation if : $2\sigma_i < \text{amount} * (x_{max} - x_{min})$, with $x_{max} - x_{min} = 2$ in the valence arousal

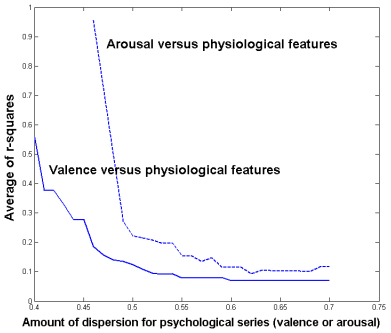

Figure 5.34: Effect of Psychological evaluation dispersion on significant correlation averages between

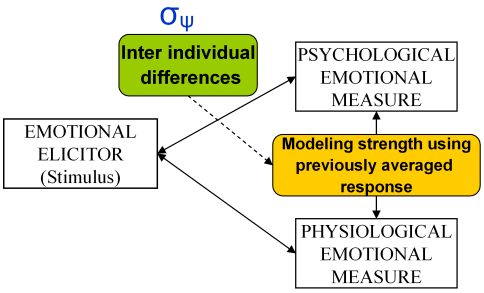

Figure 5.35: Effect of psychological dispersion on psychophysiological modeling strength.

space made of $[-1, 1]$. In previous studies (see section 2.4.3) based on social agreement, dispersion is necessarily low because pilot studies select stimuli with high level of *psychological* agreement among subject. We propose the following hypothesis, related to the possibility of using the results of the social agreement stimuli in other contexts :

Hypothesis 2 *The level of agreement of individuals about the subjective emotion elicited by a stimulus (σ_Ψ) is related to the possibilities of modeling using a user-independent approach*

Our result (figure 5.34)shows that the average of linear statistical test (r-square) is related to the dispersion (psychophysiological correlation for the population approach to 1 for stimuli with low dispersion, i.e. strong agreement). This means that for a low agreement (i.e. stimuli which elicit different subjective experiences for different subjects) the nature of the data will lead to a lesser performance in emotion recognition by using an averaged and normative approach. Thus, selecting subsets of psychophysiological representations associated to stimuli according to the agreement (inter-individual psychological differences) has effect on the modeling possibilities (see figure 5.35). The hypothesis 2 is thus confirmed.

5.6.2 Approach for prediction

5.6.2.1 Data and Choice of classifiers

After having described the statistical significant relationship between psychological and physiological data, and the possibilties of prediction according to the chosen methodology (user-depedency and subjectivity of stimuli) we describe in this section tests performed with machine learning techniques.

As shown in the figure 5.37, we tested different prediction techniques, using different data organization and different methodologies to evaluate

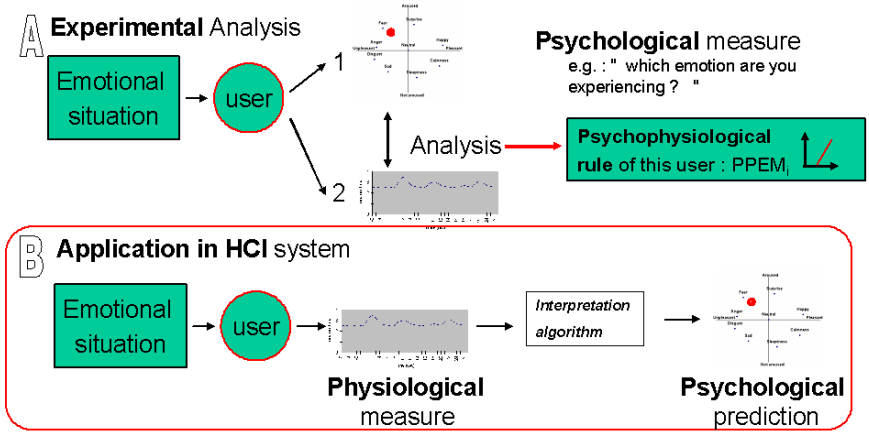

Figure 5.36: Application of the emotion estimation from physiological signal in HCI.

our model. We worked more on the methodology and data preprocessing rather than testing different machine learning algorithm as such approach has been recently largely tested (as previously stated several recent approach proposed machine learning for emotion recognition from physiology : [Changchun et al., 2005], [Wagner et al., 2005], [Lin and Hauptmann, 2006]).

Techniques :

- Categories
  - KNN: K-Nearest Neighbors (with different neighbors)
  - DA: Discriminant Analysis
- Values
  - Multivariate Modeling

Data organization:

- Averaged data: we normalized and averaged data of all subject, for each stimuli.
- Mixed data: we normalized the data and then mixed data of all subjects, sorted by stimuli.
- Intra-individual data: we applied machine learning on each indiviudal separately.
- Averaged data with psychological subjective estimation kept: we normalized and averaged physiological data of all subject, for each stimuli, by each classes defined by each subject.

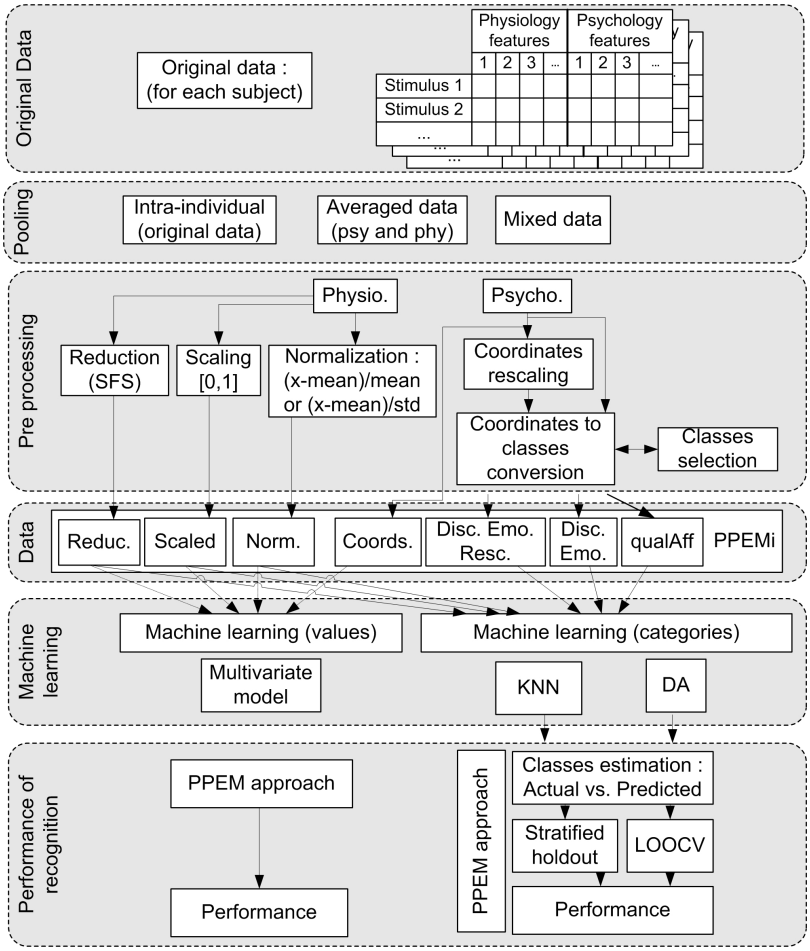

Figure 5.37: Machine learning process for the affective state prediction testing.

Methodologies:

- ≻ Dimensional: we aimed at estimating continuous values representing the position in the valence-arousal space directly from the machine learning.
- ≻ Discrete: we aimed at direclty predicting classes from the data, by using discrete machine leanring techniques

5.6.2.2 Learning set, test set and error estimation

We firstly computed for each classifier the apparent error rate (i.e. the error rate of the trained classifier on the training set). However, this rate is only a measure of the ability of the classifier to fit the data, and not an evaluation of its capability to classify untrained data ([Efron, 1986]). Therefore we used instead errors estimation procedures on untrained data. We use the Leave-One-Out Cross validation (for which the error rate is a mean of the error estimation of each classification produced, see [Elisseeff and Pontil, 2003]), random division of test set, and stratified holdout. Stratified holdout ensure that the training and test sets contains each target classes (by splitting learning and test base with respect to the distribution of the classes). At the opposite random division does not provide a similar distribution in learning and test sets.

5.6.3 Averaged data prediction results

. First approach consisted on averaging both the psychological and physiological measures associated to each stimuli, for each stimuli. We thus considered a set constituting the psychophysiological map, made of 61 element , $\{(x_j, y_j), S(j)\}$, in a PPEMi, with i an averaged subject.

To perform classification we converted the coordinates into discrete emotion, discrete emotions on rescaled coordinates or qualitative affect (see 5.4, $\{((x_j, y_j), C_j^c), S(j)\}$). Then, we used theses classes as the target classes for the prediction.

This first approach led into missing predicted cases as the distribution of classes was not the same in the learning and test sets, due to the specific experimental set-up (the psychological evaluation is performed by subject themselves and therefore we can know in advance the distribution of the classes). We thus selected the elements for learning and test with all the cases, following stratified holdout approach. Figures 5.38, 5.39 and 5.40 plot the distribution of the classes nearest discrete emotions, nearest discrete emotions from rescaled coordinates (for both, classes are assigned as following : Sad,1; Happy,2; Calmness,3; Surprise,4; Neutral,5; Fear,6; Sleepiness,7; Disgust,8; Anger,9), and qualitative affect(classes are assigned as following(N,1; HvLa,2; HvHa,3; LvLa,4; LvHa,5).

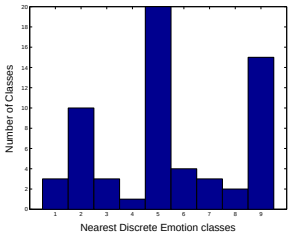

Figure 5.38: Distribution of discrete emotions classes from averaged psychological coordinates.

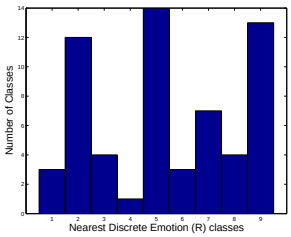

Figure 5.39: Distribution of discrete emotions classes from averaged psychological rescaled coordinates.

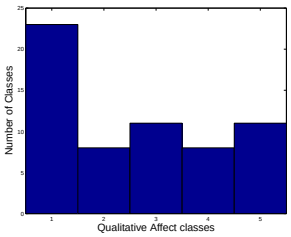

Figure 5.40: Distribution of qualitative affect classes from averaged psychological coordinates.

We firslty selected the classes with the maximum number of cases. The selection was Happy,2; Neutral,5; Anger,9 for discrete emotions and N,1; HvHa,3; LvHa,5 for qualitative affect. Moreover we distributed each class instances according to the ratio used to divide the cases set into learning and test sets (i.e. startified holdout). Figures 5.41 plots the results.

5.6.4 Prediction at intra individual level

Each pairs of psychological features associated to an emotional representation were considered as an element of the $PPEM_i$, associated to each subject i. We use here the formalism proposed into section 5.2.3 . We considered S the group of specific physiological patterns, represented as sets of 29 features (10 SC-related features, and 19 HR-related features) values derived from the physiological signal. 61 elements of each $PPEM_i$ could be considered, as 61 psycho-physiological emotional expression associated to stimuli were measured. Thus, $S^f_{n_f}(j)$,$j = 1, \ldots, 61$. We used a constant number of features for each elements stored in each the $PPEM_i$. Thus, each set of physiological feature by element is defined $S^f_{n_f}$, $f = 1, \ldots, 29$. As only one value is used within each feature, $n_f = 1$.

The psychological part of an element was not here considered as an exact coordinate, but as an emotion classes representation, based on the coordinates. Thus, we considered C^c_j as the converted coordinate of the (x_j, y_j) psychological part of the element j (see section 5.4.2.2 about conversions of valence arousal coordinates into emotion classes) into the class c. Nearest d.emo (C^1) and nearest d.emo Rescaled (C^1) classes are constituted by 7 classes instances : {Sad, Happy, Calmness, Surprise, Neutral, Fear, Sleepiness, Disgust, Anger}. QualAff (C^3) class is consituted by 5 instances : { LvHa,LvLa,HvHa,HvLa,N }. KmeansAff class (C^4) is not constituted by a fixed number of instances as it depends on the output of the kmeans algo-

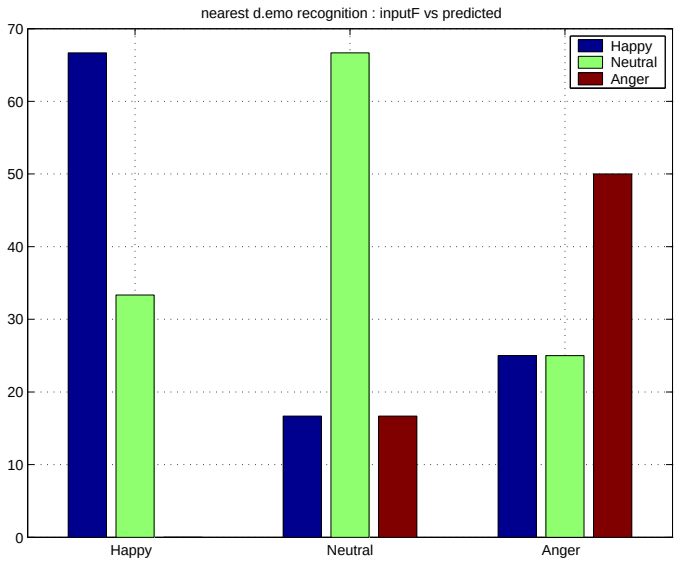

Figure 5.41: K-NN prediction for discrete emotions (K=1), using averaged data. The learning and test bases were built to have the same distribution of classes instances.

rithm (see section 5.4.2.2 for details).

To test prediction using an instance-based $PPEM_i$, we considered 75% of the elements (i.e. 45 elements, see (5.14)). Modifying the learning/testing quantity of data didn't changed significantly the prediction results.

$$PPEM_i = \left\{((x_j, y_j), C_j^c), S(j)\right\} \quad j = 1, \ldots, 45 \tag{5.14}$$

Prediction of affective representation were performed on the basis of physiological features $S(j)$. We estimate the recognition rate using all physiological features and used random division of earning and test sets.

Using all the features the maximum of recognition for each class is rather low. Only kmeansAff(C^4) gives a maximum value of 62,50%. Other maximums are below this value

5.6.5 Results optimization : Intra-individual, Mixed and Averaged approach

We optimized the recognition and compared different performances related to the involvment of the concept of inter-indiviudal differences. We first adopted the optimized machine learning techniques. We first selected the features using a Sequential Forward features Selection (SFS). This technique is a wrapper for classifiers, i.e. it select the features to get the best result

possible by the classifier. We then reducted the features using a Fisher projection which is an efficient reduction technique. To obtain more efficient target classes, we simplified it by converting coordinates into smpler qualitative affect. We divided valence in valence high and valence low, and arousal in arousal high and arousal low.

We obtained good results using such optimized approach. We first tested the $PPEM_i$ recognition, i.e. the intra-individual level. We thus build one classifier by subject. The recognition rates for valence only (high or low) using heart features was 76.15% (average of all recogition rates of subject, 7% of standard deviation. The recognition rates for arousal only (high or low) using skin conductance features was 67.17% (average of all recognition rates of subject, 5% of standard deviation. With such an optimized approach, we compared the methodological approaches. We built a mixed database by assembling all the psychological and physiological data into one large training and test set. The number of element was of 2013 cases (33 selected subjects with 61 elements). We built the averaged database by considering a $\text{PPEM}_{\text{average}}$ build with the simplified classes for the physiological part.

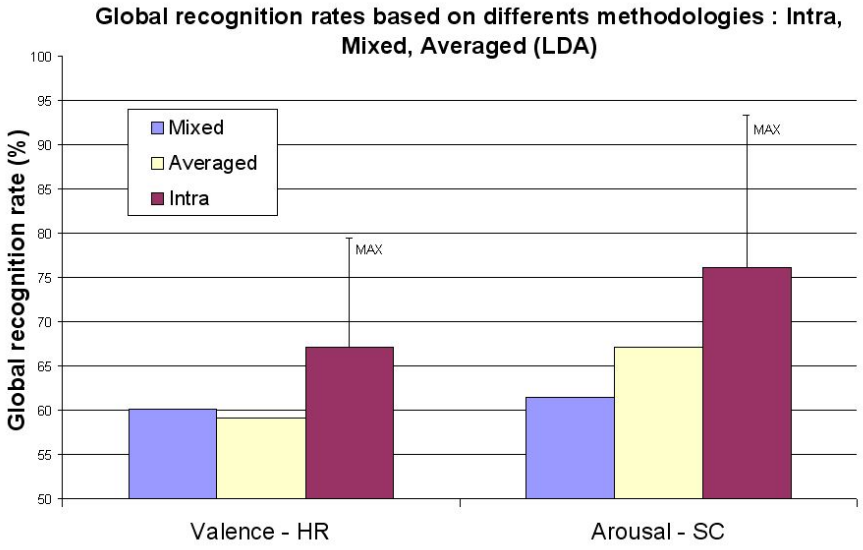

Figure 5.42: Recognition rates using optimized machine learning technique (LDA, with SFS and Fisher reduction), showing the effect of the methodology : intra-indiviudal level, Mixed population, or Averaged population.

A presented in the figures 5.43 and 5.43 we found that the intra inividual approach with $PPEM_i$ is better than the $\text{PPEM}_{\text{average}}$ with an averaged approach. The mixed approach is near the averaged approach, but often above. The best recognition rates obtained using $PPEM_i$ is 93% for arousal simplified classes on the basis of skin conductance related features.

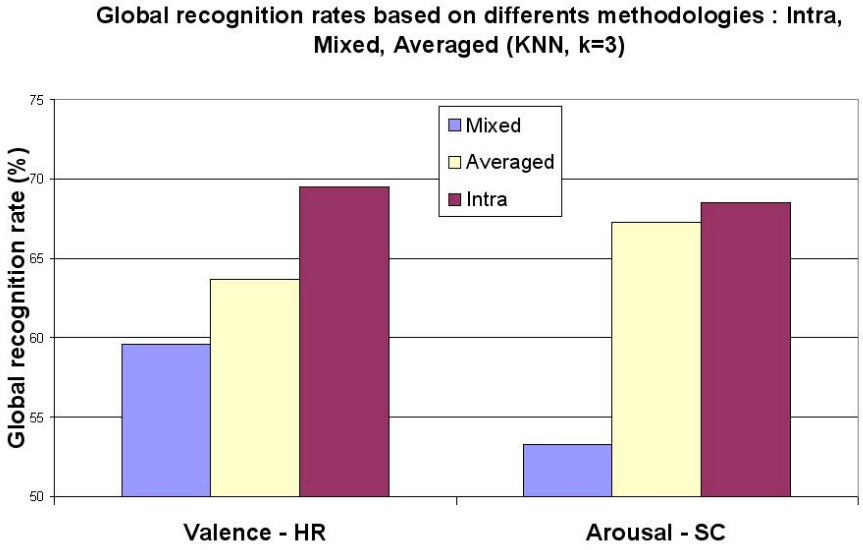

Figure 5.43: Recognition rates using optimized machine learning technique (KNN, with SFS and Fisher reduction), showing the effect of the methodology : intra-indiviudal level, Mixed population, or Averaged population.

5.6.6 Prediction PPEM$_i$ using multilinear approach

We tested possibilities of prediction at intra-individual level using multilinear regression model, based on least squares method, to obtain direclty coordinate and not discrete classes. An example of prediction for arousal is presented in figure 5.44.The interest of such model is to provide a set of coefficient associated to each input feature. This allow to compare model among subjects, and build a different PPEMi for each user on the basis of this coefficient. For each subject, the form of the multivariate model is (see (5.15), for an example with 2 features estimating valence or arousal) :

$$y = a_0 + a_1 x_1 + a_2 x_2 \tag{5.15}$$

In the equation (5.15) y is the valence or arousal to estimate ; x_1 and x_2 are two features values ; a_0, a_1 and a_2 are the associated coefficient which stands for the model of a subject. Differences between the coefficients provides differences between the psychophysiological mappings of physiological features onto psychological representation of emotion between subjects.

Table 5.16 provides the valence estimation coefficients for each subject, for each *HA* features. Each number of feature in the column is associated to features name (see table 5.15).

Column	Feature name
1	nberoflBI
2	HRAverage
3	HRMax
4	HRMin
5	meanLF
6	meanMF
7	meanHF
8	minLF
9	minMF
10	minHF
11	maxLF
12	maxMF
13	maxHF
14	meanDerivativeLF
15	meanDerivativeMF

Table 5.15: Heart rate features name.

Subj.	1	2	3	4	5 height1
2.7	3.0	-71.6	45.9	1.8	
2	-0.3	-0.3	60.6	-64.6	5.6
3	-0.5	2.4	49.1	-68.9	14.2
4	-0.5	1.6	9.4	-1.9	-3.5
5	0.5	-4.6	10.6	16.5	-22.3
6	3.3	1.5	-49.4	15.9	10.9
7	4.1	0.6	-38.6	-5.7	11.3
8	-0.6	-1.7	-6.8	7.5	1.3
9	-10.0	-1.2	52.6	10.4	14.7
10	-0.7	0.5	103.5	-78.8	-19.5
11	0.2	1.4	85.5	-67.1	-22.8
12	1.0	0.7	-10.6	2.8	-0.5
13	-3.8	4.4	69.8	-41.1	-5.2
14	0.1	0.4	24.9	-20.8	-5.3
15	-3.5	-0.3	-43.8	67.6	8.1
16	-0.5	1.3	-35.8	18.8	16.6
17	-3.0	1.0	27.6	-14.9	9.7
18	0.1	1.0	-8.0	1.1	9.1
19	-2.7	5.7	-19.9	11.9	26.3
22	4.1	-2.5	-18.9	-10.4	-3.3
24	-7.4	1.7	25.2	17.0	16.0
25	-2.0	2.2	-11.2	37.1	-13.0
26	1.6	0.6	11.3	-27.2	0.3
29	-3.6	2.2	37.3	0.3	-10.3
30	-3.2	1.2	28.4	-1.9	-3.1
31	1.7	3.6	-50.9	9.4	24.6
32	6.2	0.7	-93.6	46.5	0.8
33	1.3	2.6	-25.9	6.7	8.2
35	2.2	0.0	109.6	-115.2	-11.9
36	-1.2	1.9	14.9	-3.9	-6.1
37	1.4	1.7	-8.6	-15.6	8.5
39	-0.7	1.3	6.4	3.4	-5.3
40	1.3	0.1	-29.7	27.5	-8.2

Table 5.16: Coefficient of multilinear regression model to predict Valence from the set of heart rate related features. Each line represents the set of coefficients for each subject. Each column is a feature (see table for 5.15 key of name

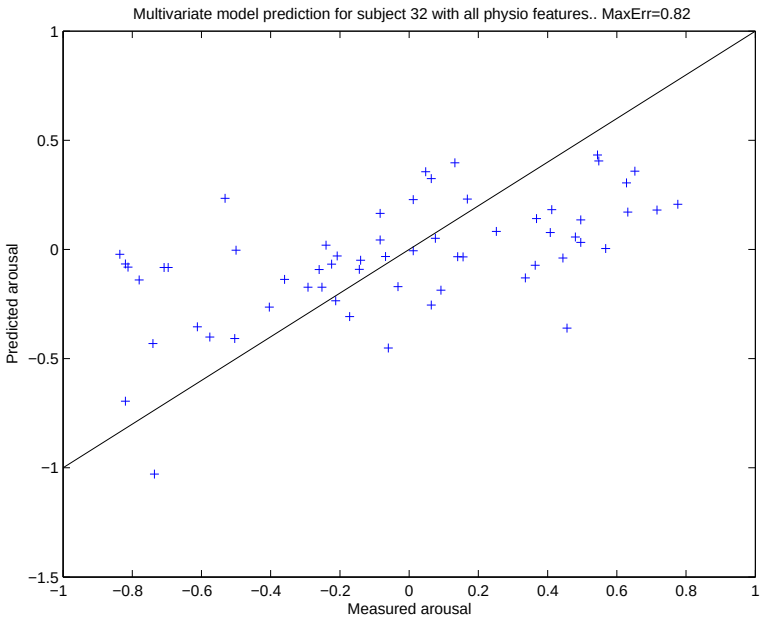

Figure 5.44: Example of prediction of arousal using all the physiological features for a subject.

5.6.7 PPEM average : from this study and from litterature results

We built a PPEM average using the results of this study using both method. First, we built a PPEM direclty from the averaged results of our study, to represent the $\text{PPEM}_{\text{average}}$ of the studied population. Secondly, we tuned the $\text{PPEM}_{\text{average}}$ from the litterature, by assigning values to those which remained to specify empirically (see section 5.2.4).

5.6.7.1 Building $\text{PPEM}_{\text{average}}$ without comparison to the litterature, for the studied population

Figure 5.45 provides the flowchart for the construction of PPEMaverage, for the population we studied. The method is quite similar to the intra-individual ANOVA performed by grouping all physiological features for each group of emotion class instance.

We built PPEMaverage using this method (see e.g. table 5.17 which contain the PPEMaverage for the studied population, using QualAff as classes of references). For each average value, the coefficient of variation is expressed (standard deviation expressed in percentage of the mean). This coefficient gives an indication of the inter-individual differences (the more the coefficient is high, the more the inter individual difference is high). However, as the method is based on intra-individual analysis, this variation already take into

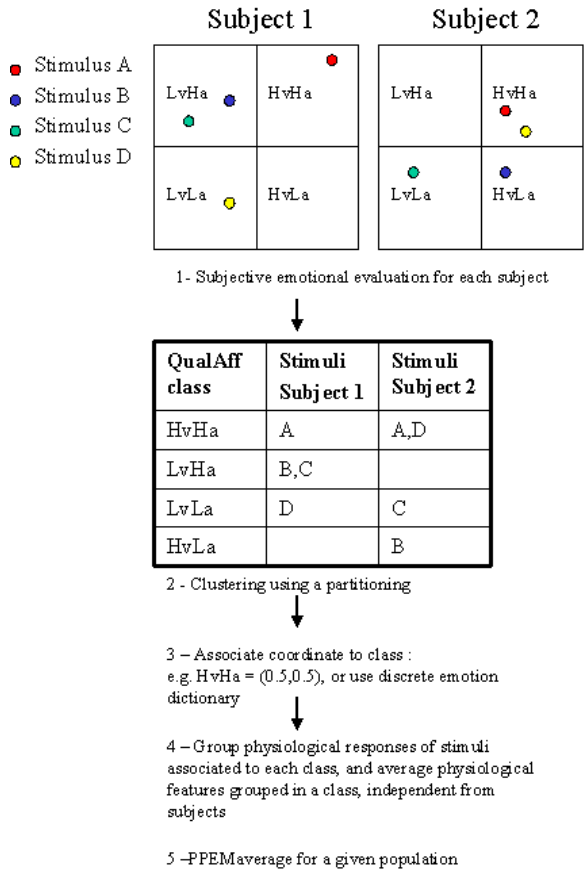

Figure 5.45: Proposed method to build PPEMaverage for the studied population.

accounts the inter-individual differences in the experience of same stimuli.

5.6.7.2 Merging $PPEM_{average}$ based on discrete and dimensional representation

We combine the designed $PPEM_{average}$ built from litterature (see section 5.2.4), by using the discrete-dimensional dictionary built in section 5.4.2.2. Any discrete emotion C_j^c was associated to its coordinate defined by $((x_j, y_j), C_j^c)$. This approach lead into a unique $PPEM_{average}$, using both dimensional and discrete emotional representations (see table 5.18).

5.7 Software engineering for real time emotional features sensing

We implemented the emotional feature extraction using heart activity and skin conductance for a (near to) real time detection. to extract emotional features is described in this section. An API was set up to allow programmers to perform emotion recognition. Moreover, each class could be considered as a stand-alone applications (e.g. Heart Rate Sensor, Heart Rate Analyzer).

The algorithms for SCRs and HRV detection were implemented in Java 1.5.0. The algorithms works in pseudo real time, i.e. with a short-term recognition. For the HRV a minimum situated around 10 seconds of delay is needed depending on the heart floating window size chosen. For the SCRs an event is produced at the end of the exponential decay of the SCRs (in general aroud 2-3 seconds after the events responsible for the SCR elicitation).

5.7.1 Design of HRV real time recognition

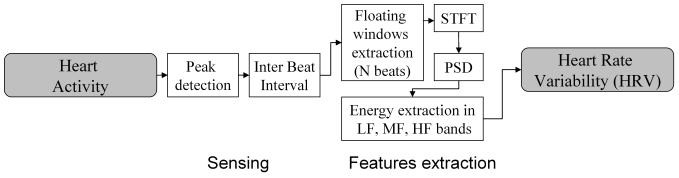

Figure 5.46: Measurement of Heart Rate Variablity in frequency domain.

The HRV algorithm used for offline recognition (see figure 5.46 for a summary) was used directly for the online computation. As the minimum IBI is superior to 0.3 seconds (i.e. inferior to 200 beats.minutes-1), the duration to compute the HRV on a floating window of IBI should be inferior to 0.3 seconds(for an overlap of 1 IBI). If the computation exceed 0.3 second an error is produced and the program is stopped. The algorithm was tested successfully on a Pentium 4 , 2 GHz, 523 Mb of RAM, using java version 1.5.0.

PPEMav. QualAff	**N**	c.var.	**HvLa**	c.var.	**HvHa**	c.var.	**LvLa**	c.var.	**LvHa**	c.var.	Average	c.var.
nberofSCR	2.23	166	2.00	186	2.81	173	1.63	159	1.40	160	2.03	178
SCRsRelativeNbr	0.10	109	0.09	110	0.10	114	0.09	116	0.10	119	0.10	115
time start	3.43	178	2.98	194	4.24	267	2.81	185	2.77	270	3.28	240
meanAmplitude	0.02	178	0.02	179	0.02	156	0.02	190	0.02	146	0.02	171
meanRiseTime	0.99	91	0.97	94	1.01	89	0.94	101	0.98	103	0.98	96
meanDuration	1.64	107	1.74	112	1.67	106	1.55	114	1.68	119	1.65	112
maxAmplitude	0.04	172	0.03	219	0.03	161	0.03	195	0.03	152	0.03	181
SCAverage	1.03	83	0.92	72	1.00	75	0.93	69	0.98	73	0.97	74
SCMaxAmplitude	1.07	81	0.96	70	1.04	74	0.96	69	1.01	72	1.00	73
SCLOnsetOffsetDiff	-0.00	877	-0.01	716	0.00	8574	-0.00	1686	-0.00	1366	-0.00	1690
nberofIBI	61.03	67	66.39	72	73.65	73	59.72	66	59.22	70	64.56	71
HRAverage	0.79	11	0.75	12	0.76	12	0.76	13	0.77	11	0.76	12
HRMax	0.94	27	0.84	17	0.89	28	0.87	22	0.88	19	0.88	23
HRMin	0.79	14	0.74	13	0.76	14	0.76	15	0.77	12	0.76	14
meanLF	0.05	318	0.02	137	0.04	328	0.03	166	0.03	195	0.03	279
meanMF	0.05	322	0.02	142	0.04	334	0.03	170	0.03	190	0.03	285
meanHF	0.05	361	0.01	186	0.04	580	0.03	318	0.02	437	0.03	501
minLF	0.01	952	0.00	332	0.00	668	0.00	329	0.00	316	0.00	904
minMF	0.01	957	0.00	352	0.00	683	0.00	347	0.00	326	0.00	931
minHF	0.01	659	0.00	195	0.00	374	0.00	459	0.00	298	0.00	733
maxLF	0.20	276	0.10	245	0.18	351	0.14	261	0.12	263	0.15	309
maxMF	0.20	278	0.10	249	0.18	358	0.14	266	0.12	265	0.15	314
maxHF	0.22	430	0.06	553	0.22	677	0.13	574	0.10	520	0.15	645
meanDerivativeLF	0.00	6410	0.00	17812	-0.00	3493	-0.00	8417	-0.00	10835	-0.00	8491
meanDerivativeMF	0.00	6420	0.00	22152	-0.00	3482	-0.00	7899	-0.00	10953	-0.00	8338
meanDerivativeHF	0.00	5907	-0.00	1278	-0.00	1665	-0.00	1853	-0.00	1659	-0.00	2531
sympathiovagalBalance	2.26	87	2.83	81	2.72	84	2.82	93	2.43	77	2.65	86
relativeMFPower	0.56	36	0.58	34	0.59	32	0.57	36	0.59	31	0.58	34
totalVariability	0.15	322	0.05	137	0.12	411	0.09	206	0.08	251	0.09	342

Table 5.17: PPEMaverage for the studied population, using QualAff as emotion representation.

a_j	b_j	x_j	y_j	C_j^c	S_j
		-0.3875	-0.5059	1	$HR+' > 0$
		-0.3875	-0.5059	1	$SC+' < 0$
		0.6625	0.1176	2	$HR+' > 0$
		0.1125	0.6	4	$HR+' > 0$
		-0.4875	0.5176	6	$HR+' > 0$
		-0.4875	0.5176	6	$SC+' > 0$
		-0.65	-0.2118	8	$HR+' > 0$
		-0.65	-0.2118	8	$SC+' > 0$
		-0.6125	0.0706	9	$HR+' > 0$
		-0.6125	0.0706	9	$SC+' > 0$
$> 0^*$					$HR+' > 0$
	$> 0^*$				$SCR+' > 0$
	$> 0^*$				$SC+' > 0$

Table 5.18: Merged $\text{PPEM}_{\text{average}}$ from discrete and representational emotion representation.

5.7.2 Design of SCRs real time recognition

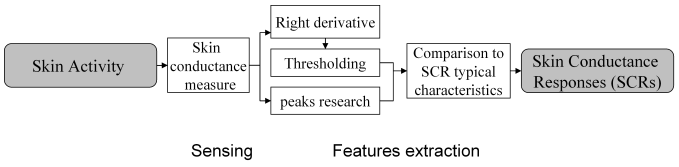

Figure 5.47: Detection of Skin Conductance Responses

The extraction of SCRs in real time required several modifications from the offline recognition system. This method gives priority to the extraction of a high rise in the SC signal (by studying the derivative). However, computing time on a Pentium 4, 2 Gh, 512 Mo RAM resulted in a minimum time of 31.6 (superior to 31.25 ms which corresponds to 32 frame per seconds). Thus we adapted the previous algorithm for real time (see figure 5.47).

The steps are mostly the same that the online one. First we add a new sc value to the buffer and discard the first of the buffer. Then, we smooth the data and search negative peak in the smoothed buffer. Then, we compute the right derivative and search for positive peaks in the derivative situated after the negative peak above a threshold. We finally search a positive peak in the smoothed buffer. If all these conditions are verified this means that we are in presence of a SCR : we have identified a left negative peak, a thresholded derivative value and a rigth positive peak.

The figure 5.48 presents a screenshot of the software implementation for emotional features extraction. On the left is presented the Heart Rate

Analyzer system, based on Heart Rate Sensor class we designed. On the right is presented the skin conductance Analyzer. Note that any skin conductance device could be connected if it can output a skin conductance value as a double with fixed timestamp.

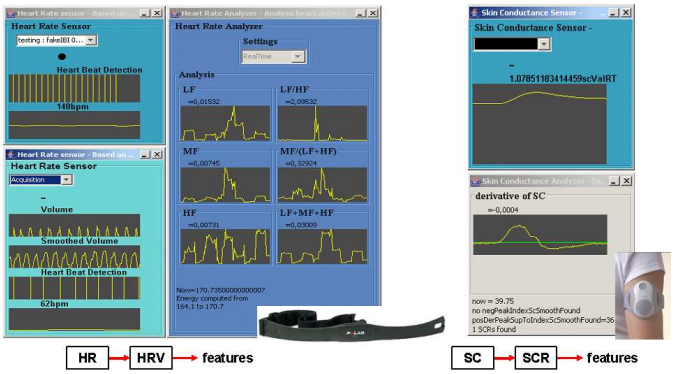

Figure 5.48: The implemented software performing short-term physiological emotional features analysis.

5.8 Conclusion

We provided a new methodology enabling (1) to tailor the interpretation of physiological components of emotion in terms of psychological descriptor of emotion (namely the valence and arousal dimension or affective discrete labels). Moreover, we provided a system to measure skin conductance and heart rate trough computer, and extract the heart rate variability and skin conductance responses. We presented a possible representation of what we call the $PPEM_{\text{average}}$, in order to constitute a computational state-of-the-art of psychophysiological rules published. Moreover, we performed an experiment based on the presented methodology, involving 40 subjects. The analysis of this experiment lead into several results, confirming the presented approach and allowing to build adapted $PPEM$.

Especially we shown that inter-individual differences are an important phenomenon to consider in emotion estimation from physiological signals. Discrete and dimensional representation can be both conisdered as ouptut to represent the subjective affective experience of individuals.

We analyzed different existing approaches by taking into consideration the user dependency criterion (user dependent versus user independent data) and the Subjectivity of stimuli criterion (social agreement versus subjective rating). Empirically, we showed that (1) physiological inter-individual differences are not related to psychological inter-individual differences and that (2) choice of stimuli based on psychological agreement does not involve the fact results are generalizable for subjective stimuli.

We obtained a maximum of 93% of recognition for qualitative Affect classes, at intra-individual level.

Chapter 6

Java API for Measuring and Modeling Affective States and Emotions Related to Multimedia Contents

6.1 Introduction

The E.A.R. (to Multimedia) and P.P.E.M. models are partially implemented into a unique set of packages (we will denotes it as EARMultimedia software). We provide here the whole architecture, including implemented work, and specifications (i.e. potential incoming implementation for which skeleton is already present in the code), to give reader an overview of the system rather than a complete system, and demonstrate the possibilities of applications it involve.

The EAR Multimedia software is the platform to make experiments and implement results of the E.A.R. and P.P.E.M. and use the E.A.R. to select, modify and potentially design the PE. Computing here is used as a tool at the service of the simulation of human behavior, but could be simultaneously considered as an usable application in affective computing.

This set of package is only intented to be an *experimental* set of tools for fundamental research on emotion, focused on the both approaches presented in this thesis. An effort has been made on the object-oriented design of classes and on the documentation to share research knowledge in the affective computing research community.

It is made of a Java set of packages, along with an API to allow user to build simulations and applications. Despite directly use the provided API, and follow the *modes* of the software, the software is easily extendable under two forms, according to the possibilities given by the Java language. New java classes could be built both by extending existing class or by implementing provided interfaces.

We provide in this chapter an overview of the java packages, the archi-

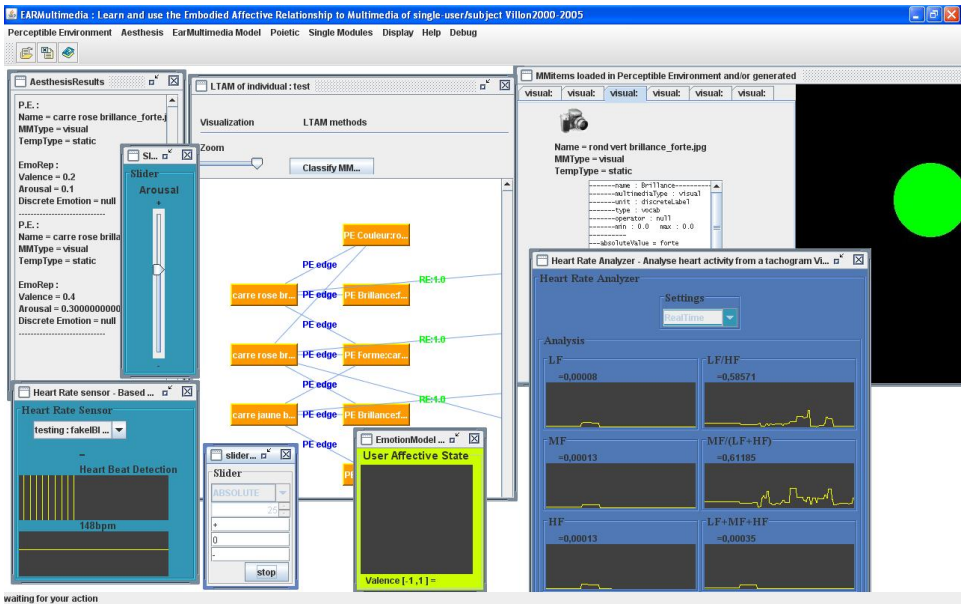

Figure 6.1: A screen capture of classes of the EarMultimedia API.

tecture of each packages, the predefined modes of software, and finally the direct application domain of EARMultimedia.

This section only gives an overwiew of the interfaces, classes and packages of the API (made of a total of 102 classes and interfaces). Therefore the reader is invited to consult the API documentation for a more detailled description. The API documentation and packages could be found at `http://ovillon.free.fr/thesis/`. Figure 6.1 presents a screen capture of the GUI of the software.

6.2 Java EARMultimedia API

EarMultimedia comes with five packages, plus one transversal set of tools, organized as following (Figure 6.2). The use of this package is made to provide some *builds*, but user is free to build its own experimental application using the packages.

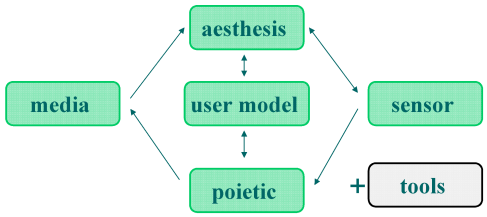

Figure 6.2: The EarMultimedia packages organization.

The five principal components are : *multimedia control and formalization* (media), *direct and indirect affective state measurement* (sensor), *EAR measurement* (aesthesis), *emotion modeling* (user model), *multimedia selection, modification and generation* (poietic)[1] In each package, several classes allow to launch a GUI based application to perform several tasks, but all classes of a package could be used independently.

All the software is written in java, and a complete API documentation (Figure 6.3) is provided for programmers.

Figure 6.3: API documentation.

As it is written in java, the software inherits of flexibility of this language. Extension of classes is possible (e.g. customize the PerceptibleEnvironment class in the media package to allow new formats to be added) as well as implementation of provided interfaces (e.g. using Sensor interface in the sensor package to connect any emotional measure devices). Moreover, java offers to connect to several other language (C, C++, etc...) using the Java Native Interface. Indeed, it is a cross-platform language, which could easily be embedded into devices oriented to personalization, like personal digital assistant. Thus, the provide API is extensible and could be easily connected to other applications.

6.3 Architecture and components

We will describe here what are the main components of the packages (we do not discusses tools, which implements mainly buffering of data, file Input/Output, real time graph plotting, multimedia real time analysis), and their architecture.

[1]The poietic package is actually not provided as it is not enough developped and not discussed in detail in this thesis. We however present it here to give an overview of its relationship with other packages. The role of the poietic package is to select, modify and even design multimedia contents according to EAR contents, which could be developped by collaborating with artist, designer, etc... but this is out of the scope of this thesis

6.3.1 Multimedia control and formalization (media)

This package provides perceptible environment analysis, synthesis, formalization and displaying (see Figure 6.4). The core element of this package is the MMitem (multimedia item).

A MMItem implements both the media and the formalization of this media (in xml) and support sound, image and video. From a multimedia content file, it could extract perceptible formalization.

The system of interface of Renderers and Formalizers (presented in section 4.5) let developper embed several representations of the P.E. discussed in previous section, as MPEG-7 analysis, IPEM Toolbox neural coding, musicXML, etc.... From a set of (synthesis) formalization, it can build a file (using engine as midi for sound) fulfilling the formalization.

Moreover, from a specific 'universe' specification file, it can compute possible combinations (using the Possibles class) from a set of primitives (of type variables/values, as color hue, sound intensity, etc...), and builds MMItems accordingly.

This is useful for researchers aiming at measuring user affective reaction into a specific environment, like modeling an instrument of music possibilities, or the types of colors and textures which could be displayed by a GUI interface. Indeed, the package propose a SlideShow class for multimedia displaying, and a PerceptibleEnvironment class which keeps track of all the existing MMItems.

The section 4.5 contains details about the mechanisms implemented into this package.

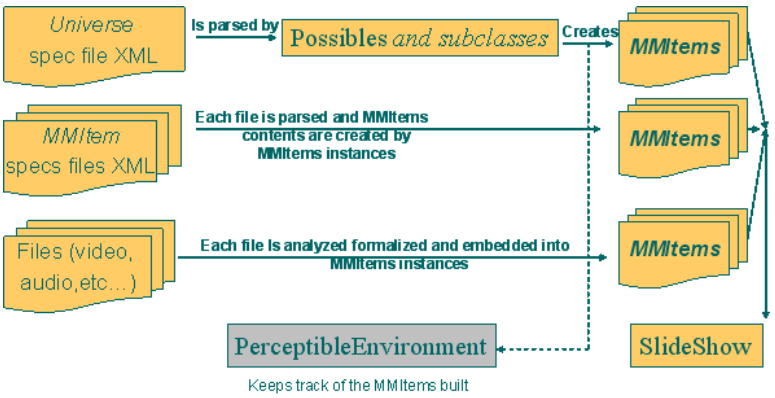

Figure 6.4: Architecture of the Media package

6.3.2 Affective state measurement (sensor)

. This package serves as an (in)direct measure of the user's subjective emotional experience. It support methodologies based on the relation between

first (subjective experience self-report) and third (any measure we can interpret as an indice of the subjective experience) person approaches of user emotion. The software engineering is designed to support (near to real time) emotion features extraction.

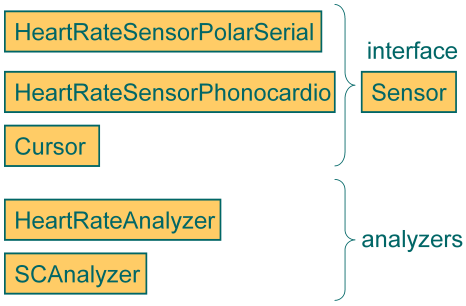

Figure 6.5: Architecture of the Sensor package.

It allows the acquisition of heart rate from phonocardiograph and ECG system, and provide analyzers for heart rate analysis as well as skin conductance analysis. The package supports both analysis in (near to) real time and offline. An interface, sensor, specify the minimal requirement for any other sensor, from sensing control to synchronization abilities for accurate measure when used with other devices. These interface could be implemented to build a new measure of affective state. Implementors of this interface are the classes HeartRateSensorPhonocardiographSoundcard, HeartRateSensorPolarSerial, SCSensorGeneric, SliderSensor.

6.3.3 EAR measurement (aesthesis).

The aesthesis is an old Greek term which refers to the aesthetical experience in presence of specific artifacts (see section 4.3). Here, the package associates media (dis)playing and direct or indirect assessment of user affective experience. Thus it mainly associates sensor instances (earmultimedia.sensor) with media instances (earmultimedia.media).

However, it implements specific aesthesis system like the software drag-and-drop multimedia, for direct assessment of MMItems in the valence*arousal emotional space (dndMultimedia is a multimedia extension of dndSounds, see [Villon, 2003], and is downloadable as a single executable at http://ovillon.free.fr/dndMultimedia/). The aesthesis outputs a AesthesisResult instance, which is a set of MMItems, and Sensor's outputs.

6.3.4 EAR simulation and use (user model).

The package user model is made for three tasks (see Figure 6.6). Firstly, it hosts the (a) representation of emotion (called EmotionRepresentation),

which could be any discrete model of emotion, or a dimensional one. However, as we seen in previous sections, the emphasis is made on dimensional model, to inherit of the computational facilities. Then, it deals with (b) the user's process and memory contents (EAR, mostly LTAM). The EAR algorithms are embedded into the EarMultimedia and LTAM class. Finally, it hosts user's psychophysiological emotional map (PPEM), i.e. what allow converting sensor output's in term of emotion, when they are indirect emotional measures (i.e. like heart rate, see sensor package).

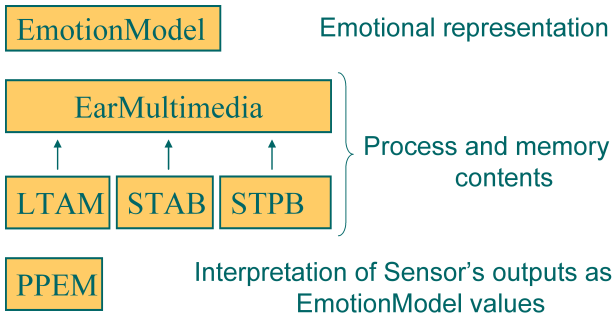

Figure 6.6: Architecture of the UserModel package

6.3.5 Multimedia selection, modification and generation (poietic).

This package is still a project. The main functionality is to select, modify, or synthetize PE according to the current state of the user (earmultimedia.sensor), the LTAM of this user (earmultimedia.usermodel), and possibility to manipulate the environment (earmultimedia.media). This is presented here to see the potential connection with other packages.

6.4 Application domain

In firsts section, we introduced several applications which tend to embeds what we call the EAR into their system. We think that introducing a research of depth in design of the emotional relationship of users with their multimedia environment to systems, according to human being could benefits to several domains.

The presented model add the notion of user's synthesized emotional memory from personal past experiences with multimedia content for any user modeling application. Thus, it could extend systems aiming at embedding affective modeling of user, as the Multimodal Affective User Interface ([Lisetti and Nasoz, 2002]). As the MAUI framework is designed to manipulate agent, not focusing on multimedia contents, it could extends the

MAUI concept, as well as other systems using affect in HCI. According to the above-presented modes, EARMultimedia could have several applications. The Simulation mode could be complementary with approaches of personalized multimedia content delivery, based on Multimedia Indexing and Retrieval, like the work of [Hanjalic and Xu, 2005].

The Poietic mode could be developped on this basis to extends and make novels form of interactive art or help in design field like sound design field. More generally , it helps to take into account personalized user affective relationship with its environment, in any interactive designs, where we expect a specific affective effect on the user.

6.5 Conclusion

The presented API is an experimental tool to assess and model affective state and emotion related to multimedia contents. It constitute a set of (beta) libraries to build and test potential issues of the presented models. It can allow researchers to investigate on emotion measure and modeling.

Chapter 7

Conclusion and Future Work

7.1 Summary and Contributions

In this Ph.D. thesis, two user model regarding emotion measure and emotion modeling associated to a practical implementation have been presented.

We shown that several fields of experimental researches focuses on the human ability to associate affective experience to media, along with the fact that computing could take benefits of theses researches in several applications involving affects and emotion. The inter-individual difference among individual reacting to same media had been poorly investigated and is still a barrier to the development of personalized applications like multimedia content delivery and HCI design. We proposed to consider the notion of Embodied Affective Relationship as a mean to take into account cultural and personal individual past affective experience responsible for the individualisation of affective experience to perceptible environment. We shown a formalized model of this notion, along with an implementation, to simulate the EAR of an individual with a controlled environment and then use it into multimedia application.

Moreover, we proposed the approach of the $PPEM$ to overcome inter-individual difference modeling in emotional measure. We provided a system to measure skin conductance and heart rate trough computer, and extract the heart rate variability and skin conductance responses. We presented a possible representation of what we call the $PPEM_{\text{average}}$, in order to constitute a computational state-of-the-art of psychophysiological rules published. Moreover, we performed an experiment based on the presented methodology, involving 40 subjects. The analysis of this experiment lead into several results, confirming the presented approach and allowing to build adapted $PPEM$. we shown that inter-individual differences are an important phenomenon to consider in emotion estimation from physiological signals. Discrete and dimensional representation can be both conisdered as ouptut to represent the subjective affective experience of individuals.

We analyzed different existing approaches by taking into consideration the user dependency criterion (user dependent versus user independent data)

and the Subjectivity of stimuli criterion (social agreement versus subjective rating). Empirically, we showed that (1) physiological inter-individual differences are not related to psychological inter-individual differences and that (2) choice of stimuli based on psychological agreement does not involve the fact results are generalizable for subjective stimuli. We obtained a maximum of 93% of recognition for qualitative Affect classes, at intra-individual level.

The work performed during this thesis lead into the following publications (1 journal, 6 international conferences with committee, 1 national conference with committee, 1 research report):

1. Camurri, A., Castellano, G., Cowie, R., Glowinski, D., Knapp, B., Krumhansl, C. L., **Villon, O.**, and Volpe, G. (2007a). The premio paganini project: a multimodal gesture-based approach for explaining emotional processes in music performance. In *In Gesture in Human-Computer Interaction and Simulation, GW 2007, Selected Revised Papers, Lecture Notes In Computer Science*. to appear.

2. **Villon, O.** (2007). Toward a computational model of subjective affective states associated with multimedia contents. Technical report, Institut Eurecom. RR-07-200.

3. Camurri, A., Castellano, G., Cowie, R., Glowinski, D., Knapp, B., Krumhansl, C. L., **Villon, O.**, and Volpe, G. (2007b). The premio paganini project: a multimodal gesture-based approach for explaining emotional processes in music performance. In *The 7th International Workshop on Gesture in Human-Computer Interaction and Simulation 2007, 23-24-25 May 2007, Lisbon, Portugal*, pages 65–67.

4. **Villon, O.** and Lisetti, C. (2006a). Affective multimedia interaction grounded on a cognitive science approach : interpreting indirect measures of emotion and modeling the affective relationship to multimedia contents. In *3rd HUMAINE EU Summer School Casa Paganini-InfoMus Lab, DIST, University of Genova, Italy, September 22-28, 2006*.

5. **Villon, O.** and Lisetti, C. (2007a). Toward recognizing individual?s subjective emotion from physiological signals in practical application. In *CBMS 2007 : 20th IEEE International Symposium on COMPUTER-BASED MEDICAL SYSTEMS*.

6. **Villon, O.** and Lisetti, C. (2007b). A user model of psycho-physiological measure of emotion. In *UM2007 User Modeling: Proceedings of the Eleventh International Conference*, Springer's LNAI series.

7. **Villon, O.** and Lisetti, C. (2006c). 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, 6-8 September 2006*, pages 269–276, Hatfield, United Kingdom. IEEE.

8. **Villon, O.** and Lisetti, C. (2006b). Toward building adaptive user's psycho-physiological maps of emotions using bio-sensors. In Reichardt, D., Levi, P., and Meyer, J.-J. C., editors, *1st Workshop on Emotion and Computing at KI2006 : Current Research and Future Impact, Bremen, Germany, June 19th, 2006*, pages 35–38.

9. **Villon, O.** and Lisetti, C. (2005). 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.

7.2 Future Works

Several extension of the work presented could be performed. This work and the implementation specifications could open a systematic mean to measure and model individual's affectice states and emotion related to multimedia contents in various scenarios of HCI, CMC, Interactive Art and PMCD.

The PPEM approach may be extended to a generic indirect measure of affective state and emotions. The input was dedicated to the physiological one in this thesis, and the ouptut the abstract representation of affective states and emotion of an individual. The approach could be extended to enable any kind of input containing affective information and which could be considered as expressing affective state and emotions. For instance the study of performing arts like expressive gestures analysis (e.g. computational analysis of dance) could use such approach to model the average affective states elicited by specific kind of expressive structure.

The Embodied Affective Relationship to Multimedia might open novel forms of interaction. For instance, implementing the Poietic mode may be combining with any generative approach, like SaXex ([Cañamero et al., 1999]) system or contribute to the computational models of expressivity (www.infomus.org). This generic model might be adapted for specific domain.

Finally, we hope that the provided API will be used and extended by researchers of the Affective Computing domain, User-Modeling domain and Computationnal Modeling of performing art. The notion of API is made to be used by other and the set of interfaces is ready for allowing specific experimental applications for computational measure and modeling of emotion elicited by multimedia contents.

Bibliography

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

[Aittomaki and Salmenpera, 1997] Aittomaki, J. V. and Salmenpera, M. T. (1997). Association between r-wave amplitude of the electrocardiogram and myocardial function during coronary artery bypass grafting. *Journal of Cardiothoracic and Vascular Anesthesia*, 11(7):856–860.

[Alexander et al., 2005] Alexander, D. M., Trengove, C., Johnston, P., Cooper, T., August, J. P., and Gordon, E. (2005). Separating individual skin conductance responses in a short interstimulus-interval paradigm. *J Neurosci Methods*, 146(1):116–123.

[Anttonen and Surakka, 2005] Anttonen, J. and Surakka, V. (2005). Emotions and heart rate while sitting on a chair. In *CHI 2005, Portland, Oregon, USA, April 2-7*, pages 491–499.

[Arbib and Fellous, 2004] Arbib, M. A. and Fellous, J.-M. (2004). Emotions: from brain to robot. *Trends in Cognitive Sciences*, 8(12):554–561.

[Balkenius and Morén, 1998a] Balkenius, C. and Morén, J. (1998a). A computational model of emotional conditioning in the brain. In *SAB'98 Workshop : Grounding Emotions in Adaptive Systems.*

[Balkenius and Morén, 1998b] Balkenius, C. and Morén, J. (1998b). Computational models of classical conditioning: a comparative study. In Pfeifer, R., Blumberg, B., Meyer, J.-A., and Wilson, S. W., editors, *From animals to animats 5 :Proceedings of the Fifth International Conference on Simulation of Adaptive Behavior (SAB'98, August 17-21, 1998, Zurich, Switzerland)*, Complex Adaptive Systems series, Cambridge, MA. MIT Press.

[Barbieri et al., 2003] Barbieri, R., Matten, E., and Brown, E. (2003). Instantaneous monitoring of heart rate variability. volume 1, pages 204–207 Vol.1.

[Basilico and Hofmann, 2004] Basilico, J. and Hofmann, T. (2004). Unifying collaborative and content-based filtering. In *Proceedings of the twenty-first international conference on Machine learning - ACM International Conference Proceeding Series Vol. 69*, volume 69, page 9.

[Bastard, 2004] Bastard, G. (2004). Modélisation et estimation des émotions de l'utilisateur à partir de paramètres faciaux. Master's thesis, Université de Nice Sophia-Antipolis, Institut EURECOM, Département Communications Multimédia. Rapport de stage pour le diplôme d' études approfondies image-vision de l'Université de Nice Sophia-Antipolis.

[Büchel et al., 1999] Büchel, C., Dolan, R. J., Armony, J. L., and Friston, K. J. (1999). Amygdala-hippocampal involvement in human aversive trace conditioning revealed through event-related functional magnetic resonance imaging. *J Neurosci*, 19(24):10869–10876.

[Blanchard et al., 2001] Blanchard, C., Blanchard, R., Fellous, J., Guimarães, F., Irwin, W., Ledoux, J., McGaugh, J., Rosen, J., Schenberg, L., Volchan, E., and Cunha, C. D. (2001). The brain decade in debate: Iii. neurobiology of emotion. *Braz J Med Biol Res*, 34(3):283–93.

[Blood and Zatorre, 2001] Blood, A. J. and Zatorre, R. J. (2001). 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):11818–11823.

[Bomers and Pfisterer, 2005] Bomers, F. and Pfisterer, M. (2005). www.jsresources.org. official web site to learn the Java Sound API.

[Bourdieu and de Saint-Martin, 1976] Bourdieu, P. and de Saint-Martin, M. (1976). Anatomie du goût. *Actes de la recherche en sciences sociales*, 11:2–112.

[Bradley and Lang, 2000] Bradley, M. and Lang, P. (2000). Affective reactions to acoustic stimuli. *Psychophysiology*, 37(2):204–15.

[Broersen, 2000] Broersen, P. (2000). Facts and fiction in spectral analysis. *Instrumentation and Measurement, IEEE Transactions on*, 49(4):766–772.

[Bureaud, 1999] Bureaud, A. (1999). Utopies distribuées. net.art, web art. *Art Press*. N° Hors-série, dir. Norbert Hillaire, Internet all over, l'art et la toile.

[Cañamero et al., 1999] Cañamero, D., Arcos, J., and de Mántaras, R. L. (1999). Imitating human performances to automatically generate expressive jazz ballads. In *AISB'99 Symposium on Imitation in Animals and Artifacts, Edinburgh, 6-9 April 1999*, pages 115–120.

[Cacioppo and Tassinary, 1990] Cacioppo, J. and Tassinary, L. (1990). Inferring psychological significance from physiological signals. *American Psychologist*, 45(1):16–28.

[Carré, 2002] Carré, M. (2002). *Systèmes de Recherche de Documents Musicaux par Chantonnement*. PhD thesis, Ecole Nationale Supérieure des Télécommunications.

[Carvalho et al., 2003] Carvalho, J., Rocha, A., Junqueira, L.F., J., Neto, J., Santos, I., and Nascimento, F. (2003). A tool for time-frequency analysis of heart rate variability. In *Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, volume 3, pages 2574–2577 Vol.3.

[Castiglioni et al., 2005] Castiglioni, P., Cerutti, C., di Rienzo, M., Elghozi, J.-L., Honzikova, N., Janssen, B., Kardos, A., Mrowka, R., Parati, G., Persson, P., and Quintin, L. (2005). Glossary of terms used in time series analysis of cardiovascular data : http://www.cbi.dongnocchi.it/glossary/. web site. Working Group on Blood Pressure and Heart Rate Variability of the European Society of Hypertension.

[Castiglioni and Rienzo, 1996] Castiglioni, P. and Rienzo, M. D. (1996). On the evaluation of heart rate spectra: the lomb periodogram. *Computers in Cardiology*.

[Castiglioni et al., 2002] Castiglioni, P., Rienzo, M. D., and Yosh, H. (2002). A computationally efficient algorithm for online spectral analysis of beat-to-beat signals. *Computers in Cardiology*, pages 417– 420.

[Changchun et al., 2005] Changchun, L., Rani, P., and Sarkar, N. (2005). An empirical study of machine learning techniques for affect recognition in human-robot interaction. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2-6 Aug.*, pages 2662– 2667.

[Christie, 2002] Christie, I. C. (2002). Multivariate discrimination of emotion-specific autonomic nervous system activity. Master's thesis, Faculty of the Virginia Polytechnic Institute and State University.

[Christie and Friedman, 2004] Christie, I. C. and Friedman, B. H. (2004). Autonomic specificity of discrete emotion and dimensions of affective space: a multivariate approach. *Int J Psychophysiol*, 51(2):143–53.

[Clifford, 2002] Clifford, G. D. (2002). *Signal Processing Methods for Heart Rate Variability*. PhD thesis, University of Oxford.

[Courvile et al., 2004] Courvile, A. C., Daw, N., and Touretzky, D. (2004). Similarity and discrimination in classical conditioning: A latent variable account. *Advances in Neural Information Processing Systems*, 17:313–320.

[Crawford and Henry, 2004] Crawford, J. R. and Henry, J. D. (2004). The positive and negative affect schedule (panas): Construct validity, measurement properties and normative data in a large non-clinical sample. *British journal of clinical psychology*, 43:245 – 265.

[Critchley, 2002] Critchley, H. (2002). Electrodermal responses: What happens in the brain? *The Neuroscientist*, 8:132–142.

[Cupchik, 1994] Cupchik, G. C. (1994). Emotion in aesthetics: Reactive and reflective models. *Poetics*, 23(1-2):177–188.

[Cuthbert et al., 2000] Cuthbert, B., Schupp, H., Bradley, M., Birbaumer, N., and Lang, P. (2000). Brain potentials in affective picture processing: covariation with autonomic arousal and affective report. *Biol Psychol*, 52(2):95–111.

[Darwin, 1997] Darwin, C. J. (1997). Auditory grouping. *Trends in Cognitive Sciences*, 1:327–333.

[Dawson et al., 2001] Dawson, M., Schell, A., and Filion, D. (2001). The electrodermal system. In Cacioppo, J., Tassinary, L., and Berntson, G., editors, *Handbook of psychophysiology*, page 53?84. Cambridge: Cambridge University Press.

[Delalande et al., 1996] Delalande, F., Formosa, M., Frémiot, M., Gobin, P., Malbosc, P., Mandelbrojt, J., and Pedler, E. (1996). *Les Unités Sémiotiques Temporelles : Éléments nouveaux d'analyse musicale.* Éditions MIM Documents Mesurgia.

[Delorme and Thorpe, 2003] Delorme, A. and Thorpe, S. J. (2003). Spikenet: an event-driven simulation package for modelling large networks of spiking neurons. *Network: Comput. Neural Syst.*, 14:613–627.

[Donahoe and Vegas, 2004] Donahoe, J. W. and Vegas, R. (2004). Pavlovian conditioning: the CS-UR relation. *J Exp Psychol Anim Behav Process*, 30(1):17–33.

[Dumouchel, 1991] Dumouchel, R. (1991). Le spectacteur et le contactile. *Cinémas*, 1(3). Nouvelles technologies : nouveaux cinémas ?, dir. Michel Larouche.

[Efron, 1986] Efron, B. (1986). How biased is the apparent error rate of a prediction rule. *Journal of the American Statistical Association*, 81(394):461–470.

[Eifert et al., 1988] Eifert, G. H., Craill, L., Carey, E., and O'Connor, C. (1988). Affect modification through evaluative conditioning with music. *Behaviour Research and Therapy*, 26(4):321–330.

[Electrophysiology, 1996] Electrophysiology, T. F. o. t. E. S. o. C. t. N. A. S. o. P. (1996). Heart rate variability : Standards of measurement, physiological interpretation, and clinical use. *Circulation*, 93(5):1043–1065.

[Elisseeff and Pontil, 2003] Elisseeff, A. and Pontil, M. (2003). Leave-one-out error and stability of learning algorithms with applications. In et al., J. S., editor, *Advances in Learning Theory: Methods, Models and Applications*, volume 190 of *NATO Science Series III: Computer and Systems Sciences*. IOS press.

[Fellous, 1999] Fellous, J. (1999). Neuromodulatory basis of emotion. *The Neuroscientist*, 5:283?294.

[Field et al., 2001] Field, A., Hartel, P., and Mooij, W. (2001). 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.

[Fiorito and Simons, 1994] Fiorito, E. and Simons, R. (1994). Emotional imagery and physical anhedonia. *Psychophysiology*, 31:513–521.

[Foote et al., 2002] Foote, J., Cooper, M., and Nam, U. (2002). Audio retrieval by rhythmic similarity. In *Third International Symposium on Musical Information Retrieval (ISMIR), September 2002, Paris*, Paris.

[Fowles, 1986] Fowles, D. (1986). The eccrine system and electrodermal activity. In MGH Coles, E Donchin, S. P., editor, *Psychophysiology*, pages 51–96. Guilford Press, New York.

[Fowles et al., 1981] Fowles, D., Christie, M., Edelberg, R., Grings, W., Lykken, D., and Venables, P. (1981). Committee report. publication recommendations for electrodermal measurements. *Psychophysiology*, 18(3):232–9.

[Friedman and Thayer, 1998] Friedman, B. H. and Thayer, J. F. (1998). Autonomic balance revisited: Panic anxiety and heart rate variability. *Journal of Psychosomatic Research*, 44(1):133–151.

[Gao et al., 2004] Gao, S., Lee, C.-H., and Tian, Q. (2004). Indexing with musical events and its application to content-based music identification.

In *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, volume 3, pages 846 – 849. 23-26 Aug. 2004.

[Gaudreau et al., 2006] Gaudreau, P., Sanchez, X., and Blondin, J.-P. (2006). Positive and negative affective states in a performance-related setting: Testing the factorial structure of the panas across two samples of french-canadian participants. *European J of Psych Assessment*. Accepted for publication.

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

[Good, 2002] Good, M. (2002). Musicxml in practice: Issues in translation and analysis. In *Proceedings First International Conference MAX 2002: Musical Application Using XML*, pages 47–54, Milan.

[Gratch and Marsella, 2004] Gratch, J. and Marsella, S. (2004). A domain-independent framework for modeling emotion. *Cognitive Systems Research*, 5(4):269–306.

[Guttman and Kalish, 1956] Guttman, N. and Kalish, H. (1956). Discriminability and stimulus generalization. *Journal of Experimental Psychology*, 51(1):79–88.

[Guyader et al., 2002] Guyader, N., Le Borgne, H., Hérault, J., and Guérin-Dugué, A. (2002). Towards the introduction of human perception in a natural scene classification system. In *IEEE International workshop on Neural Networks for Signal Processing (NNSP'2002), pp 385-394, Martigny Valais, Switzerland, September*, pages 385–394, Martigny Valais, Switzerland.

[Haag et al., 2004] Haag, A., Goronzy, S., Schaich, P., and Williams, J. (2004). Emotion recognition using bio-sensors: First steps towards an automatic system. *LNCS*, 3068:36–48.

[Hamann, 2001] Hamann, S. (2001). Cognitive and neural mechanisms of emotional memory. *Trends Cogn Sci*, 5(9):394–400.

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

[He et al., 1995] He, J., Kinouchi, Y., Yamaguchi, H., and Miyamoto, H. (1995). Exercise-induced changes in r wave amplitude and heart rate in normal subjects. *J Electrocardiol*, 28(2):99–106.

[Healey and Picard, 1998] Healey, J. and Picard, R. (1998). Digital processing of affective signals. In *Acoustics, Speech, and Signal Processing, 1998. ICASSP '98. Proceedings of the 1998 IEEE International Conference on*, volume 6, pages 3749 – 3752. IEEE.

[Healey et al., 1998] Healey, J., Picard, R., and Dabek, F. (1998). A new affect-perceiving interface and its application to personalized music selection. In *Proceedings of the 1998 Workshop on Perceptual User Interfaces PUI'98 (4-6 November 1998)*, San Francisco, CA.

[Healey et al., 1999] Healey, J., Seger, J., and Picard, R. (1999). Quantifying driver stress: developing a system for collecting and processing bio-metric signals in natural situations. *Biomed Sci Instrum*, 35:193–8.

[Herlocker, 2000] Herlocker, J. (2000). *Understanding and Improving Automated Collaborative Filtering Systems*. PhD thesis, University of Minnesota.

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

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

[Houwer et al., 2001] Houwer, J. D., Thomas, S., and Baeyens, F. (2001). Associative learning of likes and dislikes: a review of 25 years of research on human evaluative conditioning. *Psychol Bulletin*, 127(6):853–69.

[Hughes and Stoney, 2000] Hughes, J. and Stoney, C. (2000). Depressed mood is related to high-frequency heart rate variability during stressors. *Psychosom Med*, 62(6):796–803.

[Iwahama et al., 2004] Iwahama, K., Hijikata, Y., and Nishida, S. (2004). Content-based filtering system for music data. In *Applications and the Internet Workshops, 2004. SAINT 2004 Workshops. 2004 International Symposium on*, pages 480–487. IEEE.

[Izsó et al., 1999] Izsó, L., Mischinger, G., and Láng, E. (1999). Validating a new method for ergonomic evaluation of human computer interfaces. *Per. Pol. Soc. Man. Sci.*, 7(2):119–134.

[Jennings et al., 1981] Jennings, J. R., Berg, W. K., Hutcheson, J. S., Obrist, P., Porges, S., and Turpin, G. (1981). Publication guidelines for heart rate studies in man. *Psychophysiology*, 18:226–23.

[Jonsson and Sonnby-Borgstrom, 2003] Jonsson, P. and Sonnby-Borgstrom, M. (2003). The effects of pictures of emotional faces on tonic and phasic autonomic cardiac control in women and men. *Biological Psychology*, 62(2):157–173.

[Kang, 2003] Kang, H.-B. (2003). Affective content detection using hmms. In *MULTIMEDIA '03: Proceedings of the eleventh ACM international conference on Multimedia*, pages 259–262, Berkeley, CA, USANew York, NY, USA. ACM Press.

[Kensinger, 2004] Kensinger, E. A. (2004). Remembering emotional experiences: The contribution of valence and arousal. *Reviews in the Neurosciences*, 15:241–251.

[Kettunen, 1999] Kettunen, J. (1999). methodological and empirical advances in the quantitative analysis of spontaneous responses in psychophysiological time series. Technical report, university of Helsinki Department of Psychology research report 21, Helsinki.

[Kettunen and Keltikangas-Jarvinen, 2001] Kettunen, J. and Keltikangas-Jarvinen, L. (2001). Intraindividual analysis of instantaneous heart rate variability. *Psychophysiology*, 38(4):659–668.

[Khalfa et al., 2002] Khalfa, S., Peretz, I., Blondin, J.-P., and Robert, M. (2002). Event-related skin conductance responses to musical emotions in humans. *Neuroscience Letters*, 328(2):145–149.

[Kim et al., 2004] Kim, K. H., Bang, S. W., and Kim, S. R. (2004). Emotion recognition system using short-term monitoring of physiological signals. *Medical & Biological Engineering & Computing*, 42.

[Klinnert, 1984] Klinnert, M. D. (1984). The regulation of infant behavior by maternal facial expression. *Infant Behavior & Development*, 7:447?465.

[Kohrs and Mérialdo, 1999] Kohrs, A. and Mérialdo, B. (1999). Using color and texture indexing to improve collaborative filtering of art paintings. In *CBMI?99, 1st European Workshop on Content-Based Multimedia Indexing, October 25-27 1999, Toulouse, France.*

[Krumhansl, 2002] Krumhansl, C. L. (2002). Music: A link between cognition and emotion. *Current Directions in Psychological Science*, 11(2):45–50.

[Lane et al., 1997] Lane, R., Reiman, E., Bradley, M., Lang, P., Ahern, G., Davidson, R., and Schwartz, G. (1997). Neuroanatomical correlates of pleasant and unpleasant emotion. *Neuropsychologia*, 35(11):1437–44.

[Lang et al., 1997] Lang, P., Bradley, M., and Cuthbert, B. (1997). International affective picture system (iaps): Technical manual and affective ratings. NIMH Center for the Study of Emotion and Attention 1997. instruction for IAPS experiments with Self Assement Manikin (SAM).

[Lang et al., 2005] Lang, P., Bradley, M., and Cuthbert, B. (2005). International affective picture system (iaps): Digitized photographs, instruction manual and affective ratings. Technical report, Technical Report A-6. University of Florida, Gainesville, FL.

[Lang et al., 1993] Lang, P., Greenwald, M., Bradley, M., and Hamm, A. (1993). Looking at pictures: affective, facial, visceral, and behavioral reactions. *Psychophysiology*, 30(3):261–73.

[Lang, 1980] Lang, P. J. (1980). Behavioral treatment and bio-behavioral assessment: Computer applications. In J. B. Sidowski, J. H. Johnson, . T. A. W., editor, *Technology in mental health care delivery systems*, pages 119–167. Norwood, NY: Ablex.

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

[LeDoux, 2000] LeDoux, J. E. (2000). Emotion circuits in the brain. *Annu Rev Neurosci*, 23:155–184.

[Leman et al., 2001] Leman, M., Lesaffre, M., and Tanghe, K. (2001). Introduction to the ipem toolbox for perception-based music analysis. *Mikropolyphonie*, 7.

[Leventhal and Scherer, 1987] Leventhal, H. and Scherer, K. (1987). The relationship of emotion to cognition: A functional approach to a semantic controversy. *Cognition and Emotion*, 1:3–28.

[Levey and Martin, 1975] Levey, A. B. and Martin, I. (1975). Classical conditioning of human 'evaluative' responses. *Behav Res Ther*, 13(4):221–226.

[li Tian et al., 2001] li Tian, Y., Kanade, T., and Cohn, J. F. (2001). Recognizing action units for facial expression analysis. *IEEE Transactions On Pattern Analysis And Machine Intelligence*, 23(2).

[Lim et al., 1997] Lim, C. L., Rennie, C., Barry, R. J., Bahramali, H., Lazzaro, I., Manor, B., and Gordon, E. (1997). Decomposing skin conductance into tonic and phasic components. *International Journal of Psychophysiology*, 25(2):97–109.

[Lin and Hauptmann, 2006] Lin, W.-H. and Hauptmann, A. (2006). Label disambiguation and sequence modeling for identifying human activities

from wearable physiological sensors. In *Multimedia and Expo, 2006 IEEE International Conference on*, page 19972000.

[Lisetti and Nasoz, 2002] Lisetti, C. L. and Nasoz, F. (2002). Maui: a multimodal affective user interface. In *MULTIMEDIA '02: Proceedings of the tenth ACM international conference on Multimedia*, pages 161–170. ACM Press.

[Lisetti and Nasoz, 2004] Lisetti, C. L. and Nasoz, F. (2004). Using noninvasive wearable computers to recognize human emotions from physiological signals. *EURASIP Journal on Applied Signal Processing*, 11:1672–1687.

[Liu, 2003] Liu, H. (2003). A computational model of human affective memory and its application to mindreading. Technical report, MIT Media Laboratory Interactive Experience Group Technical Report IE03-01.

[Malmivuo and Plonsey, 1995] Malmivuo, J. and Plonsey, R. (1995). *Bioelectromagnetism : Principles and Applications of Bioelectric and Biomagnetic Fields*, chapter The Electrodermal Response, pages 428–434. Oxford University Press, New York.

[Marcos and Redondo, 1999a] Marcos, J. L. and Redondo, J. (1999a). Effects of conditioned stimulus presentation on diminution of the unconditioned response in aversive classical conditioning. *Biological Psychology*, 50(2):89–102.

[Marcos and Redondo, 1999b] Marcos, J. L. and Redondo, J. (1999b). Effects of cs-us interval modification on diminution of the unconditioned response in electrodermal classical conditioning. *Biological Psychology*, 50(3):191–201.

[Marolt, 2004] Marolt, M. (2004). A connectionist approach to transcription of polyphonic piano music. *IEEE Transactions on Multimedia*, 6(3):439–449.

[Martin et al., 1998] Martin, K., Scheirer, E., and Vercoe, B. (1998). Music content analysis through models of audition. In *1998 ACM Multimedia Workshop on Content Processing of Music for Multimedia Applications, Bristol UK*.

[Mateo and Laguna, 2000] Mateo, J. and Laguna, P. (2000). Improved heart rate variability signal analysis from the beat occurrence times according to the ipfm model. *Biomedical Engineering, IEEE Transactions on*, 47(8):985–996.

[McCraty et al., 1995] McCraty, R., Atkinson, M., Tiller, W. A., Rein, G., and Watkins, A. D. (1995). The effects of emotions on short-term power

spectrum analysis of heart rate variability. *The American Journal of Cardiology*, 76(14):1089–1093.

[Meste et al., 2005] Meste, O., Khaddoumi, B., Blain, G., and Bermon, S. (2005). Time-varying analysis methods and models for the respiratory and cardiac system coupling in graded exercise. *IEEE Trans. Biomed. Eng.* A paraître en 2005.

[Meyer, 1956] Meyer, L. (1956). *Emotion and Meaning in Music*. University of Chicago Press, Chicago.

[Moller and Dijksterhuis, 2003] Moller, P. and Dijksterhuis, G. (2003). Differential human electrodermal responses to odours. *Neuroscience Letters*, 346(3):129–132.

[Moody, 1993] Moody, G. (1993). Spectral analysis of heart rate without resampling. *Computers in Cardiology, IEEE Computer Society Press*, 20:715–718.

[Mumme and Fernald, 2003] Mumme, D. L. and Fernald, A. (2003). The infant as onlooker: learning from emotional reactions observed in a television scenario. *Child Dev*, 74(1):221–237.

[Nagai et al., 2004] Nagai, Y., Critchley, H. D., Featherstone, E., Trimble, M. R., and Dolan, R. J. (2004). Activity in ventromedial prefrontal cortex covaries with sympathetic skin conductance level: a physiological account of a "default mode" of brain function. *NeuroImage*, 22(1):243–251.

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

[Nuechterlein and Dawson, 1998] Nuechterlein, K. H. and Dawson, M. E. (1998). Neurophysiological and psychophysiological approaches to schizophrenia and its pathogenesis. In Watson, S. J., editor, *Psychopharmacology on CD-ROM: 1998 Edition*. Lippincott Williams & Wilkins.

[Ohman and Soares, 1994] Ohman, A. and Soares, J. (1994). "unconscious anxiety": phobic responses to masked stimuli. *Journal of Abnormal Psychology*, 103(2):231–40.

[Ortony et al., 1988] Ortony, A., Clore, G., and Collins, A. (1988). *The Cognitive Structure of Emotions*. Cambridge University Press., Cambridge:.

[Pachet, 2003] Pachet, F. (2003). The continuator: Musical interaction with style. *Journal of New Music Research*, 32(3):333–341.

[Palomba et al., 2000] Palomba, D., Sarlo, M., Angrilli, A., Mini, A., and Stegagno, L. (2000). Cardiac responses associated with affective processing of unpleasant film stimuli. *International Journal of Psychophysiology*, 36(1):45–57.

[Pan and Tompkins, 1985] Pan, J. and Tompkins, W. (1985). A real-time qrs detection algorithm. *IEEE Trans Biomed Eng*, 32(3):230–6.

[Panksepp and Bernatzky, 2002] Panksepp, J. and Bernatzky, G. (2002). Emotional sounds and the brain: the neuro-affective foundations of musical appreciation. *Behav Processes*, 60(2):133–155.

[Paulson and Tzanavari, 2003] Paulson, P. and Tzanavari, A. (2003). Combining collaborative and content-based filtering using conceptual graphs. In Lawry, J., Shanahan, J. G., and Ralescu, A. L., editors, *Modelling with Words*, volume 2873 of *Lecture Notes in Computer Science*, pages 168–185. Springer.

[Pearce, 1994] Pearce, J. M. (1994). Similarity and discrimination: A selective review and a connectionist model. *Psychological Review*, 101:587–607.

[Pearce, 2002] Pearce, J. M. (2002). Evaluation and development of a connectionist theory of configural learning. *Animal Learning & Behavior*, 30(,2):73–95.

[Pelletier and Pare, 2004] Pelletier, J. G. and Pare, D. (2004). Role of amygdala oscillations in the consolidation of emotional memories. *Biological Psychiatry*, 55(6):559–562.

[Peter and Herbon, 2006] Peter, C. and Herbon, A. (2006). Emotion representation and physiology assignments in digital systems. *Interacting with Computers*, 18(2):139–170.

[Phelps, 2004] Phelps, E. A. (2004). Human emotion and memory: interactions of the amygdala and hippocampal complex. *Current Opinion in Neurobiology*, 14(2):198–202.

[Picard, 1997] Picard, R. (1997). *Affective Computing*. MIT Press, Cambridge, Mass.

[Picard et al., 2001] Picard, R., Healey, J., and Vyzas, E. (2001). Toward machine emotional intelligence, analysis of affective physiological signals. *IEEE transactions on pattern analysis and machine intelligence*, 23(10).

[Rüdiger et al., 1999] Rüdiger, H., Klinghammer, L., and Scheuch, K. (1999). The trigonometric regressive spectral analysis–a method for mapping of beat-to-beat recorded cardiovascular parameters on to frequency domain in comparison with fourier transformation. *Comput Methods Programs Biomed*, 58(1):1–15.

[Rescorla and Wagner, 1972] Rescorla, R. and Wagner, A. (1972). A theory of pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In *Classical Conditioning II, A.H. Black and W.F. Prokasy, Eds.*, pages 64–99. Appleton-Century-Crofts. 2find.

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

[Rottenberg et al., 2006] Rottenberg, J., Ray, R., and Gross, J. (2006). Emotion elicitation using films. In Coan, J. and Allen, J., editors, *The handbook of emotion elicitation and assessment*. New York: Oxford University Press.

[Russell and Mehrabian, 1977] Russell, J. and Mehrabian, A. (1977). Evidence for a three-factor theory of emotions. *Journal of Research in Personality*, 11:273–294.

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

[Sander et al., 2003] Sander, D., Grafman, J., and Zalla, T. (2003). The human amygdala: an evolved system for relevance detection. *Rev Neurosci*, 14(4):303–16.

[Sander et al., 2005] Sander, D., Grandjean, D., and Scherer, K. R. (2005). A systems approach to appraisal mechanisms in emotion. *Neural Networks*, 18(4):317–352.

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

[Sava and Durand, 1997] Sava, H. and Durand, L.-G. (1997). Automatic detection of cardiac cycle based on an adaptive time-frequency analysis of the phonocardiogram. In *Proceedings - 19th International Conference - IEEE/EMBS Oct. 30 - Nov. 2, 1997 Chicago, IL. USA.*

[Schaeffer, 1977] Schaeffer, P. (1977). *Traité des objets musicaux*. Paris: Le Seuil.

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

[Scherer, 1984] Scherer, K. (1984). *Review of Personality and Social Psychology*, volume 5, chapter Emotion as a multicomponent process: A model and some cross-cultural data, pages 37–63. Beverly Hills, CA: Sage.

[Scherer, 1988] Scherer, K. (1988). *Facets of emotion: Recent research*, chapter Appendix F. Labels describing affective states in five major langages., pages 241–243. Hillsdale, NJ: Erlbaum. Version revised by the members of the Geneva Emotion Research Group.

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

[Scherer, 2001] Scherer, K. R. (2001). Psychologie des passions. conference.

[Shannon, 1948] Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423and 623–656.

[Smeulders et al., 2000] Smeulders, A., Worring, M., Santini, S., Gupta, A., and Jain, R. (2000). Content-based image retrieval at the end of the early years. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(12):1349–1380.

[Smith, 1997] Smith, S. W. (1997). *The Scientist and Engineer's Guide to Digital Signal Processing*. California Technical Pub.

[Sorce et al., 1985] Sorce, J., Emde, R. N., Campos, J. J., and Klinnert, M. (1985). Maternal emotional signalling : Its effect on the visual cliff behavior of 1 year olds. *Developmental Psychology*, 21(1):195–200.

[Takahashi et al., 2005] Takahashi, T., Murata, T., Hamada, T., Omori, M., Kosaka, H., Kikuchi, M., Yoshida, H., and Wada, Y. (2005). Changes in eeg and autonomic nervous activity during meditation and their association with personality traits. *International Journal of Psychophysiology*, 55(2):199–207.

[Teasdale and Russell, 1983] Teasdale, J. D. and Russell, M. L. (1983). Differential effects of induced mood on the recall of positive, negative and neutral words. *Br J Clin Psychol*, 22 (Pt 3):163–171.

[Teller, 1992] Teller, P. (1992). *Emergence or Reduction? Essays on the Prospects of Nonreductive Physicalism*, chapter A Contemporary Look at Emergence, pages 139–153. Berlin: Walter de Gruyter.

[Thorpe et al., 2000] Thorpe, S., Delorme, A., VanRullen, R., and Paquier, W. (2000). Reverse engineering of the visual system using networks of spiking neurons. In *IEEE International Symposium on Circuits and Systems*, volume 4, pages 405 –408.

[Tsai et al., 2000] Tsai, J., Levenson, R., and Carstensen, L. (2000). Autonomic, subjective, and expressive responses to emotional films in older and younger chinese americans and european americans. *Psychol Aging*, 15(4):684–93.

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

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

[Villon and Lisetti, 2005a] Villon, O. and Lisetti, C. (2005a). 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.

[Villon and Lisetti, 2005b] Villon, O. and Lisetti, C. (2005b). Psycho-physiological emotional sensing. Cognitive Imaging conference, (PACA Program), November 16, 2005, Rousset (ST MicroElectronics), France.

[Villon and Lisetti, 2006] Villon, O. and Lisetti, C. (2006). 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, 6-8 September 2006*, pages 269–276, Hatfield, United Kingdom. IEEE.

[Vyzas and Picard, 1999] Vyzas, E. and Picard, R. W. (1999). Offline and online recognition of emotion expression from physiological data. In *Workshop on Emotion-Based Agent Architectures, Third International Conference on Autonomous Agents, May 1*, Seattle, WA.

[Wagner et al., 2005] Wagner, J., Kim, J., and Andre, E. (2005). From physiological signals to emotions: Implementing and comparing selected methods for feature extraction and classification. In *Multimedia and Expo, 2005. ICME 2005. IEEE International Conference on*, pages 940– 943.

[Watson and Clark, 1999] Watson, D. and Clark, L. (1999). *The PANAS-X Manual for the Positive and Negative Affect Schedule ? Expanded Form.*

[Watson et al., 1988] Watson, D., Clark, L. A., and Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The panas scales. *Journal of Personality and Social Psychology*, 54:1063–1070.

[Winter and Kuiper, 1997] Winter, K. A. and Kuiper, N. A. (1997). Individual differences in the experience of emotions. *Clin Psychol Rev*, 17(7):791–821.

[Yoshida and Yanaru, 1995] Yoshida, K. and Yanaru, T. (1995). A proposal of emotional memory model. In *Proceedings. Second New Zealand International Two-Stream Conference on Artificial Neural Networks and Expert Systems (ANNES 1995)*, pages 67 – 70.

[Zentner et al., 2005] Zentner, M., Scherer, K., and Grandjean, D. (2005). Which emotions can be induced by music? In *ISRE 2005, Bari, Italy*, Bari, Italy. Poster presented at ISRE 2005.

[Zink et al., 2004] Zink, C. F., Pagnoni, G., Martin-Skurski, M. E., Chappelow, J. C., and Berns, G. S. (2004). Human striatal responses to monetary reward depend on saliency. *Neuron*, 42:509–517.