Emotion-oriented Systems

Emotion-oriented Systems

Edited by
Catherine Pelachaud

ISTE

WILEY

First published 2012 in Great Britain and the United States by ISTE Ltd and John Wiley & Sons, Inc.

ISTE Ltd
27-37 St George's Road
London SW19 4EU
UK

www.iste.co.uk

John Wiley & Sons, Inc.
111 River Street
Hoboken, NJ 07030
USA

www.wiley.com

Library of Congress Cataloging-in-Publication Data

Emotion-oriented systems / edited by Catherine Pelachaud.
p. cm.
Includes bibliographical references and index.
ISBN 978-1-84821-258-9 (hardback)
1. User-centered system design. 2. Human-computer interaction. 3. Interactive computer systems. 4. Virtual computer systems. 5. Emotions and cognition. I. Pelachaud, Catherine.
QA76.9.H85E485 2012
004.2'1--dc23
2011034774

British Library Cataloguing-in-Publication Data
A CIP record for this book is available from the British Library
ISBN 978-1-84821-258-9

Printed and bound in Great Britain by CPI Group (UK) Ltd., Croydon, Surrey CR0 4YY

Table of Contents

Preface xiii

PART 1: FOUNDATIONS 1

Chapter 1. Contemporary Theories and Concepts in the Psychology of Emotions 3
Géraldine COPPIN and David SANDER

1.1. Introduction. 3
1.2. Emergence of a scientific approach to emotions 4
1.2.1. The emotional sequence: James-Lange versus Cannon-Bard 4
1.2.2. Schachter's two-factor theory. 6
1.3. Basic emotions theories. 7
1.3.1. Premises of basic emotions theories 7
1.3.2. Characteristics of basic emotions 8
1.3.3. Criticisms of basic emotions theories 10
1.4. Bi-dimensional theories of emotion 11
1.4.1. Premises of bi-dimensional theories of emotion. 11
1.4.2. Criticisms of bi-dimensional theories of emotion 12
1.5. Appraisal theories of emotions 14
1.5.1. Premises of appraisal theories of emotion 14
1.5.2. Specific models of this theoretical trend 15
1.5.3. Criticisms of appraisal theories of emotion. 18
1.6. Conclusion 19
1.7. Glossary. 20
1.8. Bibliography 21

Chapter 2. Emotion and the Brain 33
Andy CHRISTEN and Didier GRANDJEAN

2.1. Introduction 33
2.1.1. Emotions and the brain: the emergence of affective neuroscience as an independent discipline 34
2.2. The major role of affective neuroscience in understanding emotions 35
2.2.1. Emotion and the brain: from a unitary entity to processing, from structure to neural networks 36
2.2.2. Levels of processing in emotional processes 38
2.2.3. Emotion and cognition 39
2.3. The historical and conceptual legacy of early conceptions of emotions and the brain 40
2.3.1. Forerunners of affective neuroscience 40
2.3.1.1. Charles Darwin 40
2.3.1.2. The James-Lange peripheralist theory 41
2.4. Initial neuro-anatomical emotion theories 41
2.4.1. Canon-Bard's centralist theory 41
2.4.2. Papez's circuit 42
2.4.3. MacLean's limbic theory 43
2.5. Structures in the brain and their functions in emotional processes . . . 44
2.5.1. Amygdala 44
2.5.2. Amygdala and emotional learning processes 45
2.5.2.1. Amygdala and classical conditioning 45
2.5.2.2. The amygdala: a structure with two processing streams 46
2.5.3. The amygdala and emotional perception: hypotheses around the specificity of processing within the amygdala 48
2.5.3.1. The amygdala as a fear module 48
2.5.3.2. The amygdala and arousal 50
2.5.3.3. The amygdala as a relevance detector 51
2.5.4. The amygdala and memory processing 52
2.6. The prefrontal cortex 53
2.6.1. The prefrontal cortex and bodily signals 55
2.6.2. The prefrontal cortex and the top-down regulation of behavior . . . 56
2.6.3. The prefrontal cortex and the motivational component of emotion 57
2.7. The anterior cingulate cortex 58
2.8. The role of the insula in disgust 58
2.9. Temporal dynamic of brain processes in emotional genesis 59
2.10. Functional connectivity 60
2.10.1. Investigations of the connectivity using brain imaging techniques (MRI) 60

2.10.2. Investigations into connectivity using electroencephalographic (EEG) techniques 61
2.10.3. Benefits of brain connectivity studies 62
2.11. Conclusion 63
2.12. Bibliography 64

PART 2: NON-VERBAL BEHAVIOR 77

Chapter 3. Emotional Corpora: from Acquisition to Modeling 79
Laurence DEVILLERS and Jean-Claude MARTIN

3.1. Introduction. 79
3.2. Building corpora: “acted”, “induced” and real-life emotions. 81
3.2.1. Acted data 82
3.2.2. Induced data 83
3.2.3. Real-life data 84
3.2.4. Comparison of different types of data 85
3.3. Current emotional corpora 86
3.4. Coding schemes 86
3.4.1. Emotional annotation protocols 89
3.4.1.1. Definitions of annotated features. 89
3.4.1.2. Annotation 90
3.4.2. Annotating context 91
3.4.2.1. Validation 92
3.5. Complex emotions in spontaneous data 93
3.6. Applications for corpora 97
3.6.1. Detecting and deciphering emotions in speech 97
3.6.2. Designing an expressive agent from corpora. 98
3.7. Conclusion 100
3.8. Bibliography 101

Chapter 4. Visual Emotion Recognition: Status and Key Issues 107
Alice CAPLIER

4.1. Introduction. 107
4.2. What is a facial expression? 109
4.2.1. Definition 109
4.2.2. Description. 109
4.2.3. Ekman’s universal expressions 111
4.2.4. An ideal system 112
4.3. Overview of facial expression recognition methods 112
4.3.1. Databases. 112
4.3.2. Preprocessing: extracting the face 114
4.3.3. Extracting facial characteristics 115

4.3.4. Classification 117
4.3.5. Performance 117
4.4. Spontaneous facial expressions 118
4.4.1. Position of the problem 118
4.4.2. Databases. 119
4.4.3. Recognizing spontaneous expressions. 122
4.5. Expression intensity 124
4.6. Dynamic analysis 126
4.7. Multimodality 128
4.8. Conclusion 131
4.9. Bibliography 132

Chapter 5. Recognition of Acoustic Emotion 139
Chloé CLAVEL and Gaël RICHARD

5.1. Introduction. 139
5.2. Principles of automatic emotion-recognition systems 140
5.3. Acoustic descriptors. 141
5.3.1. Voiced versus unvoiced content 142
5.3.2. A temporal unit for emotional analysis 143
5.3.3. Prosodic descriptors 143
5.3.3.1. Fundamental frequency (pitch). 144
5.3.3.2. Intensity 145
5.3.3.3. Rhythm descriptors. 146
5.3.4. Voice quality descriptors 146
5.3.4.1. Normalized amplitude quotient 147
5.3.4.2. Frequency modulation (jitter). 147
5.3.4.3. Amplitude modulation (shimmer) 148
5.3.4.4. Rate of unvoiced windows 148
5.3.4.5. Harmonic to noise ration 148
5.3.5. Cepstral and spectral descriptors 149
5.3.5.1. Formant parameters 149
5.3.5.2. Mel-frequency cepstral coefficients 149
5.3.5.3. Bark band energy 150
5.3.5.4. Spectral centroide. 151
5.4. Automatic emotion classification 151
5.4.1. Choosing descriptors 152
5.4.1.1. Normalizing descriptors 152
5.4.1.2. Reduction in data representation space 152
5.4.2. Learning algorithms 154
5.4.2.1. Separators with vast margin 154
5.4.2.2. Gaussian mixture models 155
5.5. Performance and assessment. 157
5.5.1. First factor: data and classes of emotions 157

5.5.2. Second factor: the problem of “ground truths”. 159
5.5.3. Third factor: manual preprocessing 160
5.5.4. Fourth factor: learning algorithms 160
5.5.5. Fifth factor: learning conditions 161
5.6. Conclusion 161
5.7. Bibliography 163

Chapter 6. Modeling Facial Expressions of Emotions 169
Sylwia Julia HYNIEWSKA, Radosław NIEWIADOMSKI
and Catherine PELACHAUD

6.1. Expressive conversational agents 169
6.2. Expressions and their emotional states 170
6.2.1. Expressing discrete emotions 171
6.2.2. Dimensional approaches to emotional expression. 171
6.2.3. Componential expression of emotions 173
6.3. Computational models for facial expressions of emotions 174
6.3.1. A discrete representation of facial expressions 174
6.3.2. Dimensional representation of facial expressions 174
6.3.3. Componential approaches to facial expressions 176
6.3.4. Mixtures of emotions and social constraints 178
6.3.5. Sequences of emotional expressions. 181
6.4. Conclusion 183
6.5. Acknowledgements 184
6.6. Bibliography 184

Chapter 7. Emotion Perception and Recognition 191
Ioana VASILESCU

7.1. Introduction. 191
7.2. Perception in vocal communication of emotion 193
7.3. Experimental paradigms and emotion-oriented automatic systems . . . 194
7.3.1. Experiments validating emotional content and/or
annotation strategies 195
7.3.1.1. Experimental design and stimuli 195
7.3.1.2. Test populations 196
7.3.1.3. Perceptual categorization into emotional classes 196
7.3.1.4. Choosing emotional labels 197
7.3.1.5. Discussion 199
7.3.2. Tests for validating measurable parameters
of emotional information 202
7.3.2.1. Discussion 205
7.3.3. Tests comparing human and automated emotion recognition 205
7.3.3.1. Discussion 208

7.4. Conclusion 208
7.5. Bibliography 209

PART 3: FUNCTIONS 215

Chapter 8. The Role of Emotions in Human−Machine Interaction 217
Valérie MAFFIOLO and Magalie OCHS

8.1. Introduction. 217
8.2. Interactive information and assistance systems. 219
8.2.1. Uses of emotions in interactive systems 220
8.2.2. Current research and tools. 220
8.2.2.1. Collecting and annotating expressions of emotions. 221
8.2.2.2. Automatic emotion recognition 222
8.2.2.3. Interface expressivity 223
8.3. Video games 227
8.3.1. The importance of emotions in video games 228
8.3.1.1. Emotions in gaming systems 228
8.3.1.2. Player emotions. 229
8.3.2. Current research and tools. 231
8.3.2.1. Towards believable emotional virtual characters 231
8.3.2.2. Towards games that account for player emotions. 233
8.4. Intelligent tutoring systems (ITS). 235
8.4.1. The importance of emotions in ITS 235
8.4.1.1. The learner's emotions. 235
8.4.1.2. Emotions in ITS 235
8.4.2. Current research and tools. 235
8.5. Discussion and research perspectives 237
8.6. Bibliography 238

Chapter 9. Music and Emotions 247
Donald GLOWINSKI and Antonio CAMURRI

9.1. The growing importance of music in society 247
9.2. Recognizing emotions and structural characteristics in music 249
9.2.1. Understanding listeners' emotional reactions 249
9.2.2. A categorical or dimensional approach? 250
9.3. Rules for modeling musical expression of emotions. 251
9.4. Towards a continuous measure of emotional reactions to music. 252
9.5. Multimodality in musical experience. 253
9.5.1. A multimodal research platform for musical expression 255
9.5.1.1. Level 1 257
9.5.1.2. Level 2 258
9.5.1.3. Level 3 259

9.5.1.4. Level 4 259
9.5.1.5. Overview. 259
9.6. Multimodal emotional synthesis in a musical context 260
9.7. The social active listening paradigm: the collective aspect of emotion 262
9.7.1. Example: Mappe per Affetti Erranti 262
9.8. Conclusion and perspectives. 263
9.9. Bibliography 263

Chapter 10. Literary Feelings in Interactive Fiction 271
Marc CAVAZZA and David PIZZI

10.1. Introduction: emotions and feelings. 271
10.2. French novels and the representation of feelings 273
10.3. *Madame Bovary*: plot and scenes 275
10.4. Interactive fiction and emotional planning. 280
10.5. Linguistic interaction and emotions. 284
10.6. Emma Bovary's virtuality 290
10.7. Conclusion 294
10.8. Bibliography 295

Chapter 11. The Design of Emotions: How the Digital is Making Us More Emotional 299
Annie GENTÈS

11.1. Representing, interpreting and evoking emotions. 299
11.2. Emotion, *mimicry* and technical devices 301
11.2.1. Representing emotions and catharsis. 301
11.3. Devices as an alternate source of emotion: photography 301
11.4. Art and computers: formal beginnings 303
11.5. The human behind the mechanics and the mechanics behind the human 305
11.6. Mirror interaction as an emotional vehicle. 307
11.7. *Trompe l'œil* versus explicit expression 309
11.8. Three-dimensional universes: an empathetic experience 311
11.9. Empathy and identifying emotions 315
11.9.1. From empathy to shared emotions 315
11.10. Making human–machine interaction and dialog effective 317
11.11. Conclusion: "revenge of the emotions". 318
11.12. Bibliography. 318

List of Authors 321

Index 325

Preface

This book is composed of a series of studies focusing on emotions, and in particular on human−machine interaction systems. Interaction systems are no longer simply efficient, precise and fast information units. They aim to provide the end-user with an emotional experience (the iPhone being a good example). They need to consider and even act on end-users' emotions (such as "serious games", which are designed to create an optimum learning environment) and to display emotions (through Embodied Conversational Agents). This new generation of systems can detect the user's emotional state, can identify trigger factors, act upon these emotions or even express an emotion.

Emotions are a complex phenomenon. They involve the evaluation of an event according to several dimensions including physiological, physical and cognitive changes. Emotions have been studied for a number of years through which several theoretical models have been explored, although there is no overall consensus on their various elements. Emotional theories also differ with regard to the (understanding of the) relation between evaluating an event and producing a physical expression and even the concept of emotional expression as a regulator or a purely social expression. However, the link between the cognitive and the physical aspects of emotions is also highlighted by a number of theories.

Emotions have been involved in the world of technology for a number of years. *Affective computing* (so-called by Rosalind Picard in 1997) addresses the computational modeling of emotions in human−machine interaction while also focusing on perception, cognition and generation. It aims to provide a machine with emotional intelligence, i.e. to provide it with the capability to monitor the end-user's emotions by detecting and expressing them. In this way, recent years have seen the development of emotion-recognition technology based on analyzing visual, acoustic, physiological and even textual signals. The user's emotional state is detected and considered during their interaction with the machine which then displays these emotions. The machine can therefore reason about the implications of carrying out a task, the task itself, solving a problem, making a decision, etc.

Emotional computational models require a theoretical basis in emotional models. This allows us to determine the choice of variables to be studied, to formalize complex phenomena and capture their intensity. Indeed, these computational models are becoming increasingly effective; they can test theoretical models, even predict specific results and validate theoretical models. This double exchange between theoretical and computational models is therefore of mutual benefit.

Emotion-oriented systems, by their very nature, are a multidisciplinary area of study. This book does not attempt to provide an exhaustive overview of the topic, but rather to examine a series of areas, ranging from psychology to technology and art, which explore the subject. The aim of this book is to provide the necessary foundations for anyone wishing to study this area of research. Each chapter is dedicated to a different area of study in the domain, which can serve as a pointer to further study. We have chosen to cover several areas to underline the diversity of emotional studies in emotion-oriented systems as well as demonstrate the importance of accounting for emotions, theoretical models and natural data when building these systems.

This book is composed of four parts, each comprising several chapters. Part 1, “Foundations”, is made up of two chapters focusing on theoretical models and affective neuroscience. Part 2, “Non-verbal behavior”, composed of five chapters, will examine the annotation, analysis and synthesis of emotional acoustic and visual behavior. Finally, Part 3, “Applications”, is composed of four chapters examining the role of emotions in human–machine interaction in the arts: in music, literary narration and, lastly, visual arts.

Chapter 1, “Contemporary Theories and Concepts in Emotional Psychology”, written by Géraldine Coppin and David Sander, offers an historical overview of various trends in emotional psychology. These trends are placed within their scientific context, indicating their significance in current research. This chapter also presents different contemporary theoretical models and their concepts. Three theories are further explained: those of basic emotions and bi-dimensional appraisal theories, for each of which their premises and characteristics are detailed. This is also accompanied by a critical analysis, demonstrating their individual limitations. These models, referred to throughout this book, are followed by a glossary of key concepts.

Chapter 2, “Emotion and the Brain”, by Andy Christen and Didier Grandjean, examines the relationship between the cerebral mechanisms underpinning the emotional process. It also examines the relatively new area of affective neuroscience. Following an historical overview of pioneering works, the chapter focuses on the brain's structure and functions in emotional processes. This is illustrated looking at various parts of the brain (e.g. the amygdala or the prefrontal cortex) active in specific emotions, their role in the cognitive process, and the development of memories or even emotionally-linked physical behavior. These studies are based on a large number of neuroscientific techniques.

The following chapter, "Emotional Corpus: From Acquisition to Modeling", by Laurence Devillers and Jean-Claude Martin, introduces Part 2 of the book. This chapter underlines the importance of using corpora to construct models based on observed data. It explains the different stages involved in creating and analyzing a corpus. In this way, it also highlights the difference between simulated, induced and experienced emotional data. The chapter then offers a summary of existing *corpora*. The annotation of *corpora* involves the creation of a precise coding system for several levels of abstraction and temporality. The importance of validating annotations and providing several measures for this purpose is also presented. The chapter concludes by presenting several applications originating from *corpora* annotations.

In Chapter 4, "Visual Emotion Recognition: Status and Key Issues", Alice Caplier presents the latest practices in the field by analyzing work on emotion recognition analysis. The chapter begins by defining facial expressions and introducing Facial Action Coding System (FACS). Analyzing facial expressions requires the ability to detect a face in an image, which therefore means that we need to be able to locate and extract the facial characteristics. As with Chapter 3, there is a distinction between simulated and natural data, which raises the problem of classifying the expressions that need to be recognized. These expressions no longer correspond to the finite set of so-called six basic emotions. Furthermore, facial expressions are not static images − they are in fact dynamic and their temporal evolution is vital for recognizing the emotions they convey. Finally, the chapter concludes by examining multi-modal emotion recognition.

While Chapter 4 examines facial expressions, Chapter 5, "Acoustic Emotion Recognition" by Chloé Clavel and Gaël Richard, examines acoustic indicators in emotion. It details a range of high-level voice descriptors, such as prosody and voice quality, as well as low-level spectral and cepstral descriptors, which are defined and accompanied by their mathematical representations. Automatic classification of emotions using acoustic descriptors is carried out using learning algorithms. It is essentially achieved using supervised methods, i.e. the data (in this case acoustic) are sent to the learning program at the same time as the emotional labels that we want the machine to recognize. The chapter concludes by detailing the factors used to evaluate and compare acoustic recognition models.

The preceding chapters examine emotion analysis and recognition; the next will deal with synthesis.

Chapter 6 “Modeling Emotional Facial Expressions”, written by Sylwia Julia Hyniewska, Radoslaw Niewiadomski and Catherine Pelachaud, raises the question of how to generate emotional behavior in virtual agents. The chapter begins by presenting the three main approaches currently employed and explains how they differ in dealing with emotional facial expressions, followed by a synopsis of facial expression computational models. This overview is structured according to the emotional representation models on which these models are based. Facial expression models describing blends of emotions are also explored. As previously highlighted in Chapter 4, emotions are not expressed by static images but by dynamic sequences of expressions. A computational model of multimodal expression sequences concludes the chapter.

While Chapters 4 and 5 focus on automatic emotion recognition, Chapter 7, “Emotional Perception and Recognition” by Ioana Vasilescu, examines human perceptions of emotion. This chapter examines experimental perception studies used to determine objective characteristic emotional indicators and to understand how human perceptions can be exploited to identify emotions according to set categories. Chapter 7 focuses primarily on the voice, examining acoustic and prosodic indicators. Lexical and dialogical indicators are also studied in relation to voice quality. The results of perception studies can serve as a basis for automatic classification models.

Part 3 consists of four chapters, dedicated to human−machine interaction, music, interactive narration and visual art, respectively.

Chapter 8, “The Role of Emotions in Human−Machine Interaction”, by Magalie Ochs and Valérie Maffiolo, examines the notion of emotional intelligence within the framework of various types of applications involving human−machine interaction. The chapter analyzes how these applications can assess the emotions of the user and the machine. These applications indicate that emotions do not have the same function. Accounting for user emotions in interactive systems allows the system to have a better understanding of the user in order to better adapt to his or her needs. An area in which emotions play a particularly important role is that of “serious games” and video games. The emotions expressed by characters allow the game to feel more life-like and believable. In addition, accounting for the player’s emotions can increase their engagement in the game by stimulating and/or controlling specific emotions.

Chapter 9, “Music and Emotions” by Donald Glowinski and Antonio Camurri, addresses the theme of communicating emotions through music in a way that allows both the composer and listener to feel them. The influence of music on listeners has also been studied. The methods used to determine listeners' and interpreters' emotional states frequently rely on questionnaires. Emotions are represented according to diverse emotional theories (presented in Chapter 1) and it should be noted that in recent years the study of multimodality in musical experiences has made head way. Studies have shown that emotions are not only communicated by sounds, but also by listeners' multimodal behavior, creating a multisensory experience. The authors present a technical platform that allows analysis of musicians' multimodal behavior. The chapter concludes by presenting musical installations in which participants can interact with musical subject matter, creating a new social listening paradigm.

Another view of emotions is presented in Chapter 10, "Literary Feelings: the Psychology of Characters in Interactive Fiction", by Marc Cavazza and David Pizzi. The authors focus not on communicational emotions, but the emotions experienced during cultural experiences, particularly during interactive narration. They use the word "feeling" rather than "emotion", arguing for the necessity of narrative linguistic analysis to understand the nature of feelings. A study of the *Madame Bovary* narrative using drafts by Gustave Flaubert examines the ontology of feelings felt by the narrative's characters. The second half of the chapter demonstrates an interactive narrative model in which the virtual characters' feelings guide their actions, which, in turn, modifies their feelings. This is illustrated by the example of an interactive narration system based on *Madame Bovary*.

The last chapter of this volume, Chapter 11 "Designing Emotions and how Digitalization Awaits Us" by Annie Gentès, poses the problem of representing emotions with regard to the emotional potential of digital media. To better understand this hypothesis, the author focuses on media such as painting, photography and video, highlighting how interactive devices question how our own identity relates with the machine and our emotions when we interact with it. The creators of these devices rely on two premises: the subject of the work itself and the spectator's dynamic interaction with it. Interacting in virtual worlds therefore offers an emotional experience based on actions, objects and virtual characters and our personal actions in this virtual world. Interactive works seek a form of empathy from their audience.

I would like to thank Radoslaw Niewiadomski, in particular, for helping me finalize this edition of this book, and Sylwia Hyniewska for her comments.

Catherine PELACHAUD
October 2011

PART 1

Foundations

Chapter 1

Contemporary Theories and Concepts in the Psychology of Emotions

Is emotion more than a chapter heading?
Bentley, 1928

Everyone knows what an emotion is,
until asked to give a definition.
Then, it seems, no one knows.
Fehr and Russell, 1984

1.1. Introduction

Following the decisive change of the 1980s, which was marked by a veritable explosion in the scientific study of emotions, emotion is now considered to be a determining explanatory factor in human behavior. In this context, it seems important for any area of research into the functioning of psychological processes and their modeling, as is the case in informatics, to consider current theories and concepts in the psychology of emotions [SAN 09a, SAN 09b]. The central role of emotion in the cognitive system is illustrated by the fact that emotion occupies a "privileged status" in the human brain [DAV 04]; the majority of psychological mechanisms are either necessary for emotion, as such, or are influenced by emotion or involved in regulating emotions. However, as the above quotations suggest, conceptual precision of emotion remains a problem [ALV 02], so much so that there

Chapter written by Géraldine COPPIN and David SANDER.

Emotion-Oriented Systems Edited by Catherine Pelachaud

are almost as many definitions of emotion as there are theories and/or researchers working on this topic [KLE 81, STR 96]. In this introductory chapter, we will present the main contemporary theories and concepts of emotion, touching on the notions established towards the end of the 19th Century in which modern approaches are rooted. For each major current theory, we will present its premise, its main assumptions and its characteristics as well as the criticisms that have been aimed at it. As we will see, there is huge variation in the contribution to these theories towards discovering the functional architecture of emotional mechanisms which consists of characterizing the functional sub-systems responsible for the different stages of processing needed to produce emotions and specifying the organization and interaction of these sub-systems [FEL 05, KOR 11, SAN 02].

1.2. Emergence of a scientific approach to emotions

1.2.1. *The emotional sequence: James-Lange versus Cannon-Bard*

The first "scientific" emotional theory was proposed by William James in 1884 [JAM 84] and Carl Lange in 1885 [LAN 85], who each separately proposed a revolutionary "peripheralist" approach to emotion. This first theoretical conception was the focus of great controversy concerning the mechanisms responsible for triggering emotional feeling (the "sequence problem" [CAN 77]). According to James and Lange, what had previously been considered as the consequence of an emotion, they now considered to be the cause. According to them, the elicitation of a particular emotion is determined by the perception of a specific peripheral arousal pattern, i.e. we experience fear because we are trembling. It should be underlined that, according to James, this definition of emotion is only applicable to emotions that accompany a "specific physical reaction" and not all emotions, resulting in misinterpretations of this theory in later years. In 1885, Lange [LAN 85], summarized the problem of their theory in the following way:

> "If I begin to tremble because I am threatened by a loaded pistol, does first a physical process occur within me, does terror arise, is that what causes my trembling, palpitation of the heart, and confusion of thought; or are these bodily phenomena produced directly by the terrifying cause, so that the emotion consists exclusively of the functional disturbances in my body?"

Clearly, the second position is the one supported by James and Lange. However, whilst their proposition was in complete conflict with the general consensus and traditional concepts of emotion, these authors recaptured the time-honored idea that each emotion has its own *pattern* of physiological changes [RIM 86]. It should be noted at this point that James' opinions were much more nuanced than this with regard to emotion appraisal theories, as we will see later in this chapter [ELL 94].

In contrast to this, the so-called "centralist" approach to emotion, supported by Cannon [CAN 27] and Bard [BAR 28], sees the triggering of a specific emotion as being determined by processing a stimulus in the "central" nervous system, with the peripheral arousal *pattern* being neither specific nor causal. This theory therefore highlights the importance of the central nervous system, specifically the thalamus, in triggering a given emotion. As such, the physiological changes are not considered to be a cause, but rather a consequence of emotion. Cannon raised numerous objections to the James-Lange theory and carried out a number of empirical studies that aimed to disprove it; for example, he observed that similar diffuse visceral reactions occur with a number of emotions as well as in non-emotional states (i.e. digestion or fever); he also reported that suppressing visceral afferents did not suppress emotions [CAN 27]. However, as Fraisse [FRA 63] has notably found, the criticisms raised by Cannon were not without fault and therefore did not warrant an outright rejection of the James-Lange theory [FEH 70].

The theories of James-Lange and Cannon-Bard, both based on a physiological approach but radically opposed with regard to their conception of the temporal sequence of emotion, have had a considerable impact on research of emotion, on the one hand initiating research into the causal relationship between physiological changes and emotion, and on the other hand, studies into the importance of "cognition" in emotion.

The James-Lange theory has had a revolutionary effect due to the originality of the concepts that it proposed, as well as due to the fact that it was empirically testable and therefore refutable. Their theory has also strongly influenced current researchers of emotion. Today, Damasio's theory of somatic markers [DAM 94] revives the James-Lange idea of the causal role of physical changes in emotion by hypothesizing that there are "somatic markers" i.e. physiological reactions, associated with past emotional events. These markers are supposed to be activated when a new event is processed and would influence decisions in relation to their potential consequences. On the other hand, James' theory could be considered to be the basis of facial feedback theories of emotion, which assumes that facial movements regulate emotional feeling. According to James [JAM 92], "every voluntary and dispassionate display of what we believe between the manifestation of a specific emotion should produce this emotion in us". *Embodiment Theories of Emotion* [NIE 07] have also been influenced by the James-Lange theory since they propose that cerebral representations of a given emotion involve retesting perceptive, motor and somato-sensory components in relation with this emotion.

The importance of the James-Lange/Cannon-Bard debate surrounding awareness of the role of cognition in emotion is illustrated by the work undertaken by Schachter, who is among one of the most influential pioneering contributors in the field of affective sciences.

1.2.2. *Schachter's two-factor theory*

According to Schachter's bifactorial theory [SCH 64], an emotion is determined by an interaction between two components: a physiological arousal and a cognition regarding the recognition of the situation triggering this physiological arousal. As such, physical arousal is considered to be undifferentiated by nature, diffuse and non-specific to an emotion in determining the intensity but not the quality of the emotion. The interpretation of the situation would lead to the identification of the emotion felt.

Schachter and Singer [SCH 62] wrote that "it is cognition which determines whether the physiological state of arousal will be labeled as 'anger', 'joy', 'fear' or 'other'". Schachter and Singer therefore share James-Lange's idea that a physiological arousal is necessary in order for an emotion to be produced while also agreeing with Cannon-Bard that physiological changes are not specific to a particular emotion. (Whilst they do not exclude the possibility of physiological changes differing according to the emotion, they believe that these differences would be too subtle to have a psychological effect.) Note that the temporal coincidence between the two components is not a sufficient condition for triggering an emotion: the person needs to establish a link between the physiological arousal and a relevant explanation for the latter.

Schachter and Singer's famous experiment [SCH 62] is typically cited as fundamental experimental proof of this theory. This has essentially suggested that when a person does not have information that is likely to explain why they are physically stimulated, they will rely on information available in the situation and context in order to make sense of their physiological arousal (which in this experiment was triggered by an injection of epinephrine). However, when the person is not in a particular state of physiological arousal (without an injection of epinephrine), or where he/she has an adequate explanation for it (information on the consequences of the injection of epinephrine), they will not look for factors in the environment that explain this arousal. However, the effects predicted by Schachter and Singer's proposition are not systematically observed (see [REI 83] for a review of experiments in the same vein as Schachter and Singer). On the other hand, the component determining which stimuli trigger the physiological arousal in the first place is not specified; this theory therefore does not explain the emotion-triggering process.

Schachter, by introducing the existence of a cognitive components associated with physiological arousal as a determinant of emotional quality, is clearly one of the forerunners of a cognitive approach to emotions. Furthermore, his theory has the advantage of considering the social dimension of emotion, since emotion stems in part from the use of information taken from the social environment. Another theoretical trend, falling within the evolutionary perspective, has also highlighted the social dimension of emotion by focusing specifically on the communicative function of emotions through their expressions.

1.3. Basic emotions theories

1.3.1. *Premises of basic emotions theories*

Some theoreticians with an evolutionary perspective, who believe that evolution has played a central role in shaping the emotions' characteristics and functions, have stressed the notion of emotions adaptation. These characteristics and functions appeal to programs that govern the body's major systems, such as physiology, the motor system, as well as numerous cognitive mechanisms such as attention, learning or even memory; emotion is therefore a high level-organizing process [COS 00]. Matsumoto and Ekman [MAT 09] define emotions as transient, bio-psychological reactions designated to aid individuals in adapting to and coping with events that have implications for survival and well-being. Within the framework of Izard's theory of differential emotions, emotions constitute the primary motivational system of human behavior [IZA 09]. Following attention on the adaptive character of emotions, some researchers have suggested the existence of a limited number of fundamental universal emotions, each having an evolutionary function: "basic emotions", or alternatively "primary", "fundamental" or occasionally "discrete" emotions [EKM 82, IZA 77, TOM 80].The majority of authors who adopt this approach consider anger, fear, joy, sadness and disgust to be basic emotions, although this is a contentious subject, in particular with regard to surprise. More complex emotions would therefore originate as a mixture of these basic emotions [ORT 90]. This theoretical approach is based on the key discoveries of Darwin in terms of the facial expression of emotions [DAR 72]. In his book, *The Expression of Emotions in Man and Animals*, Darwin describes emotional facial expressions as innate and universal and emphasized not only their communicative function, but also their evolution in relation to the direct environment. However, note that this notion of basic emotions had already been suggested by Descartes [DES 31], who distinguished six primary emotions: admiration, love, hatred, desire, joy and sadness; with all other emotions being composed of these six or other forms of them.

1.3.2. *Characteristics of basic emotions*

Basic emotions, according to Ekman [EKM 92] the current most important proponent of this view, are characterized by a set of properties. Above all, basic emotions are different from other types of emotions such as, for example, pro-social or moral emotions [MAT 09]. However, basic emotions do share common properties. As such, a basic emotion would be present in non-human species, be triggered rapidly and automatically, and appear spontaneously and for a short duration. Furthermore, it has specific trigger conditions. As previously mentioned, if, according to this approach, emotions are considered to have evolved to respond to fundamental tasks for survival that present a phylogenetic adaptive advantage, then it is logical to believe that there are distinct universal trigger events for basic emotions (e.g. the loss of a loved one would be a universal condition triggering grief).

Moreover, a basic emotion has specific autonomous *patterns* (see [EKM 83] for a view on how emotions can be differentiated on the basis of the activity of the autonomic nervous system). However, empirical studies in this area suffer from a major methodological flaw; inducing strong emotions in a laboratory is almost impossible for both ethical and practical reasons. Consequently, the specific and systematic differences for different emotions have rarely been possible to demonstrate. Stemmler, Heldmann, Pauls and Scherer [STE 01], who have studied the psycho-physiological responses induced by fear and anger in an ecologically-valid framework, avoiding this difficulty, report that patterns of emotional responses studied did not overlap. Even if the empirical arguments remain limited overall, it seems justifiable to argue in favor of the existence of typical emotional *patterns*, linked to action tendencies characterizing major emotions such as fear and anger. This reaction of the autonomic nervous system, which may originate in appraisal, is particularly functional if we consider the role of emotion in preparing for actions adapted to a particular situation.

Furthermore, basic emotions have specific neural patterns. Today, the majority of research on emotion in cognitive neuroscience is focused on researching the specific cerebral regions that elicit basic emotions. There is, as such, a corpus of data suggesting that signals linked to fear are processed by the amygdala (see Öhman and Mineka's suggestion of the existence of a "module of fear" [ÖHM 01]), while the signals linked to disgust are processed by the insular cortex [CAL 01]. As such, a major writer in the field of affective neuroscience, Joseph Ledoux, has written that "different classes of emotions are underpinned by separate neural systems" [LED 96]. Equally, reinforcing this modular approach, Öhman [ÖHM 99] concludes that the different emotional systems have different evolutionary histories and deserve to be seen as independent rather than parts of a general area of emotion.

Finally, basic emotions have specific expressive patterns − perhaps their most important characteristic, since neo-Darwinian theories are essentially focused on determining basic emotions by studying emotional facial expressions (see Figure 1.1). Facial expressions are considered by Ekman as the pivot in communication between human beings [EKM 89]. Nevertheless, it seems that, as Darwin proposed, emotional facial expressions are also formed through interaction with our physical environment. As such, Susskind and his colleagues [SUS 08] have shown that the facial expression of fear, in direct opposition to the facial expression of disgust, allows us to increase sensory acquisition (through opening the eyes, nose and mouth). Again, the adaptive value of such configurations seems clear. For Izard and King [IZA 09], expressive behavior is also fundamental in child development because it contributes to the emergence of interpersonal skills and encourages appropriate social behavior. However, Ekman *et al.* are focused on demonstrating the universality of emotional facial expressions in the majority of their work. They have notably been able to demonstrate that Western facial expressions were well recognized in a preliterate New-Guinea culture, which has had very little contact with the outside world, therefore excluding the possibility that the people of this culture had learnt to recognize the expressions of other populations [EKM 71].

Figure 1.1. *Representation of the facial expressions considered to be characteristic of basic emotions. Figure taken from Matsumoto and Ekman [MAT 09] (by permission of Oxford University Press)*

1.3.3. *Criticisms of basic emotions theories*

The notion of basic emotions is a concept that has given way to important discussion in the field of affective sciences and has been the object of rigorous criticism. According to Mandler [MAN 84], restricting the numerous possibilities that can result from the appraisal of events, a mechanism supposed to be responsible for emotional genesis, to a limited number of emotions makes little sense. Some academics even deny the existence of the categories making up these basic emotions [ORT 90, WIE 92]. Scherer [SCH 93b] adopts a much more nuanced position in his view that the structure of emotions is not only more flexible, but also much more varied, as is proposed within the basic emotions theories framework. He also suggests the term "modal emotions" to designate frequently-triggered emotions following other similar views of appraisal, a process again considered to be an emotional trigger. Nevertheless, as will be made clear in the rest of this chapter, the depth of emotional states that his model proposes is much more vast.

Furthermore, basic emotion theoreticians typically maintain that emotional facial expressions result from neuromotor programs that trigger the emotional facial expressions in their entirety, with the different facial muscles employed for this expression having their apex simultaneously. In contrast with this view, some supporters of appraisal theories of emotion believe that the results of appraisals can be linked to specific facial movements [SCH 92, SMI 97]. As such, evaluating an event as an obstacle to our goals would be linked to activity in the *corrugator supercilii* muscle, which allows frowning. Thus, it is possible to predict which facial changes will be produced as a result of specific types of evaluations. These changes will be added to the appraisal process as and when they happen and could produce a final configuration similar to the prototypical configurations proposed by basic emotions (see Figure 1.1). The conceptualization of the expressive process proposed by Scherer and Smith emphasizes the dynamic nature of emotional facial expressions and enables experiments to better understand the underlying mechanisms in the expression and the recognition of emotional facial expressions [SAN 07, WEH 00].

Besides facial expression, the premise of functionally independent modular systems specific to each emotion has also been criticized. The amygdala could be thought of as a "relevance detector", not specific to fear [SAN 03]. Furthermore, an increasing amount of empirical evidence suggests that the insular cortex is not specifically linked to disgust [SCH 02] but is involved in a number of processes based on interoceptive information (for further details concerning cerebral bases, see Chapter 2).

Let us note finally that even if it seems reasonable to think that peripheral psycho-physiological responses differ, this does not constitute an argument in favor of basic emotions theories. This physiological reaction could come from the appraisal of a situation and not an reflex-like program. Remember, however, that theoreticians of basic emotions have integrated a version of the idea of appraisal by proposing the existence of an *auto appraiser*. This refers to an automated appraisal system, integrated in a "basic emotion system" (see Figure 1.2), proposed as a mechanism triggering emotions [EKM 04b, MAT 09].

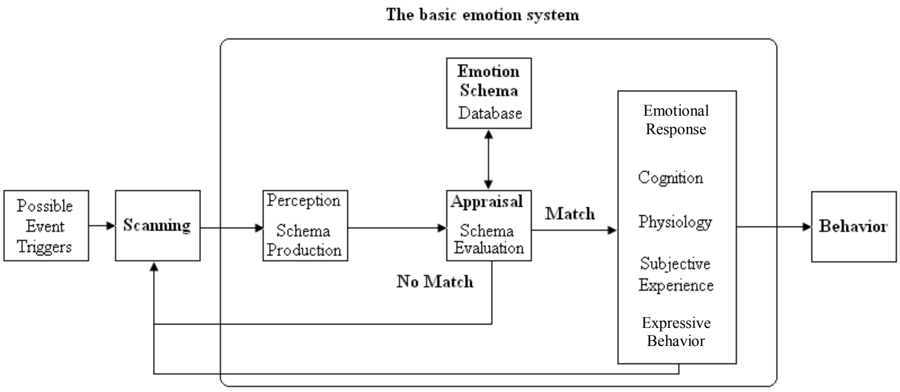


Figure 1.2. *Representation of the basic emotion system [MAT 09] (by permission of Oxford University Press)*

1.4. Bi-dimensional theories of emotion

1.4.1. *Premises of bi-dimensional theories of emotion*

Scientific literature, as is often the case in everyday life, frequently refers to categories such as joy and fear. However, according to Feldman-Barrett [FEL 06], establishing emotional categories such as fear or anger "now presents a major obstacle to understanding what emotions are and how they work". Feldman-Barrett suggests that emotional responses do exist, that they may be functional and are very likely to be the result of evolution. However, this does not necessarily mean that *anger*, *sadness* and *fear* are useful categories. In order to support this critique, Feldmen-Barrett adopted a different research tradition, a dimensional approach, in which the affect is described in relation to independent elementary dimensions, which are core phenomenological properties in affective experience [RUS 99], dimensions that can be combined.

Current dimensional theories are based on the general idea of Wundt's theory [WUN 97] about emotional experience, which identifies three basic dimensions to describe emotional feeling (pleasure/displeasure, excitement/inhibition and tension/relaxation). Emotional feeling may be permanently represented on a more or less important level on each of these three dimensions. According to the model proposed by Russell [RUS 80], it is possible to represent emotions using a circle in which two axes alone are necessary: the dimension of valence indicating pleasure/displeasure and the arousal dimension (weak/strong), which represent the affect as a subjective experience on a continuum [FEL 99], see Figure 1.3. This circular model is called *circumplex* and corresponds to a mathematical formalism to represent the current structure of a group of stimuli around a circle [FEL 09]. Currently this approach is probably the most commonly used for measuring subjective emotional experience [FON 09]. This representation is found in different cultures and is potentially universal, although this is not always confirmed by empirical data [TER 03].

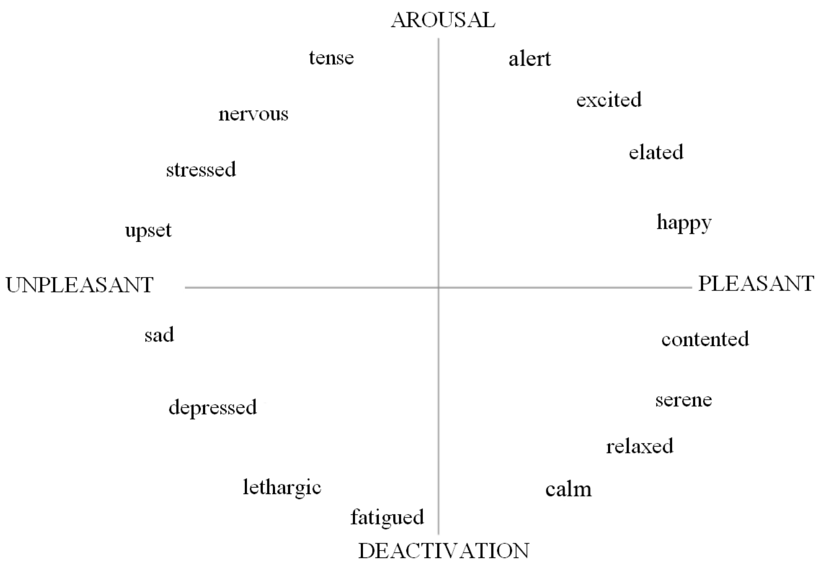


Figure 1.3. *Representation of the circumplex model with the horizontal dimension indicating valence and the vertical dimension indication arousal [FEL 09] (by permission of Oxford University Press)*

1.4.2. *Criticisms of bi-dimensional theories of emotion*

Some have criticized the ability of this model to differentiate between emotions; for example, fear and anger are found in the same place in the circle because both of these emotions are particularly negative and intense. However, on a subjective, expressive and behavioral level, these two emotions are very different. Being based on verbal reports, this breakdown considers that the affect's structuring overlaps the structure generally underlying language. This model, which seeks to distance itself from common language categories such as fear or anger, therefore displays a paradox.

On the other hand, there is no real consensus on the elementary dimensions proposed in different dimensional models, and some authors include, in addition to the dimensions of valence and arousal, a dimension of control [OSG 62]. For others, it is the dominance dimension, conceptually very close to that of "control" or "coping", which should be included [MEH 96]. Others, such as Duffy [DUF 41] propose that emotions can be reduced to the notion of "energy level". Fontaine, Scherer, Roesch and Ellsworth [FON 07] have however demonstrated that for English, French and German, four dimensions are necessary to satisfactorily represent the similarities and differences in meaning of emotional words when the different emotional components are accounted for and not simply verbal accounts. The majority of authors who currently adopt this reduced set of complex data representing the emotional experience in elementary dimensions focus on the two dimensions of valence and arousal. According to Russell [RUS 03], "the finding of two broad dimensions is so ubiquitous, and current descriptive models so similar, that the word consensus is now appearing in writings on this topic (Watson and Tellegen, 1985)".

Furthermore, Cacioppo, Gardner and Berntson [CAC 97] propose that the evaluative processing of stimuli with regard to their valence is not bipolar (and therefore not a continuum), but that distinct evaluative and motivational systems underlie the evaluation of the positive or negative character of a stimulus. Beyond these alternative approaches, some researchers, such as Watson and Tellegen [WAT 99], call into question the validity of the bi-dimensional approach to emotion since "Russell's own data indicated significant problems with the model". Russell and Caroll [RUS 99] responded to this criticism by admitting that their model was only an approximation but still a "convenient and heuristic one, the best we know of". The heuristic character is also debatable, this approach being centered on a specific aspect of emotion, namely subjective feeling, and only makes brief reference to other, no less trivial, emotional components, such as elicitation mechanisms. However, "emotions are adaptive responses to the world, not simply abstract sensations, as dimensional theories insinuate" [ELL 03].

One of the major problems with this approach is probably the lack of agreement on a clear definition of the concepts used. Reisenzein [REI 94] highlights that there are at least four different theoretical viewpoints with regard to emotional intensity within bi-dimensional theories of valence/arousal, which assume the independency of these two dimensions. The most common position argues that emotional quality is determined by a combination of specific values of pleasure/displeasure and arousal/calm; this however does not explain emotion intensity, nor does it account for emotions characterized by the same combination valence/arousal but with different intensities. The second position, which could be seen as an attempt to compensate for this problem, adds the dimension of intensity to that of valence and arousal; nevertheless, this does not prove more satisfactory because this intensity dimension is unspecified (is it the intensity of pleasantness, arousal or both?). A third conceptualization suggests that emotional quality is determined by the valence dimension while emotional intensity is determined by the arousal dimension. Finally, a fourth proposition − which could itself be further divided into two sub-propositions − hypothesizes that both emotion quality and intensity will be determined by the valence and arousal dimensions. Beyond this conceptual ambiguity, the significance given to some concepts is neither justified theoretically nor empirically. For example, the *core affect*, defined by Russell [RUS 03] as a neurophysiological state that can be unconsciously accessed as a feeling reflecting an integral mixture between the feelings of valence (pleasure/displeasure) and arousal (asleep-activity) is conceptualized as "primitive" by Russell and Feldman-Barrett, although no empirical study has proven this proposition.

1.5. Appraisal theories of emotions

1.5.1. *Premises of appraisal theories of emotion*

The appraisal approach towards emotions was developed along two lines: the first conceptualizes emotion as an information-processing system just like any other mechanism [OAT 87, THA 02]; the second approach supposes that an evaluative type of cognitive processing (appraisal) is at the root of emotional elicitation. It is this second approach that we will examine further.

Despite the popularity of Ekman's view of basic emotions, appraisal theories of emotions dominate the field of study in terms of how emotions are generated and differentiated from one another. These theories propose in effect that the evaluation that the individual makes about a stimulus, event or situation determines the elicitation of an emotion [SAN 05, SIE 07]. These models propose that organisms constantly explore their environment, reacting to relevant stimuli. The major contribution of these theories has been the specification of a set of standard criteria that are presumed to underpin the emotional appraisal process. When an event occurs, an individual will evaluate the relevance of this event based on a specific number of criteria.

Most theoreticians agree on the evaluation criteria of novelty, intrinsic pleasantness, predictability, goal-relevance, the possibility of managing consequences of the event (or *coping potential*) and the compatibility with personal or social norms (or *normative significance*) [ELL 03]. The combinations of these evaluations, which are often automatic and unconscious [KAP 01, MOO 07, MOO 09, MOO 10], can cause different emotions. The idea that each different evaluation structure corresponds to a different emotion was originally proposed by Arnold [ARN 60], considered to be the pioneer of appraisal models. Her idea has since been adopted and elaborated by other scholars [FRI 86, ROS 84, SCH 84b, SMI 85] and has contributed much to the significance attached to the "appraisal" mechanism in affective sciences today. It currently seems well established that different emotions can be described in terms of different *patterns* of appraisal with various experiments corroborating this idea [FRI 87, SCH 93b, WEI 85]. Note that, in 1894, William James declared that "as soon as an object has become thus familiar and suggestive, its emotional consequences, *on any theory of emotion*, must start rather from the total situation which it suggests than from its own naked presence".

Appraisal [ARN 60, LAZ 66, LAZ 84], is the most commonly used term to describe the cognitive process which elicits emotion. Lazarus equally talks of "transaction" [LAZ 84] because he proposes that emotions stem from the mutual influence of a person and their environment [LAS 78]. Lazarus [LAZ 68] conducted a series of studies aiming at discovering the determinants of the appraisal process, following which he identified three components: primary appraisal, secondary appraisal and re-appraisal. Primary evaluation involves recognizing a stimulus and its significance for the individual's well-being while secondary appraisal is related to analyzing the resources that individual can use to respond to the situation. Re-appraisal is a process allowing modifications to the primary and/or secondary appraisal while the interaction between the individual and their environment occurs. Lazarus' proposition that there is significant variability in interpretation of the environment according to individuals, resulting in emotions being expressed with a large degree of variability, is found in modern appraisal theories. His idea that emotion is a continual process is also widespread today.

1.5.2. *Specific models of this theoretical trend*

Frijda [FRI 06] considers emotions to be triggered by significant events that are defined as occurrences that touch on one or more of the subject's concerns. Note that the word "concern" is here used to mean the mindset of someone who is particularly interested in someone or something, engaged in something or wants to reach a specific goal.

One of Frijda's contributions to the study of emotion is precisely this notion of "concern", defined as a disposition to desire the occurrence or not of a given type of situation. As such, Frijda considers that emotions are based on evaluating the relevance of events [FRI 86]. According to Frijda, the emotional process is made up of different processing stages that lead to an action tendency that is the main component of emotion in his eyes. Action tendency then allows us to respond to the situation [FRI 86]. Frijda considers that emotion involves states of preparation for action that are induced by evaluating events as important to the individual.

For Scherer [SCH 89], who elaborated on the component processes model, the appraisal process is a rapid succession of stimuli-processing stages (also known as *stimulus evaluation checks* or *SEC*, see Figure 1.4) that are at the root of emotion genesis. Emotion is defined as "an episode of interrelated, synchronized changes in the states of all or most of the five organismic subsystems in response to the evaluation of an external or internal stimulus event as relevant to major concerns of the organism" [SCH 01a]. In this context, we will favor the term "emotional episode" rather than "emotional state" to emphasize the dynamic nature of emotion [SAN 09c]. Scherer's definition, which emphasizes the notion of synchronized changes in an organism's various sub-systems, allows us to rethink the emotional sequence (as discussed earlier in this chapter) as a question of dynamic interaction between the five components of emotion. This definition also has the benefit of not referring to only one aspect of emotion, which has been a recurrent problem in previous descriptions of emotion. Scherer [SCH 84] proposes that emotion is multidimensional and comprises five components:

a) a stimulus- or situation-appraising component enabling the elicitation and differentiation of emotions, an integral part of the emotional process [CLO 00].

However, as Lazarus has indicated [LAZ 91], there are other components of emotion other than appraisal, such as:

b) a physiological component (bodily changes);

c) a motor expression component, both in terms of facial and vocal expression as well as posture and gestures;

d) a motivational component including action tendencies (e.g. approaching or avoiding); and finally

e) a subjective feeling component reflecting felt emotional experience [SCH 93a, SCH 01].

It seems important to stress this last component, because the terms "emotion" and "feeling" have often been used interchangeably in the past. Wundt [WUN 97], for example, considered that there is no clear demarcating line between feeling and emotion. The majority of researchers in affective sciences agree that feeling is one component of emotion and not emotion as such.

One the most current and influential approaches in the psychology of emotion today is encapsulated by appraisal theories of emotions. As with all basic emotion theories, this approach underlines the adaptive function of emotions. If we compare these two theoretical trends, it seems that basic emotion theories have, in recent years, been influenced by notions from appraisal theories. For example, the significance attached by Scherer to the synchronization of emotional components is also found in the work of Matsumoto and Ekman [MAT 09] and their notion of coordination. They hypothesized that because emotions have developed to help humans to prepare for action, the responses associated with emotion (physiology, expressive behavior, cognitions and feeling) need to be *organized* and *coordinated*. Equally, Ekman [EKM 04a] talks about *auto-appraiser* mechanisms. According to him, we need to have automatic appraising mechanisms which continually scan the world around us, detecting when something important for our well-being or survival occurs. Ekman's acceptance of the existence of an *appraisal* process, whether voluntary or automatic, is not new [EKM 77], the formalization of this process within basic emotions theories has never achieved the kind of precision proposed by appraisal theories. The notions that (1) an event needs to be important to the individual to induce an emotion; and (2) the speed of the evaluative process (which can occur rapidly and unconsciously); are two key concepts in appraisal theories of emotions [GRA 09] that seem to be accepted in recent versions of basic emotion theories.

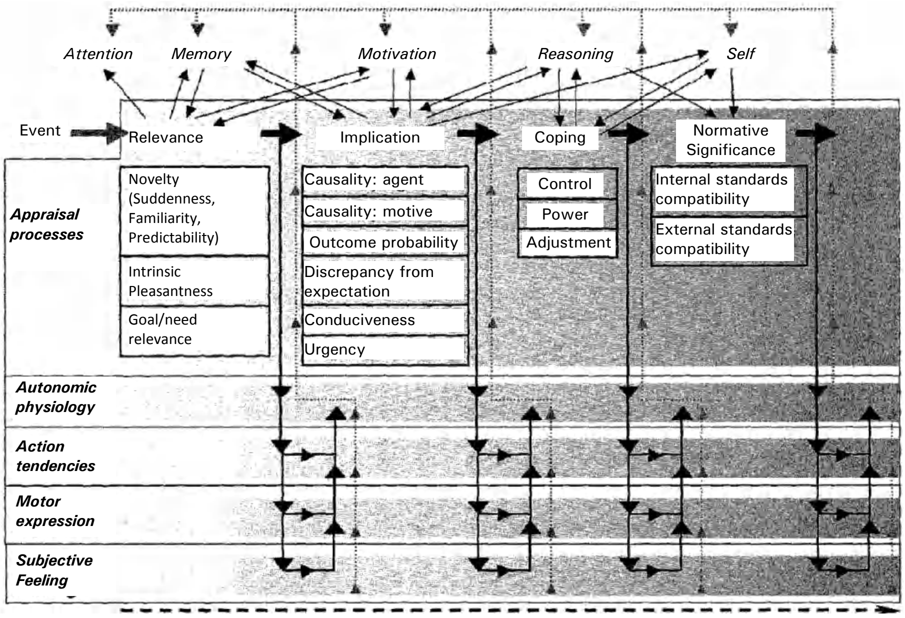


Figure 1.4. *The component processes model*

However, in contrast to basic emotions and bi-dimensional theories that say little about emotional trigger mechanisms, emotion appraisal theories have contributed greatly to our understanding of genesis governing mechanisms and differentiating between emotions. These theories have also been highly useful for understanding intra- and interpersonal differences, notably cultural and social, in the triggering and differentiation of emotions.

1.5.3. *Criticisms of appraisal theories of emotion*

The main criticism of these theories is their potentially excessive cognitivism. As such, while researchers opposed to this approach accept that appraisal theories explain specific types of emotional reaction, they deny the necessity of this process and propose that in a number of cases emotions are produced by non-cognitive factors [BER 93, LED 93, OHM 78, ZAJ 84]. A key question concerning the "emotion/cognition" debate lies in the definition of the term "cognition" (see the glossary in section 1.7). If we accept that appraisal can occur rapidly on several cognitive levels, including automatic and unconscious processes, then the criticism of excessive cognitivism no longer holds up (see [LEV 87] for their proposition of levels of processing in the appraisal process [MOO 09, MOO 10, ROB 09]).

Another criticism of these theories concerns emotional disorders. For example, for a sufferer of arachnophobia, acquiring the *explicit knowledge* that a spider is harmless does not stop them from being scared [GRI 04]. The notion of levels of processing therefore seems equally important for responding to this criticism. However, it should be noted that this "multi-level" approach to the appraisal process has also been criticized because the levels proposed could be seen as being too diverse to cover the general concept of appraisal [GRI 04]. One route to explore, which may allow us to understand the different levels of processing in phobias, is perhaps that of divergences (or similarities) in appraising the same experienced, remembered or imagined event without specific information or an imagined event when informed that this situation is extremely likely to make us fearful.

Olsson, Nearing and Phelps [OLS 07a] and Olsson and Phelps [OLS 07b] have underlined the importance of social learning. According to them, witnessing someone being frightened by a given stimulus could trigger a fear response just as strongly than if we had ourselves experienced this stimulus. Today, and to our knowledge, no study has systematically compared the above-mentioned four conditions. The experimental studies that have taken place have focused on the appraisal process as an emotional trigger by remembering past emotional events or imagining them (for further information on the methods used, see [SCH 01b]). For both ethical and practical reasons in experimental psychology (for further information on this question, see part one of the *Handbook of Emotion Elicitation* *and Assessment* [COA 07, KUP 07a]), methods of inducing natural emotions are not systematically used in the psychology of emotions. However, the combination of experimental techniques for manipulating and measuring responses enables precise hypotheses regarding the appraisal process to be empirically tested [GRA 08, LAN 07, SIE 07].

1.6. Conclusion

In 1885, Lange considered that it was possible to affirm without exaggeration that, scientifically, absolutely nothing was understood about emotions, and that there was no shadow of a theory on the nature of emotions in general, nor of each emotion in particular. A century later, Frijda [FRI 89] claimed that: "there was not a real theory of emotions. By 'real theory' of emotions, I mean a theory of the human body or biological systems in general, in which emotion has its own place, among other components such as information processing and adaptation." It should not be mistaken, however, that during the century that separated these two authors, the study of emotional phenomena was abandoned. The same also applies to the conceptual precision of the term "emotion".

As Plutchik remarked in 1980 that the progress made to find a good definition of emotion was not encouraging, a crucial point in the contemporary approach to emotions therefore consists of defining emotion by breaking it into several components. Scherer therefore has proposed a working definition that restricts the use of the term "emotion" to short periods of time during which the body's sub-systems are synchronized to produce an adaptive reaction to an event or situation that is relevant to that person (see [SCH 87, SCH 93b] for a more detailed discussion of this definition). This definition, by its specificity, allows us to distinguish emotion from other affective phenomena such as desire, humor, preferences, attitudes or even emotional styles, which represents a major conceptual advance.

Emotions have traditionally been analyzed using a limited list of categories. Basic emotion theories assume that there are a limited number of emotions with an evolutionary status and are in this sense "fundamental". In terms of the bi-dimensional approach, emotions are not seen as forming different categories, rather in terms of the elementary dimensions of emotions such as valence or arousal, which help explain and describe them. Finally, within the framework of appraisal theories, the evaluation of a situation or stimulus according to different criteria by an organism is the cause of the elicitation and differentiation of emotions. Consequently, basic emotion and bi-dimensional theories are not suitable for modeling the elicitation of emotional processes.

An area where such progress has had interesting repercussions has been that of emotional modeling [CAN 09, KOR 11, PEL 09, PET 09 PIC 09, TAY 09]. For example, as Sander, Grandjean and Scherer [SAN 05] have remarked, modeling neural networks, while having been very important in terms of its development in the past 30 years, has neglected the study of emotions (with the exception of [TAY 09]). However, an advantage of emotional computational analysis is that it encourages academics in affective sciences to develop explicit functional architecture [CLO 09, GRA 09, MAR 09, SAN 02, SAN 05]. The aim remains to develop emotional models that are sufficiently explicit to be tested using computational, neuroscientific and psychological methods.

1.7. Glossary

Arousal: according to Duffy [DUF 62], arousal refers to "a condition conceived to vary in a continuum from a low point in sleep to a high point in extreme effort or intense excitement". Note that arousal is not synonymous with intensity of affect, although this mistake is often made in academic writing [BAR 99]. Furthermore, its link with peripheral nervous system activity is neither direct nor linear [BAR 04]. Arousal is proposed as a fundamental dimension of emotional experience (with valence) in Russell's *circumplex* model, which belong to dimensional emotional models and is probably the most influential of them.

Cognition: while the definition of cognition has been a matter of fierce debate (for further details, see Lazarus [LAZ 84] and Zajonc [ZAJ 84]), it seems possible with some consensus to define it as a cognitive process that, either naturally or artificially, processes (not necessarily symbolically) information used for the acquisition, organization and use of knowledge in an explicit or implicit way. This definition is broader than that of Neisser [NEI 67], for whom the term *cognition* refers to all processes by which sensory input is transformed, reduced, elaborated, stored, recovered and used. Neisser's definition appears to be limited because it excludes the sensorial processes from the field of cognitive sciences, which is a subject of debate.

Emotion: the definitions of this term are so numerous that it seems impossible to cover them all. Here, we will consider emotion to be "an episode of interrelated, synchronized changes in the states of all or most of the five organismic subsystems in response to the evaluation of an external or internal stimulus event as relevant to major concerns of the organism" [SCH 01a]. The five components that this definition sets out are: appraisal, psycho-physiological changes, motor expression, action tendencies and subjective feeling.

Appraisal process: proposed as being the initiator of the elicitation and differentiation of emotions by emotion appraisal theories. It consists of the rapid and often unconscious evaluation of a stimulus or specific event against a set of cognitive criteria. This process will determine and differentiate the emotion or mixture of emotions felt subjectively.

Subjective feeling: one of the components of emotions for a long time confused with "emotion" itself, is currently theorized notably as a reflection of the changes occurring in the four other components of the emotional process [SCH 05].

Action tendency: Frijda [FRI 86] has defined action tendency as a state of preparation "with the aim of executing a specific type of action". Action tendencies (e.g. the desire to flee when scared) are one of the components in the emotional process, as opposed to actions, and are considered more as behavioral consequences of emotion. Note that there can be predictive action tendency (i.e. with anger, for example, towards a superior) where the effective behavior is a retreat (e.g. leaving a conflict with a superior rather than becoming aggressive towards him).

Valence: a key dimension of emotion (see, for example, Titchener [TIT 09]) already proposed by philosophers such as Aristotle referring to hedonism, i.e. the agreeable versus disagreeable character of an event or situation. Valence is proposed as one of the elementary dimensions of the emotional experience for the quasi-totality of emotional dimensional theories. In appraisal theories, two types of valence are distinguished for a given event: (1) intrinsic pleasantness (pleasant or unpleasant) of the event itself; and (2) the goal-conduciveness (conducive or obstructive) of this event for the individual. For example, a bar of chocolate can have an intrinsic pleasant character but may be goal-obstructive for the individual if he/she likes chocolate but is on a diet.

1.8. Bibliography

[ALV 02] ALVARADO N., ADAMS S.S., BURBECK S., *The Role of Emotion in an Architecture of Mind*, IBM Research, Zurich, Switzerland, 2002.

[AND 03] ANDERSON A.K., CHRISTOFF K., STAPPEN I., PANITZ D., GHAHREMANI D.G., GLOVER G., GABRIELI J.D.E., SOBEL N., "Dissociated neural representations of intensity and valence in human olfaction", *Nature Neuroscience*, vol. 6, pp. 196-202, 2003.

[ARN 60] ARNOLD M.B., *Emotion and Personality*, Columbia University Press, New York, USA, 1960.

[BAR 28] BARD P., "A diencephalic mechanism for the expression of rage with special reference to the central nervous system", *American Journal of Physiology*, vol. 84, pp. 490-513, 1928.

[BAR 99] BARRETT L.F., RUSSELL J.A., "Structure of current affect", *Current Directions in Psychological Science*, no. 8, pp. 10-14, 1999.

[BAR 04] BARRETT L.F., QUIGLEY K., BLISS-MOREAU E., ARONSON K.R., "Arousal focus and interoceptive sensitivity", *Journal of Personality and Social Psychology*, vol. 87, pp. 684-697, 2004.

[BEN 28] BENTLEY I.M., "Is emotion more than a chapter heading?", in: REYMERT M.L. (ed.), *Feelings and Emotions: The Wittenberg Symposium*, pp. 17-23, Clark University Press, Worcester, 1928.

[BER 93] BERKOWITZ L., "Towards a general theory of anger and emotional aggression: Implications of the cognitive neo-associationisitic perspective for the analyzes of anger and other emotions", in: WYER Jr. R.S. and SRULL T.K. (eds), *Perspective on Anger and Emotion: Vol. 6, Advances in Social Cognition*, pp. 1-46, Lawrence Erlbaum Associates, Hillsdale, NJ, USA, 1993.

[CAC 97] CACIOPPO J.T., GARDNER W.L., BERNTSON G.G., "Beyond bipolar conceptualizations and measures: The case of attitudes and evaluative space", *Personality and Social Psychology Review*, vol. 1, pp. 3-25, 1977.

[CAL 01] CALDER A.J., LAWRENCE A.D., YOUNG A.W., "Neuropsychology of fear and loathing", *Nature Neuroscince Reviews*, vol. 2, pp. 352-363, 2001.

[CAÑ 01] CAŇAMERO L., "Emotions and adaptation in autonomous agents: a design perspective", *Cybernetics and Systems: An International Journal*, vol. 32, pp. 507-529, 2002.

[CAÑ 09] CAŇAMERO L., "Autonomous agent", in: SANDER D. and SCHERER K.R. (eds), *Oxford Companion to Emotion and the Affective Sciences*, Oxford University Press, Oxford, 2009.

[CAN 77] CANDLAND D.K., "The persistent problems of emotion", in: CANDLAND D.K., FELL J.P., KEEN E., LESHNER A.I., PLUTCHIK R. and TARPY R.M. (eds), *Emotion*, pp. 1-84, Brooks/Cole, Monterey, CA, USA, 1977.

[CAN 27] CANNON W.B., "The James-Lange theory of emotions: A critical examination and an alternative theory", *American Journal of Psychology*, vol. 39, pp. 106-124, 1927.

[CLO 00] CLORE G.L., ORTONY A., "Cognition in emotion: always, sometimes or never?", in: LANE R.D. and NADEL L. (eds), *Cognitive Neuroscience of Emotion*, pp. 24-61, Oxford University Press, Oxford, 2000.

[CLO 09] CLORE G.L., PALMER J., "Affective guidance of intelligent agents: How emotion controls cognition", *Journal of Cognitive Systems Research*, vol. 10, pp. 21-30, 2009.

[COA 07] COAN J.A., ALLEN J.B., *Handbook of Emotion Elicitation and Assessment*, Oxford University Press, New York, USA, 2007.

[COPa] COPPIN G., SANDER D., “Neuropsychologie affective et Olfaction: Etudier la sensibilité de l’amygdale aux odeurs pour tester les théories de l’émotion”, in: SCHAAL B. and CHREA C. (eds), *Odeurs et Emotions*, Presses Universitaires de Dijon, France, forthcoming.

[COPb] COPPIN G., SANDER D., *Paradigm Shift in Emotion Research*, forthcoming.

[COS 00] COSMIDES L., TOOBY J., “Evolutionary psychology and the emotions”, in: LEWIS M. and HAVILAND-JONES J.M. (eds), *Handbook of Emotions* (2nd Edition), pp. 91-115, Guilford, New York, USA, 2000.

[DAM 94] DAMASIO A.R., *Descartes’ Error*, Putnam’s Sons, New York, USA, 1994.

[DAR 72] DARWIN C., *The Expression of the Emotions in Man and Animals*, John Murray, London, 1872.

[DAV 04] DAVIDSON R.J., MAXWELL J.S., SHACKMAN A.J., “The privileged status of emotion in the brain”, *Proceedings of the National Academy of Sciences*, vol. 101, pp. 11915-11916, 2004.

[DES 96] DESCARTES R., *Les Passions de l’Âme*, Flammarion, Paris, France, 1996.

[DUF 41] DUFFY E., “An explanation of ‘emotional’ phenomena without the use of the concept ‘emotion’”, *Journal of General Psychology*, vol. 25, pp. 283-293, 1941.

[DUF 62] DUFFY E., *Arousal and Behavior*, Wiley, New York, USA, 1962.

[EKM 69] EKMAN P., SORENSON E.R., FRIESEN W.V., “Pancultural elements in facial displays of emotion”, *Science*, vol.164, pp. 86-88, 1969.

[EKM 71] EKMAN P., FRIESEN W.V., “Constants across cultures in the face and emotion”, *Journal of Personality and Social Psychology*, vol. 17, pp. 124-129, 1971.

[EKM 72] EKMAN P., “Universals and cultural differences in facial expression of emotion”, in: COLE J.R. (ed.), *Nebraska Symposium on Motivation*, pp. 207-283, University of Nebraska Press, Lincoln, USA, 1972.

[EKM 77] EKMAN P., “Biological and cultural contributions to body and facial movement”, in: BLACKING J. (ed.), *Anthropology of the Body*, pp. 34-84, Academic Press, London, 1977.

[EKM 82] EKMAN P., *Emotion in the Human Face*, Cambridge University Press, Cambridge, 1982.

[EKM 83] EKMAN P., LEVENSON R.W., FRIESEN W.V., “Autonomic nervous system activity distinguishes among emotions”, *Science*, vol. 221, pp. 1208-1210, 1983.

[EKM 84] EKMAN P., “Expression and the nature of emotion”, in: SCHERER K.R. and EKMAN P. (eds), *Approaches to Emotion*, pp. 319-344, Erlbaum, Hillsdale, NJ, USA, 1984.

[EKM 89] EKMAN P., in: RIMÉ B. and SCHERER K.R. (eds), *Les Émotions*, pp. 183-201, Delachaux et Niestlé, Neuchâtel, Switzerland, 1989.

[EKM 92] EKMAN P., "An argument for basic emotions", *Cognition and Emotion*, vol. 6, pp. 169-200, 1992.

[EKM 04a] EKMAN P., *Emotions Revealed: Understanding Faces and Feelings*, Orion Books, London, 2004.

[EKM 04b] EKMAN P., "What we become emotional about", in: MANSTEAD A.S.R., FIJDA N. and FISCHER A. (eds), *Feelings and Emotions. The Amsterdam Symposium*, pp. 119-135, Cambridge University Press, Cambridge, 2004.

[ELL 94] ELLSWORTH P.C., "William James and emotion: is a century of fame worth a century of misunderstanding?", *Psychological Review*, vol.101, pp. 222-229, 1994.

[ELL 03] ELLSWORTH P.C., SCHERER K.R., "Appraisal processes in emotion", in: DAVIDSON R.J., GOLDSMITH H. and SCHERER K.R. (eds), *Handbook of Affective Sciences*, pp. 572-595, Oxford University Press, London, 2003.

[FEH 70] FEHR B., STERN J.A., "Peripheral physiological variables and emotion: The James-Lange theory revisited", *Psychological Bulletin*, vol. 74, pp. 411-424, 1970.

[FEH 84] FEHR B., RUSSELL A.J., "Concept of emotion viewed from a prototype perspective", *Journal of Experimental Psychology: General*, vol. 113, pp. 464-486, 1984.

[FEL 98] FELDMAN BARRETT L., RUSSELL J., "Independence and bipolarity in the structure of affect", *Journal of Personality and Social Psychology*, vol. 74, pp. 967-984, 1998.

[FEL 99] FELDMAN BARRETT L., RUSSELL J., "The structure of current affect: Controversies and emerging consensus", *Current Directions in Psychological Science*, vol. 8, pp. 10-14, 1999.

[FEL 06] FELDMAN BARRETT L., "Are emotions natural kinds?", *Perspectives on Psychological Science*, vol. 1, pp. 28-58, 2006.

[FEL 09] FELDMAN BARRETT L., RUSSELL J., "Circumplex models of affect", in: SANDER D. and SCHERER K.R. (eds), *Oxford Companion to Emotion and the Affective Sciences*, Oxford University Press, London, 2009.

[FEL 05] FELLOUS J.M., ARBIB M.A., *Who needs emotions? The brain meets the robot*, Oxford University Press, Oxford, 2005.

[FON 07] FONTAINE J., SCHERER K.R., ROESCH E., ELLSWORTH P., "The world of emotion is not two-dimensional", *Psychological Science*, vol. 18, pp. 1050-1057, 2007.

[FON 09] FONTAINE J., "Dimensional emotion models", in: SANDER D. and SCHERER K.R. (eds), *Oxford Companion to Emotion and the Affective Sciences*, Oxford University Press, London, 2009.

[FRA 63] FRAISSE P., "Les émotions", in: FRAISSE P. and PIAGET J. (eds), *Traité de Psychologie Expérimentale*, vol. 5, pp. 97-181, PUF, Paris, France, 1963.

[FRI 86] FRIJDA N.H., *The Emotions*, Cambridge University Press, Cambridge, 1986.

[FRI 87] FRIJDA N.H., "Emotions, cognitive structures and action tendency", *Cognition and Emotion*, vol. 1, pp. 115-143, 1987.

[FRI 89] FRIJDA N.H., in: RIME B. and SCHERER K.R. (eds), *Les Émotions*, pp. 21-72, Delachaux et Niestlé, Neuchâtel, Switzerland, 1989.

[FRI 06] FRIJDA N.H., *The Laws of Emotion*, Erlbaum, Hillsdale, NJ, USA, 2006.

[GRA 08] GRANDJEAN D., SCHERER K.R., "Unpacking the cognitive architecture of emotion processes", *Emotion*, vol. 8, pp. 341-351, 2008.

[GRA 09] GRANDJEAN D., SCHERER K.R., "Théorie de l'évaluation cognitive et dynamique des processus émotionnels", in: SANDER D. and SCHERER K.R. (eds), *Traité de Psychologie des Émotions*, pp. 41-76, Dunod, Paris, France, 2009.

[GRA] GRATCH J., MARSELLA S., PETTA P., "Modeling the antecedents and consequences of emotion", *Journal of Cognitive Systems Research*, vol. 10, pp. 1-5, forthcoming.

[GRI 04] GRIFFITHS P.E., "Towards a Machiavellian theory of emotional appraisal", in: CRUSE P. and EVANS D. (eds), *Emotion, Evolution and Rationality*, pp. 89-105, Oxford University Press, Oxford, 2004.

[IZA 71] IZARD C.E., *The Face of Emotion*, Plenum, New York, USA, 1971.

[IZA 09] IZARD C.E., KING K.A., "Differential emotions theory", in: SANDER D. and SCHERER K.R. (eds), *Oxford Companion to Emotion and the Affective Sciences*, Oxford University Press, London, 2009.

[JAM 84] JAMES W., "What is an emotion?", *Mind*, vol. 9, pp. 188-205, 1884.

[JAM 92] JAMES W., *Psychology*, Henry Holt and Company, New York, 1892.

[JAM 94] JAMES W., "The physical basis of emotion", *Psychological Review,* 1894.

[KAP 01] KAPPAS A., "A metaphor is a metaphor is a metaphor: Exorcising the homunculus from appraisal theory", in: SCHERER K.R., SCHORR A. and JOHNSTONE T. (eds), *Appraisal Processes in Emotion: Theory, Methods, Research*, pp. 157-172, Oxford University Press, New York, USA, 2001.

[KEL 75] KELLOGG R., BARON R.S., "Attribution theory, insomnia, and the reverse placebo effect: A reversal of Storm's and Nisbett's finding", *Journal of Personality and Social Psychology*, vol. 32, pp. 231-236, 1975.

[KLE 81] KLEINGINNA P.R., KLEINGINNA A.M., "A categorized list of emotion definitions with suggestions for a consensual definition", *Motivation and Emotion*, vol. 5, pp. 345-379, 1981.

[KOR 11] KORSTEN N., ROESCH E.B., FRAGOPANAGOS N., TAYLOR J.G., GRANDJEAN D., SANDER D., "Biological and computational constraints to psychological modelling of emotion", in: PETTA P. *et al.* (eds), *Handbook for Research on Emotions and Human-Machine Interactions - HUMAINE*, Springer-Verlag, Berlin, Germany, 2011.

[KUP 07] KUPPENS P., MECHELEN I.V., "Interactional appraisal models for the anger appraisal of threatened self-esteem, other-blame, and frustration", *Cognition and Emotion*, vol. 21, pp. 56-77, 2007.

[LAN 07] LANCTÔT N., HESS U., "The timing of appraisals", *Emotion*, vol. 7, pp. 207-212, 2007.

[LAN 85] LANGE C., *The Emotions*, Williams & Wilkins, Baltimore, USA, 1885.

[LAZ 66] LAZARUS R.S., *Psychological Stress and the Coping Process*, McGraw-Hill, New York, USA, 1966.

[LAZ 68] LAZARUS R.S., "Emotions and adaptation: Conceptual and empirical relations", in: ARNOLD W.J. (eds), *Nebraska Symposium on Motivation*, pp. 175-270, University of Nebraska Press, Lincoln, USA, 1968.

[LAZ 78] LAZARUS R.S., LAUNIER R., "Stress-related transactions between person and environment", in: PERVIN L.A. and LEWIS M. (eds), *Perspectives in Interactional Psychology*, pp. 287-327, Plenum, New York, USA, 1978.

[LAZ 84] LAZARUS R.S., FOLKMAN S., *Stress, Appraisal and Coping*, Springer, New York, USA, 1984.

[LAZ 91] LAZARUS R.S., *Emotion and Adaptation*, Oxford University Press, Oxford, 1991.

[LED 93] LE DOUX J.E., "Emotional memory systems in the brain", *Behavioral Brain Research*, vol. 58, pp. 69-79, 1993.

[LED 96] LE DOUX J.E., *The Emotional Brain*, Simon & Schuster, New York, USA, 1996.

[LEV 87] LEVENTHAL H., SCHERER K.R., "The relationship of emotion and cognition: A functional approach to a semantic controversy", *Cognition and Emotion*, vol. 1, pp. 3-28, 1987.

[MAN 84] MANDLER G., *Mind and Body*, Norton, New York, USA, 1984.

[MAR 09] MARINIER R.B., LAIRD J.E., LEWIS R.L., "A computational unification of cognitive behavior and emotion", *Journal of Cognitive Systems Research*, vol. 10, pp. 48-69, 2009.

[MAT 09] MATSUMOTO D., EKMAN P., "Basic emotions", in: SANDER D. and SCHERER K.R. (eds), *Oxford Companion to Emotion and the Affective Sciences*, Oxford University Press, Oxford, 2009.

[MEH 96] MEHRABIAN A., "Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament", *Current Psychology*, vol. 14, pp. 261-292, 1996.

[MOO 07] MOORS A., "Can cognitive methods be used to study the unique aspect of emotion: An appraisal theorist's answer", *Cognition and Emotion*, vol. 21, pp. 1238-1269, 2007.

[MOO 09] MOORS A., "Theories of emotion causation: A review", *Cognition and Emotion*, vol. 23, pp. 625-662, 2009.

[MOO 10] MOORS A., "Automatic constructive appraisals as a candidate cause of emotion", *Emotion Review*, vol. 2, pp. 139-156, 2010.

[NEI 67] NEISSER U., *Cognitive Psychology.* Wiley, New York, 1967.

[NIE 07] NIEDENTHAL P.M., "Embodying emotion", *Science*, vol. 316, pp. 1002-1005, 2007.

[OAT 87] OATLEY K., JOHNSON-LAIRD P.N., "Towards a cognitive theory of emotions", *Cognition & Emotion*, vol. 1, pp. 29-50, 1987.

[ÖHM 78] ÖHMAN A., DIMBERG U., "Facial expressions as conditioned stimuli for electrodermal responses: A case of "preparedness"", *Journal of Personality and Social Psychology*, vol. 36, pp. 1251-1258, 1978.

[ÖHM 99] ÖHMAN A., "Distinguish unconscious from conscious emotional processes: Methodological considerations and theoretical implications", in: DALGLEISH T. and POWER M.J. (eds), *Handbook of Cognition and Emotion*, pp. 321-352, Wiley & Sons, Hoboken, USA, 1999.

[ÖHM 01] ÖHMAN A., MINEKA S., "Fears, phobias, and preparedness: toward an evolved module of fear and fear learning", *Psychological Review*, vol. 108, pp. 483-522, 2001.

[OLS 07a] OLSSON A., NEARING K.I., PHELPS E.A., "Learning fears by observing others: the neural systems of social fear transmission", *Social Cognitive and Affective Neuroscience*, vol. 2, pp. 3-11, 2007.

[OLS 07b] OLSSON A., PHELPS E.A., "Social learning of fear", *Nature Neuroscience*, vol. 10, pp. 1095-1102, 2007.

[ORT 90] ORTONY A., TURNER T.J., "What's basic about basic emotions?", *Psychological Review*, vol. 97, pp. 315-331, 1990.

[OSG 62] OSGOOD C.E., "Studies of the generality of affective meaning systems", *American Psychologist*, vol. 17, pp. 10-28, 1962.

[PAN 98] PANKSEPP J., "Feeling the pain of social loss", *Science*, vol. 302, pp. 237-239, 1998.

[PEL 09] PELACHAUD C., "Embodied conversational agent ECA", in: SANDER D. and SCHERER K.R. (eds), *Oxford Companion to Emotion and the Affective Sciences*, Oxford University Press, Oxford, 2009.

[PET 09] PETTA P., GRATCH J., "Computational models of emotion", in: SANDER D. and SCHERER K.R. (eds), *Oxford Companion to Emotion and the Affective Sciences*, Oxford University Press, Oxford, 2009.

[PIC 09] PICARD R.W., "Affective computing", in: SANDER D. and SCHERER K.R. (eds), *Oxford Companion to Emotion and the Affective Sciences*, Oxford University Press, Oxford 2009.

[PLU 80] PLUTCHIK R., *Emotion. A Psychoevolutionary Synthesis*, Harper & Row, New York, USA, 1980.

[REI 83] REISENZEIN R., "The Schachter theory of emotion: two decades later", *Psychological Bulletin*, vol. 94, pp. 239-264, 1983.

[RIM 86] RIMÉ B., GIOVANNINI D., "The physiological patterns of reported emotional states", in: SCHERER K.R., WALLBOTT H.G. and SUMMERFIELD A.B. (eds), *Experiencing Emotion: A Cross-cultural Study*, pp. 84-97, Cambridge University Press, Cambridge, 1986.

[ROE 07] ROESCH E.R., SANDER D., SCHERER K.R., "The link between attention and emotion: A playground for psychology, neuroscience, and plausible artificial neural networks", in: MARQUES DE SA J. *et al.* (eds), *Proceedings of the International Conference on Artificial Neural Network*, pp. 859-868, 2007.

[ROS 84] ROSEMAN I., "Cognitive determinants of emotion: a structural theory", in: SHAVER P. (eds), *Review of Personality and Social Psychology*, *vol. 5, Emotions, relationships, and health*, pp. 11-36, Sage, Beverly Hills, USA, 1984.

[RUS 80] RUSSELL J.A., "A circumplex model of affect", *Journal of Personality and Social Psychology*, vol. 39(6), pp. 1161-1178, 1980.

[RUS 99] RUSSELL J.A., CARROLL J.M., "The phoenix of bipolarity", *Psychological Bulletin*, vol. 125, pp. 611-617, 1999.

[RUS 03] RUSSELL J.A., "Core affect and the psychological construction of emotion", *Psychological Review*, vol. 110, pp. 145-172, 2003.

[SAN 02] SANDER D., KOENIG O., "No inferiority complex in the study of emotion complexity: A cognitive neuroscience computational architecture of emotion", *Cognitive Science Quarterly*, vol. 2, pp. 249-272, 2002.

[SAN 03] SANDER D., GRAFMAN J., ZALLA T., "The human amygdale: an evolved system for relevance detection", *Reviews in the Neurosciences*, vol. 14, pp. 303-316, 2003.

[SAN 05] SANDER D., GRANDJEAN D., SCHERER K.R., "A system approach to appraisal mechanisms in emotion", *Neural Networks*, vol. 18, pp. 317-352, 2005.

[SAN 07] SANDER D., GRANDJEAN D., KAISER S., WEHRLE T., SCHERER K.R., "Interaction effects of perceived gaze direction and dynamic facial expression: Evidence for appraisal theories of emotion", *European Journal of Cognitive Psychology*, vol. 19, pp. 470-480, 2007.

[SAN 09a] SANDER D., SCHERER K.R., "La psychologie des émotions: survol des théories et débats essentiels", in: SANDER D. and SCHERER K.R., *Traité de Psychologie des Émotions*, pp. 1-39, Dunod, Paris, France, 2009.

[SAN 09b] SANDER D., SCHERER K.R. (eds), *Oxford Companion to Emotion and the Affective Sciences*, Oxford University Press, Oxford, 2009.

[SAN 09c] SANDER D., "Computational analysis of emotion", in: SANDER D. and SCHERER K.R. (eds), *Oxford Companion to Emotion and the Affective Sciences*, Oxford University Press, Oxford, 2009.

[SCH 62] SCHACHTER S., SINGER J.E., "Cognitive, social and physiological determinants of emotional state", *Psychological Review*, vol. 69, pp. 379-399, 1962.

[SCH 84a] SCHERER K.R., "Les émotions: fonctions et composantes", *Cahiers de Psychologie Cognitive*, vol. 4, pp. 9-39, 1984.

[SCH 84b] SCHERER K.R., "On the nature and function of emotion: a component process approach", in: SCHERER K.R. and EKMAN P. (eds), *Approaches to Emotion,* pp. 293-317, Hillsdale, Erlbaum, NJ, USA, 1984.

[SCH 87] SCHERER K.R., "Toward a dynamic theory of emotion: The component process model of affective states", *Geneva Studied in Emotion and Communication*, vol. 1, pp. 1-98, 1987.

[SCH 89] SCHERER K.R., "Les émotions: fonctions et composantes", in: RIME B. and SCHERER K.R. (eds), *Les Émotions*, pp. 97-134, Delachaux et Niestlé, Neuchatel, Switzerland, 1989.

[SCH 92] SCHERER K.R., "What does facial expression express?", *International Review of Studies on Emotion*, vol. 2, pp. 139-165, 1992.

[SCH 93a] SCHERER K.R., "Neuroscience projections to current debates in emotion psychology", *Cognition and Emotion*, vol. 7, pp. 1-41, 1993.

[SCH 93b] SCHERER K.R., "Studying the emotion-antecedent appraisal process: An expert system approach", *Cognition and Emotion*, vol. 7, pp. 1-41, 1993.

[SCH 00] SCHERER K.R., "Emotions as episodes of subsystem synchronisation driven by nonlinear appraisal processes", in: LEWIS M.D. and GARNIC I. (eds), *Emotion, Development, and Self-organization: Dynamic Systems Approaches to Emotional Development*, pp. 70-99, Cambridge University Press, Cambridge and New York, 2000.

[SCH 01a] SCHERER K.R., "Appraisal considered as a process of multi-level sequential checking", in: SCHERER K.R., SCHORR A., and JOHNSTONE T. (eds), *Appraisal Processes in Emotion: Theory, Methods, Research*, pp. 92-120, Oxford University Press, New York, 2001.

[SCH 04] SCHERER K.R., "Feelings integrate the central representation of appraisal-driven response organization in emotion", *Feelings and Emotions: The Amsterdam Symposium*, pp. 136-157, Cambridge University Press, Cambridge, 2004.

[SCH 02] SCHIENLE A., STARK R., WALTER B., BLECKER C., OTTO U., KIRSCH P., SAMMER G., VAITL D., "The insula is not specifically involved in disgust processing: an fMRI study", *Neuroreport*, vol. 13, pp. 2023-2026, 2002.

[SCH 01b] SCHORR A., "Subjective measurement in appraisal research", in SCHERER K.R., SCHORR A., and JOHNSTONE T. (eds), *Appraisal Processes in Emotion: Theory, Methods, Research*, pp. 331-349, Oxford University Press, New York, 2001.

[SIE 07] SIEMER M., MAUSS I., GROSS J.J., "Same situation – different emotions: How appraisals shape our emotions", *Emotion*, vol. 7, pp. 592-600, 2007.

[SMA 03] SMALL D.M., GREGORY M.D., MAK Y.E., GITELMAN D., MESULAM M.M., PARRISH T., "Dissociation of neural representation of intensity and affective valuation in human gustation", *Neuron*, vol. 39, pp. 701-711, 2003.

[SMI 85] SMITH C.A., ELLSWORTH P.C., "Patterns of cognitive appraisal in emotion", *Journal of Personality and Social Psychology*, vol. 43, pp. 813-838, 1985.

[SMI 97] SMITH C.A., SCOTT H.S., "A componential approach to the meaning of facial expressions", in: RUSSELL J.A. and FERNANDEZ-DOLS J.M. (eds), *The Psychology of Facial Expression*, pp. 229-254, Cambridge University Press, New York, USA, 1997.

[STE 01] STEMMLER G., HELDMANN M., PAULS C.A., SCHERER T., "Constraints for emotion specificity in fear and anger: The context counts", *Psychophysiology*, vol. 38, pp. 275-291, 2001.

[STO 77] STORMS M.D., NISBETT R.E., "Insomnia and the attribution process", *Journal of Personality and Social Psychology*, vol. 16, pp. 319-328, 1977.

[STR 96] STRONGMAN K.T., *The Psychology of Emotion: Theories of Emotion in Perspective*, 4th edition, John Wiley and Sons, New York, USA, 1996.

[SUS 08] SUSSKIND J.M., LEE D.H., CUSI A., FELMAN R., GRABSKI W., ANDERSON A.K., "Expressing fear enhances sensory acquisition", *Nature Neuroscience*, no.11, pp. 843-850, 2008.

[TAY 09] TAYLOR J.G., KORSTEN N., "Connectionist models of emotion", in: SANDER D. and SCHERER K.R. (eds), *Oxford Companion to Emotion and the Affective Sciences*, Oxford University Press, Oxford, 2009.

[TER 03] TERRACCIANO A., MCCRAE R., HAGEMANN D., COSTA P., "Individual difference variables, affective differentiation, and the structures of affect", *Journal of Personality*, vol. 71, pp. 669-704, 2003.

[THA 02] THAGARD P., "The passionate scientist: Emotion in scientific cognition", in: CARRUTHERS P., STICH S. and SIEGAL M. (eds), *The Cognitive Basis of Science*, pp. 235-250, Cambridge University Press, Cambridge, 2002.

[TIT 09] TITCHENER E.B., *A Text-book of Psychology*, Macmillan, New York, USA, 1909.

[TOM 80] TOMKINS S.S., "Affect as amplification: Some modifications in theory", in: PLUTCHIK R. and KELLERMAN H. (eds), *Emotion, Theory, Research, and Experience: Theories of Emotions*, pp. 141-165, Academic Press, New York, USA, 1980.

[WAT 85] WATSON D., TELLEGEN A., "Toward a consensual structure of mood", *Psychological Bulletin*, vol. 98, pp. 219-235, 1985.

[WAT 99] WATSON D., TELLEGEN A., "Issues in the dimensional structure of affect – effects of descriptors, measurement error, and response formats: Comment on Russell and Carroll", *Psychological Bulletin*, vol. 125, pp. 601-610, 1999.

[WEH 00] WEHRLE T., KAISER S., SCHMIDT S., SCHERER K.R., "Studying the dynamics of emotional expression using synthesized facial muscle movements", *Journal of Personality and Social Psychology*, vol. 78, pp. 105-119, 2000.

[WEI 85] WEINER B., "An attributional theory of achievement motivation and emotion", *Psychological Review*, vol. 92, pp. 548-573, 1985.

[WIE 92] WIERZBICKA A., "Talking about emotions: Semantics, culture, and cognition", *Cognition and Emotion*, vol. 6, pp. 285-319, 1992.

[WIN 05] WINSTON J.S., GOTTFRIED J.A., KILNER J.M., DOLAN R.J., "Integrated neural representations of odor intensity and affective valence in human amygdale", *Journal of Neuroscience*, vol. 25, pp. 8903-8907, 2005.

[WUN 97] WUNDT W.M., *Outlines of Psychology*, Engelman, Oxford, 1897.

[ZAJ 84] ZAJONC R.B., "On the primacy of affect", *American Psychologist*, vol. 39, pp. 117-123, 1984.

Chapter 2

Emotion and the Brain

2.1. Introduction

Since the 1990s, research into the cerebral mechanisms involved in emotional processes has seen a substantial boom, coinciding with the establishment of affective neuroscience as a field in its own right. This latest area of research relies on concepts and methods from a variety of disciplines, including philosophy, anthropology, psychology, psychiatry, biology, neurology and computational sciences as well as neurosciences in the broader sense of the term. While technological advances have strongly improved our knowledge of cerebral functions, strong traditions nevertheless still influence current conceptualizations of emotion and the brain. This new technology, combined with centuries of reflection, has therefore resulted in a more integrated perspective on emotional processes.

The main aim of this chapter is to examine the neuronal bases underlying various emotional processes and subsequently highlight that affective neuroscience is a multidisciplinary approach. As a form of introduction, we will outline a brief history of the main steps that have punctuated the understanding about the relationships between emotions and the brain. We will then spend some time describing the neuroscientific approach to emotions and how it has contributed to our knowledge about emotional phenomena. Afterwards, we will focus on some initial theoretical assumptions about the role of emotions in the brain. We will then examine the main cerebral structures and their functions in different emotional mechanisms. We will conclude by presenting some more advanced concepts about cerebral connectivity and the temporal dynamic of emotional processes.

Chapter written by Andy CHRISTEN and Didier GRANDJEAN.

Emotion-Oriented Systems Edited by Catherine Pelachaud

2.1.1. *Emotions and the brain: the emergence of affective neuroscience as an independent discipline*

This section will provide a brief overview of the major events shaping our understanding of the connections between emotion and the brain that have, in turn, led to the emergence of affective neuroscience as an independent area of research. This section seeks to explain the origins of this new approach by examining the various disciplines and methods that have influenced its conception.

In his 17th Century work, *Les Passions de l'Âme*, René Descartes [DES 49] suggested there was a dual relationship between body and spirit, or between passion and reason as he described it. Basing his reflection on knowledge about various areas of the brain, Descartes suggested that the pineal gland, in his view an indivisible structure of the brain, was the location of the soul. This philosophical conception represented the first steps towards localizing specific functions within the brain and therefore to link notions from mind to cerebral structures. While this example is of mere historical significance today, it nevertheless illustrates how philosophy can and has stimulated neuroscientific research, which in turn has produced philosophical debate.

The first clinical descriptions attesting to a concrete link between the brain and emotions appeared during the 19th Century, where the infamous case of Phineas Gage − examined by John Harlow in 1868 [HAR 68] − demonstrated that brain damage can cause major changes in socio-emotional behavior. However, systematic study and further understanding of the cerebral mechanisms involved in emotion processes did not emerge until much later. A biological and behavioral perspective has therefore dominated research in the field of emotion for an extended period [DAV 03c]. This was all changed by Charles Darwin's *The Expression of the Emotions in Man and Animals*, published in 1872, which would have a lasting impact on emotion theories and the development of affective neuroscience.

The beginning of the 20th Century saw fierce debate between peripheralist [JAM 90] and centralist [CAN 29] theories on the nature of how emotion are generated, leading to the first predictions that emotion is accompanied by changes in the activity of some brain structures. Several years later, in 1937, James Papez [PAP 37] first advanced the notion of a brain circuit being involved in emotion. On the basis of this work and that of Canon and Bard, MacLean [MAC 49] subsequently formulated his 1949 triune brain theory. Nevertheless, a tremendous turn in exploring brain mechanisms involved in emotions was initiated in the 1970s, this trend mainly growing from animal literature and around the notion of conditioning [LED 87]. Since the end of the 1980s, technological advances made in the tool that serve to investigate brain functions, such as magnetic resonance imaging, finally offers the possibility of examining the brain activity *in vivo*, but in an indirect manner. This revolution has also given rise to a colossal amount of publications into the neurobiological bases of emotional processes. This coincides with the establishment of affective neuroscience as an independent field within neuroscience [DAV 95], see Figure 2.1.

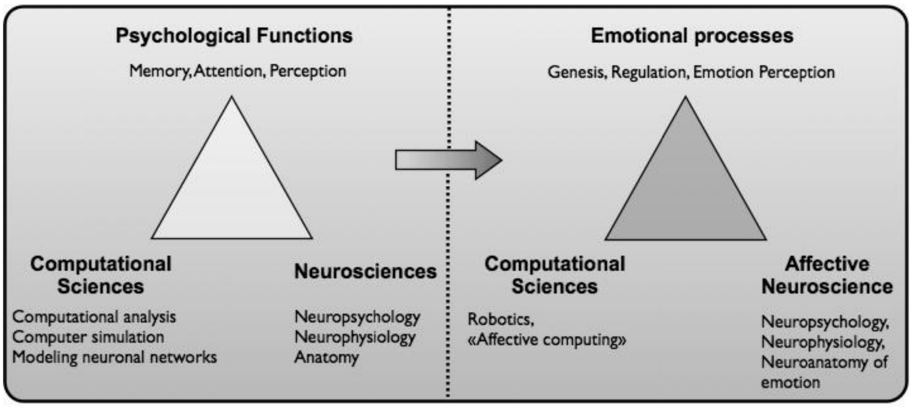


Figure 2.1. *The cognitive neuroscience triangle applied to the study of emotion, adapted from [KOS 92]*

The term "neuroscience" encompasses a wide range of disciplines devoted to the study of anatomy and cerebral functioning. Historically, the cognitive revolution which began in the 1950s, along with progress made in the field of informatics, led to the emergence of cognitive neuroscience at the end of the 1970s (Gazzaniga and Miller). This discipline aims to understand how the brain underlies cognitive functions such as language, memory, perception and attention [KOS 92]. In a similar way, affective neuroscience focuses on the neural foundations underlying affective processes and more specifically on mood and emotions [DAV 03]. These two fields of study are based on an interdisciplinary approach, since they bring together concepts and methods from psychology, computational science and neurosciences. This approach can be modeled by a triangle bringing together these three disciplines [OCH 99]. The fact that the triangle is equilateral conveys that none of the disciplines are favored for understanding psychological functions. Figure 2.1 illustrates the application of concepts from cognitive neuroscience in the study of emotion.

2.2. The major role of affective neuroscience in understanding emotions

Since the 1990s, affective neuroscience has benefited from an extraordinary convergence of data from human and animal literature, which has consolidated its privileged status in the study of emotional processes [DAV 03b]. Similar conclusions, obtained by using various methods such as the lesion approach or neuroimaging techniques, reinforce the significance of results obtained so far (see Figure 2.2.). Knowledge of how the brain conveys emotion has allowed both freedom from *a priori* conceptions on their functioning as well as to constraint psychological modeling of emotional processes leading to more detailed theories. In the following section, we will examine some fundamental contributions made by neuroscience to the development of conceptualizations about emotion. This section will therefore provide a background that will hopefully enable the reader to approach the different elements found in the rest of the chapter critically.

2.2.1. *Emotion and the brain: from a unitary entity to processing, from structure to neural networks*

Following the example of cognitive science, which has split information processes into more core components, neuroscientific research also favors such subdivisions for the construct of emotion. Indeed, current research has shown that emotion is not processed by the brain as a unitary entity but includes a heterogeneous set of processes involving widely dispersed neural networks. These processes encompass the genesis, the differentiation, the expression, the perception, the recognition, the regulation, the subjective feeling of emotion – as well as motivational processes. Each of these mechanisms can be further split into more developed subdivisions, which are beyond the scope of the present chapter. The data from the literature tend to show that this approach better accounts for emotional complexity, for different levels of emotional processing and for the multiple mental operations accompanying them.

In the literature, the study of emotional phenomena can be understood by different levels of increasing complexity, expanding from the structuralist approach to modeling of neuronal networks. In the perspective aiming to determine which brain areas are involved in a given function, the research interest focuses on functional specificity of brain-related activity during an emotional episode. Theories proposed by authors defending core or discrete emotion approaches illustrate this localizationalist perspective. Indeed, this approach argues that each discrete emotion is represented by specific neural patterns [EKM 99]. For example, a significant amount of data suggests that the amygdala is involved in processing fear-related signals, while the insula would play a specific role in processing information linked to disgust [CAL 01]. We will return to this issue in the section 2.5 of this chapter.

Neuroimaging techniques can serve different objectives, depending on whether the aim is to understand the brain activity involved in emotional processes or whether we want to identify differences at the structural level of the brain. The "structural imagery" seeks to identify, locate and measure the different parts of the central nervous system. This method has provided a wealth of medical information, where it can be used diagnostically as well as helping us to understand how the brain functions. Indeed, it brings elements that have helped interpret behavioral data established in neuropsychology. By using this method, we can then easily locate the site of a brain lesion and identify the role played by a specific brain structure in a given process. This type of technique is also used in the domain of pathology to identify biological markers, such as difference in volume and thickness of the cortical layer for example. "Functional imagery" on the other hand aims to characterize brain activity related to a specific task, which is equivalent to highlight the functions played by some brain structures in the processes induced by this task. Various techniques are used to measure brain signals with more or less satisfactory spatial and temporal precision. One of these techniques, functional magnetic resonance imaging (fMRI, see Figure 2.2a) consists of measuring the BOLD signal, reflecting Blood Oxygen Levels in the Brain. Using a mechanism called a hemodynamic response, the inflow of oxygenated blood increases in the areas requiring energy. This method allows us to characterize with a high degree of precision (spatial resolution of about one third of a millimeter at best) the brain areas involved in a given task (see Figure 2.2c). In contrast, this method has a poor temporal resolution (in the range of the second), which is greatly inferior to the dynamic of emotional processes. Electroencephalography (EEG, see Figure 2.2b) is another non-invasive method of measuring electric brain activity. While this method has a relatively imprecise spatial resolution, it does have a temporal resolution of a millisecond. As such, this tool seems to be advantageous for studying the temporal dynamic of emotional processes. Among different possible analysis, one can compute event-related potentials (ERPs, see Figure 2.2d), which are positive or negative waves reflecting the electric signature of a characteristic stage of the information processing.

Studying a brain structure's function in emotional processing is only a first step in understanding the complex brain architecture underpinning a process. Beyond this regionalist view, models try to include, within a complex neural network, the specific processes carried out by certain brain structures during the temporal unfolding of an emotion. This new, network-based approach has shown, among other things, that the processes linked to emotion do not depend solely on sub-cortical regions of the brain as, for example, Panksepp has suggested [PAN 98]. On the contrary, more recent theoretical suggestions have indicated that emotional processes involve vast neural networks that are widely distributed between cortical and sub-cortical brain areas. In this respect, Adolphs has provided an eloquent demonstration with the case of facial emotion recognition, by detailing not only the different processes and specific areas of the brain underpinning them, but also by proposing their temporal dynamics and their anatomical connections [ADO 02]. The components underpinning an emotion, such as identifying a facial expression or the awareness of bodily changes during an emotional episode would all be implemented in at least partly distinct neuronal circuits, those being therefore largely interconnected and potentially modulated by different sensory modality. Apart from MRI, EEG techniques, which emphasize the study of temporal dynamics, offer the possibility of identifying more precisely the different stages and the various components of emotional processes. It may therefore be possible to learn how brain structures exchange information temporally during an emotional episode (see, for example [GRA 08a]).

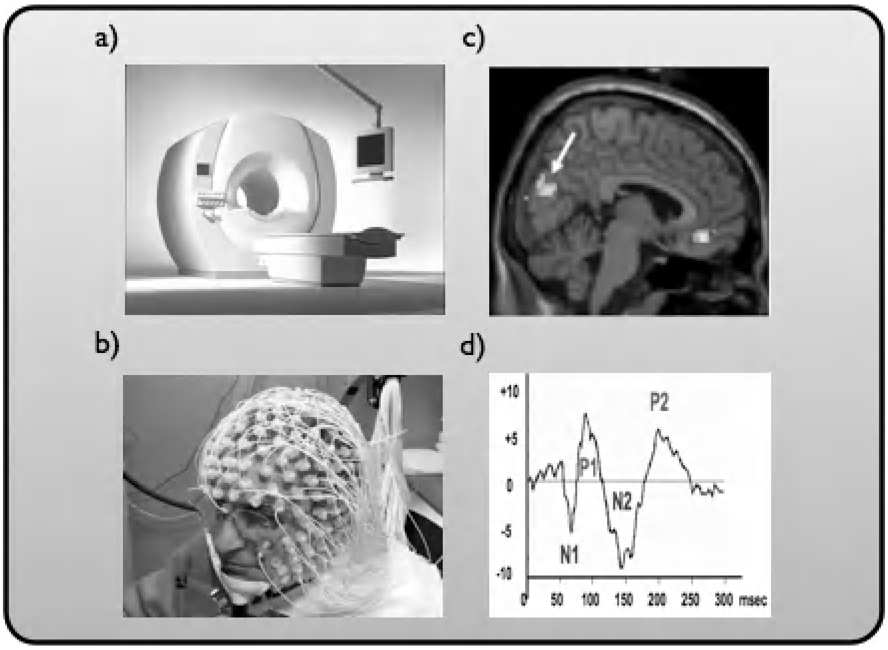


Figure 2.2. *Research techniques used to examine brain structure and function*

2.2.2. *Levels of processing in emotional processes*

For many years, emotions have been viewed within psychology as representing conscious states requiring *de facto* voluntary processing. As proof, investigations about the subjective feeling were bound to self-report methods, such as self-administered questionnaires or even measures of verbal communication. At the time, authors have concluded that without changes in this type of indicator, the genesis of emotion has not been triggered. While these methods have some significance for understanding emotional phenomena, the use of a varied set of experimental paradigms has highlighted that a number of so-called emotional processes escape our explicit consciousness. The development of emotional information processing inaccessible to the verbal report and potentially associated with largely automatic affective processes could be explained by the fact that our cognitive system (i.e. the part dealing with processes which can be verbally expressed) is of limited capacity and that such processes need to be automatic so that the organism can react with efficiency in case of emergencies.

The evolutionary approach and cognitive evaluation theories have respectively proposed that emotional stimuli have a particular importance for human survival (e.g. indicating danger) and well-being (e.g. congruency to current goals and expectations). As such, the automatic detection of a stimulus could enable the system to stop ongoing activities and to redirect voluntary attention toward the area of the environment requiring a prioritized processing [VUI 05]. A large amount of evidence from behavioral studies, from EEG and functional brain imaging, indicates that there are a large number of implicit processing involved in emotional processes. This has therefore generated a large amount of literature on the subject, particularly with regard to emotional facial expression. Based on the notion of automatic emotion encoding, some authors have suggested that emotional stimuli could be processed at a pre-attentive stage [ÖHM 01], unconsciously and relatively independently of voluntary attentional processes [GRA 05, SAN 05a].

The emotional characteristics of stimuli also have the potential to lower the threshold for perceptive processing, making their detection in the presence of distracters easier [PHE 06]. Studies of brain-damaged patients suffering from hemineglect syndrome and presenting left auditory extinction following damage to the right parietal areas, have highlighted the capacity of emotional prosody to reduce auditive extinction[1] for stimuli presented to the left ear in dichotic listening [GRA 08c]. This has therefore indicated that emotional stimuli can modulate voluntary (endogenous) and involuntary (exogenous) processes involved in spatial attention. As such, examining the cerebral architecture underlying implicit and explicit processes linked to emotion has allowed us to constrain psychological and neuropsychological models leading to the development of more detailed theories.

2.2.3. *Emotion and cognition*

The traditional view considers that affect and cognition represent two separated systems which are underpinned by independent neural networks. However, recent work in the field of neurosciences has allowed us to move away from these

1 Auditive extinction refers to the fact that the patient does not report the existence of a stimulus delivered to the left ear concomitantly to the presentation of a stimulis occurring to the right ear.

simplistic, if not erroneous, conceptualizations. A large amount of evidence currently supports the notion of an overlap between the neural circuits responsible for cognition and those dedicated to affect [DAV 03b, PES 08]. Indeed, emotion seems to exert an influence on numerous aspects of cognition, such as perception, attention, memory, moral judgment and decision-making. Results show, for example, that the activity in sensory cortices is more increased in response to emotional compared to neutral events [GRA 05]. These brain modulations of sensory cortices, which mediator would be the amygdala, would strengthen the perceptual representations for this particular kind of stimuli. Other data also reveal that emotional events can capture attention quicker than neural stimuli [BRO 08]. Similarly, researchers have been able to show that emotional memories, both positive and negative, are better remembered than neutral events [LAB 06].

In this respect, the component process model (CPM) of emotion [SCH 01] offers an interesting conceptual framework because it bypasses the debate around the primacy of emotion or cognition by integrating a cognitive component as a constitutive element triggering the emotional episode (see Chapter 1 for further details). This type of approach seems more in line with empirical data which indicate bidirectional relationships between these two systems. After having exposed this preliminary and introductory knowledge from modern neuroscience, we will now examine the main theoretical proposals that have marked the development of this field of study.

2.3. The historical and conceptual legacy of early conceptions of emotions and the brain

2.3.1. *Forerunners of affective neuroscience*

2.3.1.1. *Charles Darwin*

Thirteen years after the appearance of *On the Origin of Species* and following his studies into evolution, Charles Darwin [DAR 72] published *The Expression of the Emotions in Man and Animals*. In this book, he defends the idea of a phylogenetic continuity in emotional expression. This proposal is based on numerous observations conducted from sketches and photographs revealing that animals and humans express emotions in the same way. Today, Darwin's proposal has led to the use of animal models, which by extrapolation allow a better understanding of certain aspects of emotional phenomena in humans. Darwin also believed in the existence of a limited number of core emotions that are innate and universal. This echoes current research seeking to establish to what extent different emotions are represented by partly distinct neural substrates [EKM 73, LED 96, PAN 98].

2.3.1.2. *The James-Lange peripheralist theory*

Around a decade after Darwin, in 1884, William James [JAM 84] published *What are Emotions?*, a book in which he developed a thesis explaining the mechanism at the root of emotional genesis. According to James, emotion results from awareness of bodily changes which accompany it, as they occur. This innovative proposition at that time highlighted the importance of bodily signals in emotional processes. The majority of neuroscientific researchers have revisited this peripheralist approach, emphasizing particularly that bodily feedbacks can modulate the emotional experience. We will return to this point in section 2.6.1 of this chapter, particularly when we will present the somatic marker theory.

2.4. Initial neuro-anatomical emotion theories

2.4.1. *Canon-Bard's centralist theory*

The centralist emotion theory emerged in response to criticism aimed at James-Lange's peripheralist conception of emotion (see Chapter 1). One of the main arguments of the centralist approach came from observing the effects of brain damage on the emotional behavior of cats. Canon and Bard suggested that if emotion results from an awareness of the bodily changes, as proposed by peripheralist theories, then it should depend entirely on the integrity of the sensory cortices. Bard's results [BAR 28] demonstrated that a resection of sensory cortices in cats does not suppress the emotional response. On the basis of these data, Canon and Barn developed a theory that saw the central nervous system as the source of emotion. They formulated therefore the beginnings of an explanation about the cerebral mechanisms involved in emotion. According to Canon-Bard, the hypothalamus would be involved in the response to emotional stimuli, while neocortical regions would exert an inhibitory action on this structure. This emotional response would indeed be regulated by means of top-down control of cortical areas on the hypothalamic regions, the latter no longer existing in brain-damaged cats.

The research carried out by Canon and Bard once more demonstrated the benefits of using animal models to infer about emotional mechanisms in humans. Furthermore, they use the so-called lesional method to assess the role played by a specific brain region in emotional processes. In order to do this, they worked on the postulate that changes in emotional behavior resulting from the resection of a given brain area provided information about the processes underlain by this particular brain area.

2.4.2. *Papez's circuit*

In 1937, James Papez first proposed the notion of a neural circuit being involved in emotion. According to this model, sensory information reaching the thalamus is divided into two processing streams. Along the "stream of thinking", sensations are transmitted by the thalamus in the direction of sensory cortices and the cingulate cortex to be transformed into thoughts, memories and perceptive elements. This stream also extends from the cingulate cortex to the hippocampus and reaches the mamillary bodies located in the hypothalamus via the fornix. From here, projections would be sent to the anterior thalamus. These different structures form the so-called "Papez circuit" (see Figure 2.3). In the "stream of feeling", the information from the thalamus reaches the hypothalamus directly. This channel therefore allows the generation of emotion and accompanying bodily changes via connections between the hypothalamus and bodily systems. Furthermore, this emotional response is subject to top-down cortical regulation, which would imply a transmission of information from the cingulate cortex to the hypothalamus via the hippocampus. According to Papez, emotional response is dependent on the activity of the cingulate cortex, which is responsible for integrating information coming from the sensory cortices and from the hypothalamus. Consequently, the generation of emotions can potentially take place via the two streams, given that each of them has access to all of the structures constitutive of the Papez circuit.

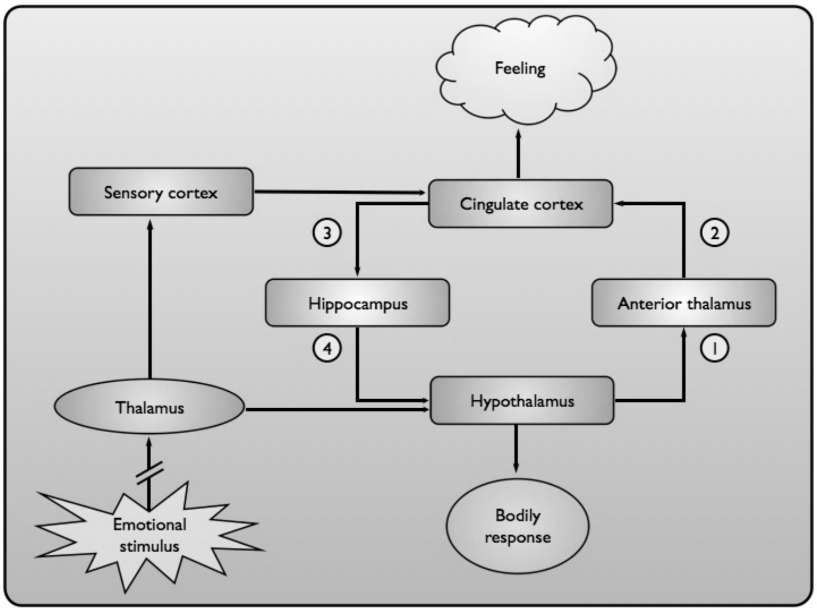


Figure 2.3. *The Papez circuit developed in 1937, borrowed from [DAL 04]*

The contribution of this model to the understanding of the neural network involved in emotion is remarkable for its time. While several anatomical paths described by Papez are supported by current data in the literature, the absence of central structures in emotional processes, such as the amygdala or specific prefrontal regions, is a major shortcoming of this model. The ideas developed by MacLean have, however, contributed to solving this problem.

2.4.3. *MacLean's limbic theory*

On the basis of the work carried out by Canon-Bard, MacLean [MAC 49] proposed in the 1950s that the human brain should be considered the product of progressive evolution, anatomically and concentrically localized. As such, MacLean formulated his triune brain theory, which subdivided the brain architecture into three distinct parts (see Figure 2.4b). The oldest part is composed of the stratial complex and the basal ganglia, which MacLean believed to be the root cause of primitive emotions such as fear or aggression. According to MacLean, the paleo-mammalian, or visceral brain constitutes the second part of the brain's architecture. MacLean includes in it the majority of areas described in Papez's circuit, such as the thalamus, hypothalamus, hippocampus and cingulated cortex.

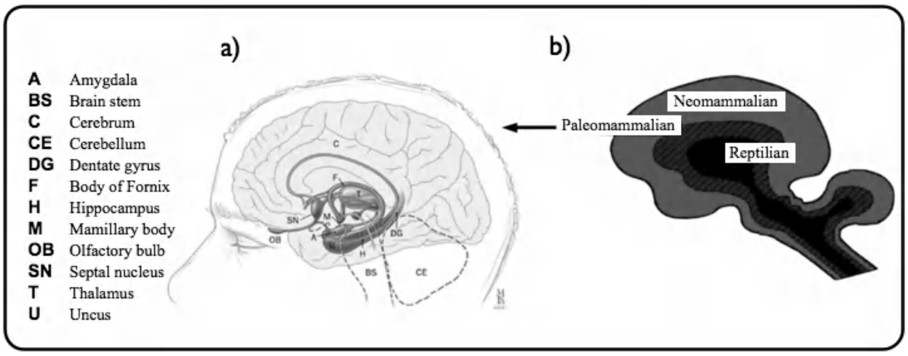


Figure 2.4. *a) View of the limbic system (b) and MacLean's triune brain*

The real innovation of MacLean's theory was to incorporate the amygdala and prefrontal cortex, which are now accepted as being determinants in emotional processes. This second subdivision would result in the notion of the limbic system and lead to MacLean's proposal that emotional processing is undertaken by a set of sub-cortical structures (see Figure 2.4a). While this notion has reflected for a long time a dominant conception about brain areas being responsible for emotional processing, more recent evidence has undermined this concept, both theoretically and empirically [CAL 01, LED 96]. Indeed, current evidence indicates that the neomammalian brain that forms the third part of MacLean's triune brain architecture represents nowadays the cerebral substrate underlying higher cognitive functions also involved in emotional processes. These data suggest that neo-cortical areas of the brain are involved in the top-down control of emotional responses.

By adopting a neo-Jamesian approach, MacLean confirmed that emotional experience comes from the integration of external sensations and bodily or internal information. According to this perspective, an emotional event triggers bodily changes, an awareness of these changes as they occur, results in an emotion. In this respect, information regarding peripheral changes reaches the brain and is integrated into a perception of the outside world. MacLean proposed that this integration mechanism depends on the activity of the visceral or paleomammalian brain, and, in particular, the hippocampus. However, more recent data have revealed that the structure MacLean considered to be essential in feeling emotion are in fact more involved in computing specific cognitive aspects, such as memory regarding the hippocampus for example.

In conclusion, the insight of these first propositions concerning emotional processes at the brain level has laid the foundations for most current research. In the following section, we will describe the main areas of the brain considered to be essential to understanding emotional processes. Among these structures, we will focus on the amygdala, the prefrontal cortex, the anterior cingulate cortex and the insula within the context of the psychological and neuropsychological theories that attempt to explain the functions carried out by these brain regions in emotional processes.

2.5. Structures in the brain and their functions in emotional processes

2.5.1. *Amygdala*

Anatomically, the amygdala is located in the anterior and medial commissure of each of the two temporal lobes (see Figure 2.5a). This structure is actually composed of a heterogeneous set of approximately 12 nuclei [AMA 92] that can each be further subdivided [LED 00]. While the spatial resolution of different imaging methods and brain damage studies do not allow us to assess this issue conclusively in humans yet; these different nuclei could each play a specific functional role [PIT 97]. Moreover, the gathering of such diverse entities into a unique structure has also been the subject of debate [SWA 98]. Nevertheless, the amygdala could be considered an integrated structure in the sense that its different nuclei maintain important reciprocal connections that seem to have evolved together during the phylogenesis. The amygdala also represents an important crossroads because it is extensively and reciprocally interconnected with numerous other brain regions, such as the sensory cortices, thalamus [AMA 92], hippocampus and hypothalamus [DAV 01] as well as the frontal, temporal and parietal lobes [BAR 07, GHA 02, GHA 07]. These numerous connections with other parts of the brain mean that the amygdala is well positioned to modulate not only the activity of the sensory regions but also the cortical areas responsible for cognitive, emotional, somatic, visceral and endocrine processes [DAM 03].

In previous years, there was a boom in studies exploring the function of the amygdala. During this time, there has been a wealth of papers collating evidence around the anatomo-functional correlates supported by the amygdala [AGG 92, AGG 00]. In this way, data have revealed that this region seems to be involved in emotional, memory and attentional processes as well as those of conditioning, social conditions and reward-based learning. We will now look at some of these functions and discuss some of the models that attempt to encapsulate them.

2.5.2. *Amygdala and emotional learning processes*

Emotional learning fulfills a central function by evaluating the predictable results of external events that are relevant to the organism's survival and therefore allow the mobilization of the necessary bodily resources for triggering and coordinating appropriate behavior [LAB 03]. This type of system is particularly useful in dangerous situations that require both rapid reaction and long-term memory of the event. Indeed, the organism may find itself faced with a similar or identically threatening situation. Thus encoding such types of events in long-term memory allows the organism not only to recognize it, but also to react to it adequately and as quickly as possible. The neuroscience has helped to highlight the neural circuits involved in this type of learning and memory, particularly by the use of the conditioning paradigm.

2.5.2.1. *Amygdala and classical conditioning*

In classical conditioning [PAV 27], a meaningless stimulus (conditioned stimulus, abbreviated as CS) acquires affective properties (i.e. creating a fear response) when it occurs shortly following a naturally or biologically threatening event (unconditioned stimulus, abbreviated as US) and that the association of the two stimuli is repeated over time. As soon as the CS–US link has been established, then the fear response alone is induced by the presentation of the conditioned stimulus. For example, if a rat hears a sound (CS) several times that is shortly followed by an electric shock (US), then the exclusive presentation of the sound (CS) creates a fear response once the CS–US link has been made. In classical or Pavlovian conditioning, we traditionally distinguish between two important phases. The first corresponds to the acquisition stage of the association between the CS and US described above. When the CS repeatedly appears without being paired with the

US, the fear response decreases until it disappears. The second phase refers to the extinction process. Neuroimaging data have shown that amygdala activation is higher during the first trials of acquisition [BÜC 98] and extinction [LAB 98]. This specific temporal pattern indicates that the amygdala is involved in learning new conditioned associations and that its role becomes less pronounced once the latter are well established [LAB 03].

2.5.2.2. *The amygdala: a structure with two processing streams*

In line with his first work on animals, LeDoux [LED 96] proposed two processing streams involving the amygdala that may underpin the process of fear conditioning (see Figure 2.5a). Indeed, he describes the characteristics of these two distinct processing streams:

– the sub-cortical, which carries sensory information directly from the thalamus to the amygdala;

– the cortical one, which takes information to the amygdala indirectly via the sensory cortices.

We will now look at some of the arguments originating from case studies and neuroimaging data that support the involvement of the amygdala in fear-conditioning processes in humans. In this vein, Andrilli *et al.* [ANG 96] have shown for example that a patient suffering from brain damage on the right amygdala seems to be completely unresponsive to fear conditioning. Contrary to control participants, he does not show any improvement in blink reflex, a defense response linked to the presence of aversive stimuli. In another study, Bechara *et al*. [BEC 95] sought to understand the role of the amygdala and the hippocampus in emotional conditioning and establishing declarative knowledge. The results revealed the existence of a double dissociation between these two structures and their underlying processes. Indeed, the patient SM, with bilateral damage to the amygdala but an intact hippocampus, had acquired factual knowledge regarding the experimental session, although he did not show any fear conditioned response. In contrast, the patient WC, with lesions to the hippocampus and intact amygdala, displayed the opposite pattern. In this case, the patient was able to acquire a fear conditioned response while having no conscious memory relating to the task he had just completed. These two studies reveal that lesions to the amygdala cause a selective impairment to emotional learning reactions, but do not alter explicit knowledge.

Evidence from neuroimaging has played a large role in reinforcing these initial data while outlining the first arguments for an unconscious processing stream in humans, compatible with LeDoux's double stream model [LED 96]. In this vein, several studies have shown that the right amygdala responds more strongly to anger facial expression presented under the threshold of consciousness [MOR 98, MOR 99]. Furthermore, authors have noted the presence of a correlation between activity in the amygdala with that in the pulvinar and the superior colliculus [MOR 99]. This experimental evidence has led researchers to believe that the emotional value of a stimulus can be extracted unconsciously or automatically via a colliculus–pulvinar–amygdala stream (see Figure 2.5b). In terms of this hypothesis, similar results have been found in blindsight patients, who have cortical blindness but whose peripheral vision organs remain intact. These patients, while blind, are capable of categorizing emotional facial expressions beyond the level of chance [DEG 99]. Other neuroimaging experiments have also indicated that amygdala and pulvinar activities covariate in the same patient [MOR 01]. The indirect evidence from this particular category of patients has provided a strong argument in favor of the existence of a double stream in humans.

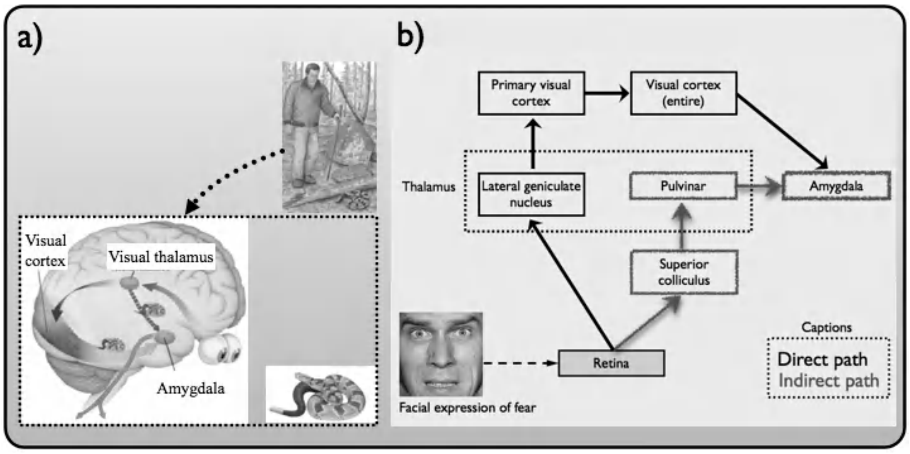


Figure 2.5. *a) Illustration of the two processing streams in an animal as proposed by LeDoux (b) and its extension in human, adapted from [LED 96]*

The second processing stream is cortical, since it conveys sensory information from the thalamus to the visual cortex and other cortical areas such as the extrastriate and fusiform cortices and the superior temporal sulcus before arriving at the amygdala (see Figure 2.5b). Neuroimaging studies have underlined this stream plays a critical role in appraising facial expression [PES 02]. The two processing streams have therefore different functional implications. The direct sub-cortical stream allows the amygdale to have a rough representation extracted rapidly and in a relative independent fashion regarding conscious and voluntary attentional processes. With regard to the indirect/cortical stream, it allows more detailed and refined analysis, which will reach the amygdala later, therefore producing slower conditioned responses. This stream relies on conscious or intentional processing to integrate the information processed at a higher cognitive level [SAN 03].

2.5.3. *The amygdala and emotional perception: hypotheses around the specificity of processing within the amygdala*

Within the field of emotional psychology there are several influential theories including discrete emotion theory, dimensional theories and appraisal theories (see Chapter 1 for more details). As previously mentioned, affective neuroscience has provided evidence that has not only allowed us to critically evaluate the predictions from these theories, but also to *in fine* constraint these different models. With this in mind, studies of processes supported by the amygdala have been highly informative in many ways. This brain structure has been extensively studied and the nature and numbers of processes undertaken by it are still the subject of debate today.

2.5.3.1. *The amygdala as a fear module*

With regard to the brain, core emotion theories propose that different neural substrates underlie each discrete emotion [EKM 99, LED 96]. The majority of research in cognitive neuroscience of emotion has therefore focused on identifying the areas of the brain implementing these different emotions. As such, evidence from various disciplines has highlighted the strong involvement of the amygdala in processing fear-related signals. The convergence of these data has led some authors to propose that the amygdala form the core of a modular cognitive system dedicated to rapid, automatic and specific processing of these threatening stimuli [MIN 02, ÖHM 05, ÖHM 03]. This module would have evolved during the evolution to allow mammals to face recurrent problems encountered in their environment [ÖHM 03]. In this way, the fear module functions as a device that automatically activates when there is a threat and triggers behavioral defense responses on the one hand and associates the psycho-physiological responses to the subjective feeling of fear on the other [ÖHM 04] (see Figure 2.6).

The majority of research carried out on animals indicates that the amygdala is not only necessary for establishing conditioned processes [LED 00] but, more broadly, is part of a specific system that triggers adaptive reactions when faced with danger. Observations made in primates with Klüver-Bucy syndrome, following bilateral amygdala resection, have revealed a permanent disruption of social behavior, and more specifically a reduction in fear responses, as well as an increase in approach behavior [WEI 56]. More recently, a series of studies carried out in brain-damaged patients has indicated that a bilateral lesion of the amygdala provokes impairments in the perception and recognition of fear facial [ADO 94, ADO 95, CAL 01] and vocal [SCO 97] expressions. Finally, results from neuroimaging data have shown that the amygdala participates in processing fear-related faces [MOR 96] and prosody [PHI 98].

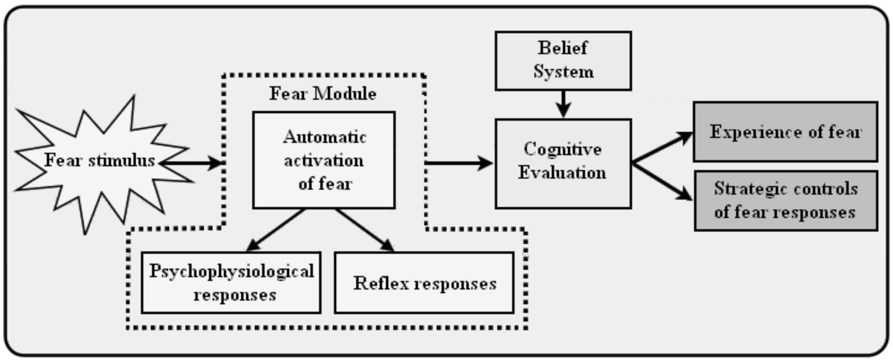


Figure 2.6. *The "fear module", adapted from [ÖHM 04]*

Some research has suggested that the activity of the amygdala is also modulated by other emotions of negative valence. These data give support to the arguments questioning the proposition of a fear module as a selective and exclusive profile of amygdalian computations. Indeed, some studies have shown changes in recognition as well as subjective feeling of emotions, such as sadness and anger as a result of amygdala lesions [FIN 00, LEV 03, RAP 00]. Similarly, some neuroimaging results have corroborated the evidence obtained by lesion studies. As such, they emphasize the presence of amygdala responses to facial expressions of sadness [BLA 99], anger [WRI 02, YAN 02] as well as disgust [SCH 02] and vocal expressions of anger [GRA 05, SCO 97].

A second body of studies, however, goes against the hypothesis of a fear module and a conception, which sees the amygdala as a structure limited to the processing of unpleasant or negative emotions. As such, several recent studies have underlined that the amygdala is involved in processing positive events (for detailed reviews see [SAN 03, SAN 05b, SER 08]). In this vein, there is an abundance of evidence from brain imaging data supporting this proposal. Indeed, such research indicates that the amygdala is activated in response to facial expressions of joy [YAN 02], to words and images with a positive content [HAM 99 and HAM 03, respectively], positive tastes [ODS 01], as well as when viewing entertaining or erotic films [SUG 05]. Finally, the amygdala takes part in the neuronal circuit which implements reward-based learning and positive reinforcement processes [MUR 07]. This structure therefore seems to be a mediator between sensory inputs and their emotional valence, as illustrated by the work of Somerville *et al.* [SOM 06]. In this study, the researchers subjected patients to the task of memorizing unfamiliar faces with an associated positive, negative or neutral description. Interestingly, the results showed that the right amygdala was only activated in response to faces associated with emotional content, independently of valence.

2.5.3.2. *The amygdala and arousal*

There is a second class of influential models in the field of emotion psychology, developed within the dimensional approach. Generally, this theoretical trend proposes that emotions can be represented in a shared, multidimensional space. Among the different models including this approach, bidimensional theory [FEL 99, RUS 99] proposes that emotions can be divided along two orthogonal axes representing the dimensions of valence (pleasant to unpleasant) and arousal (low to high). From these core concepts, some authors have suggested that different regions of the brain underpin these dimensions [HAM 03]. Studies in neuroimaging indicate that there is a dissociation in the processing of these two dimensions for olfactory [AND 03b] and gustatory stimuli [SMA 03]. The results of Anderson *et al.* [AND 03b] have revealed that amygdala activity correlated with intensity of the stimulus but not with the valence of odors, although left ventromedial and right orbital areas of the prefrontal cortex code positive and negative valence of such stimuli, irrespective of their intensity. In the same way, Small *et al.* [SMA 03] found that the medial insula and the amygdala react to the intensity of taste, while the orbitofrontal cortex showed a response pattern specific to gustatory valence.

This theoretical conception of the amygdala as a brain center dedicated to the processing of highly arousing emotions is not exempt for criticisms. Indeed, as Sander *et al.* [SAN 05b] have highlighted in their review, the concept of intensity presented in such research is not equivalent to the notion of physiological arousal proposed theoretically. Furthermore, data have indicated that the amygdala responds to the presentation of sad facial expressions, a type of event that induces weak responses in terms of arousal [BLA 99]. Moreover, Whalen *et al.* [WHA 01] have reported that the amygdala is activated differentially in response to fear and anger, which are theoretically found on the same level on the arousal dimension. Finally, in another study, Winston *et al.* [WIN 05] tested an alternative hypothesis concerning a possible non-linear activation of the amygdala in response to the valence of the stimulus. To this end, the authors measured the response of the amygdala to low or high concentration of pleasant, neutral and unpleasant smells. The results revealed that the effect of intensity on amygdala activity is not the same for all levels of valence. Indeed, the amygdala responds differentially to pleasant and unpleasant smells with high intensity but not to neutral odors. This data therefore indicated that the amygdala does not code intensity or valence *per se*, but that they integrate a combination of these two attributes, which better accounts for the emotional properties contained in a stimulus.

2.5.3.3. *The amygdala as a relevance detector*

Given the limits of classical approaches in emotion psychology to account for activation patterns in the amygdala, some authors have proposed an alternative theoretical framework that better grasps the complexity of processing carried out by this structure. This new perspective is found in appraisal theories, which propose that genesis and emotional differentiation are carried out on the basis of cognitive evaluation processes concerning the significance and consequences of an external event in relation with an individual's goals and needs (see Chapter 1 for further information). On the basis of these theoretical considerations, Sander *et al.* [SAN 03] have proposed to consider the amygdala as a central structure in the detection of relevant events. From this viewpoint, they suggested that there are two reflection axes that account for the functions carried out by the amygdala [SAN 03, SAN 05b]. On the one hand, they proposed that different parts of the amygdala can implement distinct processes and, on the other hand, that a detailed analysis of the tasks and stimuli to which this structure responds could lead to the identification of a common computational profile for the different nuclei it encompasses.

Thereby, a review of the literature in neuroscience tends to show that the adoption of such a perspective better explains the results observed in these studies. An interesting point in this approach lies in the fact that intra-subject variables can modulate amygdala responses. In fact, the conclusions of some studies have indicated that the amygdala was activated in all participants in response to fear facial expressions, but they responded differently when presented with positive facial expressions, depending on the participants' degree of extraversion [CAN 02]. Furthermore, some data have indicated that appraising an event is susceptible to modulate the activity of the amygdala. Canli *et al*., for example, have found that the degree of activation in the left amygdala during encoding can predict the performance of a memory task, but only for scenes that have subjectively been evaluated as being more emotionally intense [CAN 00]. Furthermore, other works led us to think that a person's current needs influence responses within the amygdala [HIN 04, LAB 01]. For example, Hinton *et al*. [HIN 04] observed an increased activity of the amygdala when participants imagined tasting their favorite food in a restaurant, but only when they were hungry and not when full.

Furthermore, a large body of data suggests that during the evolutionary process the amygdala has extended the scope of its computations to integrate relevant events, as outlined by appraisal theories. Nowadays the decoding and expression of social signals takes on a determining importance in the generation of appropriate behavior. The expansion of processes undertaken by the amygdala could therefore reflect the necessity to implement more refined emotional responses in socially complex contexts and, as a result, to favor the adaptation of an individual to his or her environment. Numerous studies have shown that the amygdala participates in processing complex social signals [ADO 03a, ADO 03b, ADO 08] as well as in analyzing some *a priori* non-emotional attributes. Among the latter, the gaze [GEO 01], group belonging [HAR 00, PHE 00], facial familiarity [DUB 99], trustworthiness [KOS 05, TOD 08] as well as the ambiguity of facial expressed cues [ADA 03] all enhance amygdala activity. The processes undertaken by the amygdala therefore seem to have become more complex through evolution. Recent studies have indicated that the amygdala is part of a larger neural network involved in social judgments [ADO 01, WIN 03] as well as in decision-making [DEM 06]. These processes of a higher degree of complexity require, among other things, an appraisal about the relevance and consequences of the choices one individual can make.

This insight into the processes supported by the amygdala shows the need to design innovative and ecological experimental paradigms in order to test the predictions made by different theories within the field of emotion psychology. It also demonstrates the need to improve the models about the normal and pathological functioning of such processes.

2.5.4. *The amygdala and memory processing*

The human memory is not a single, indivisible entity, but is made up of several systems containing, among others, short- and long-term memory, declarative[2] and non-declarative memory[3] as well as semantic[4] and episodic memory[5]. It also involves a number of processes, such as encoding,[6] consolidating and remembering events. Cognitive neuroscience of emotion has begun to reveal the psychological processes and neural mechanisms that can explain why emotional memories about episodes of life often acquired a privileged status in the human brain, via mechanisms that reinforce memory consolidation for example.

---

2 This type of memory is involved in storing and retrieving memories that a person can consciously express using language.

3 This type of memory, also called procedural memory, refers to memories of skills, expertise and associations whose details are not consciously accessible and cannot be verbally explained (e.g. we cannot simply and precisely describe how to ride a bike).

4 Semantic memory is part of the long-term memory and refers to general or factual knowledge that we have about the world and ourselves. It is independent of the spatio-temporal context of encoding (e.g. we know that Paris is the capital of France without having to remember the date, place or even the people present when this information entered into our memory).

5 This memory, which is also linked to long-term memory, refers to the storing and retrieval of personally experienced memories for which we have a spatiotemporal context (detailed description of the memory).

6 Encoding refers to the process of memorizing information.

In line with the results obtained through lesion studies [ADO 97, CAH 96], research in neuroscience has suggested that the amygdala can mediate the effect of memory facilitation for emotional stimuli by potentiating long-term memory traces for this type of material [CAN 00, HAM 99]. Neuroimaging studies have the advantage of being able to investigate the nature of memory processes in which the amygdala seems to participate. Indeed, Hamann *et al*. [HAM 99] have observed that activity of the amygdala during the encoding phase correlates with the performance of a retrieval task achieved several weeks later. Furthermore, additional data have revealed that emotional events of positive and negative valence both benefit from deeper encoding as well as better retrieval as compared with neutral events (for a review, see [LAB 06]). These studies suggest that the amygdala plays a key role at the storing stage as well as during the consolidation of emotional memories. The numerous connections that the amygdala has with other structures involved in long-term memory, such as the hippocampus, the entorhinal, perirhinal and parahippocampal cortices, make the modulation of these processes possible.

Furthermore, data from literature on animals have indicated that the amygdala also takes part in the retrieval process of information stored in memory [MAR 01a, MAR 01b, SMI 04a, SMI 04b]. Concomitant activation of the amygdala with the retrieval of an emotional memory would allow us to consolidate again the memory trace representing this particular event [NAD 00]. The results of recent scientific studies in humans have suggested that the amygdala is involved in the processes of retrieving emotional episodes. Indeed, a participant can rely on different strategies to gain access to a stored episodic memory. This can either be based on a feeling of familiarity or on the process of remembering, these two processes requiring both distinct memory systems and neural substrates [YON 05]. In particular, the remembering process involves retrieving information about the target and its spatial and temporal encoding context, which contains sensory–perceptive and affective details. Studies have revealed that the amygdala and the hippocampus are activated particularly in response to emotional items retrieved by a remembering process as compared with those originating from a feeling of familiarity [DOL 04, SHA 04]. These results demonstrate that the role of the amygdala in memory processes is not limited to encoding and consolidation but also extends to retrieving emotional memories stored in episodic memory and, more specifically, those that are phenomenologically rich.

2.6. The prefrontal cortex

The prefrontal cortex is located within the frontal part of the brain anterior to motor and premotor areas. It contains a set of vast heterogeneous cortical territories [ROL 04] some of them underlying among other things the executive aspects of cognitive processes as well as a wide range of affective mechanisms. Among the structures involved in emotional processes, the orbitofrontal cortex [PRI 07], the anterior cingulate cortex [BUS 00] and the insula [CAL 01] are the most commonly mentioned areas. Given the large surface covered by this set of brain regions, the prefrontal cortex is further divided into functional subdivisions that are specific to certain areas and common to several of them (see Figure 2.7) [PES 08, RUL 02].

Historically, the first anatomo-functional correlates in the prefrontal cortex emerged from the study of a patient suffering from a cerebral stroke of this brain area acquired consecutively to a site accident. Phineas Gage barely survived an explosion that thrust a metal bar 3 cm in diameter through his left eye and the top of his skull. Doctor Harlow's descriptions [HAR 68] revealed that Phineas Gage, a man known for his kindness, work ethic, intelligence and social standing, was "no longer Gage". Following his accident, Gage was described as scornful, impatient, quick to anger and impulsive. The lesion to the brain had therefore caused significant changes both in Gage's personality and his emotional behavior. In a more recent study, Damasio *et al.* [DAM 94] sought to disentangle controversies about the probable location of Gage's brain lesions. To this end, the authors have combined a reconstruction of Gage's skull and the use of modern brain imaging techniques. The results revealed bilateral damage to the prefrontal cortex, which would explain the difficulties in emotional processing and decision-making, symptoms frequently observed in patients presenting such brain lesions.

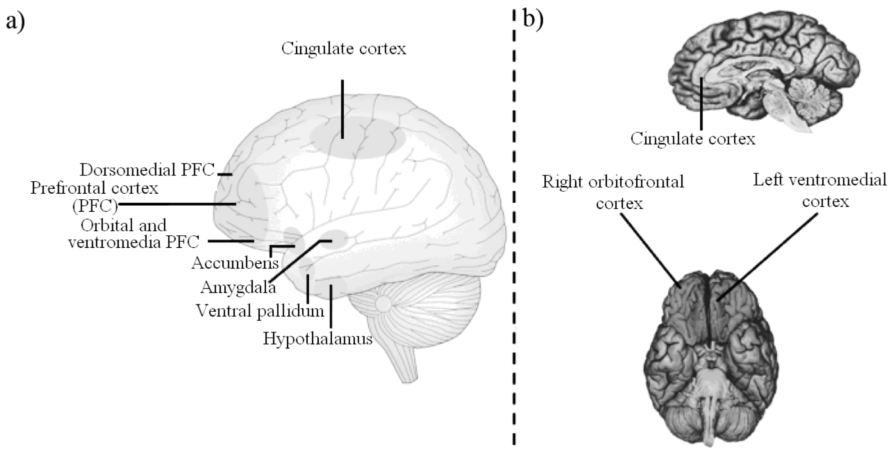


Figure 2.7. *Illustration of the main structures of the so-called "emotional brain". These two figures present the structures that play a central role in emotional processing. (a) Lateral view of the brain that does not show the depth of the different structures nor their bilateral distribution (adapted from [DAL 04]). (b) Locations of the prefrontal structure, such as the anterior cingulate cortex, the orbitofrontal cortex and the ventromedial prefrontal cortex (adapted from [DAV 00])*

Nowadays, knowledge about the functions under laid by different areas of the prefrontal cortex is still a subject of debate and has received increasing interest in research which will be described briefly in the next section.

2.6.1. *The prefrontal cortex and bodily signals*

The importance of bodily signals in emotional genesis has already been the object of well-known debate between partisans of James-Lange's peripheralist view and Canon-Bard's centralist approach. On the basis of James-Lange's initial work, Schachter and Singer [SCH 62] proposed that bodily signals are not the only determiners of emotion, but that they interact with cognitive and contextual factors which can modulate the subjective feeling. For example, they observed that identical patterns of arousal correlate with emotional states classified according to different labels, such as anger, joy or fear. Schachter's two-factor theory [SCH 62] has had a long-lasting influence on theoretical concepts in emotional psychology, since it conciliates the viewpoints of both centralist and peripheralist theories.

Following these works, Damasio *et al*. [DAM 96] formulated their hypothesis about somatic markers to account for the role played by bodily feedback in the reasoning and decision-making processes. In their theory, somatic markers represent associations established early on in personal development between certain types of complex situations and the physiological and emotional reactions they have elicited As such, a physiological signature can act as a sort of marker of a situation and its consequences as being favorable (or not) to an individual. The authors think that the integration of interoceptive (internal sensations) and exteroceptive (external stimuli) information is carried out in the ventromedial prefrontal cortex. When an individual is faced with a type of situation that he has previously encountered, the ventromedial prefrontal cortex triggers a pattern of physiological activation as well as the associated somatic state, these processes being either conscious or unconscious. Somatic markers therefore guide an individual's decision-making process in a situation of incertitude where a logical analysis of the possible choices is insufficient and for which the behavioral decision is based on the emotional properties of the stimulus. In the absence of somatic markers, all the alternatives to a situation have an equivalent value and, as such, logical reasoning involves an evaluation of the results for each of these possible options. This pattern of behavior is therefore frequently observed in patients with brain damage to the ventromedial prefrontal cortex who display difficulties in uncertain situations (e.g. social situations) where the subtle emotional value of several stimuli must be processed and integrated [SAV 91].

A well-known experiment elaborated by Bechara *et al.* [BEC 94] submitted patients with damage to the ventromedial prefrontal cortex to a so-called gambling task in which patients were asked to sit in front of four decks of cards identical in appearance. Each pack contained cards that allow you to win or lose money. However, among these four decks of cards, two are considered as "advantageous" packs as they allow the participants to win money in the long term (low immediate wins but less money lost), and the other two were "disadvantageous", because they would cause long-term financial losses (more significant immediate gains but much bigger losses). The instruction given to the participants was to increase their amount of money, unless the experimenter provides them an explicit rule concerning the decks' contents. To reach their objective, the participants had on the one hand, to inhibit the immediate reward (the biggest profit coming from the "disadvantageous" decks) and on the other hand to become more sensitive to delayed rewards. The results of the study showed that the participants belonging to the control group progressively and implicitly integrated the rules of the games, probably based on their bodily sensations and without having necessarily a conscious explanation concerning the choice of cards that they had made. Furthermore, physiological measures of peripheral activity indicated an increase in skin conductance for the control participants before taking a card from one of the "disadvantageous" decks. The participants from the control group therefore progressively learned to make associations between their bodily sensations and the beneficial or harmful consequences linked to their choices. In this sense, they would have developed somatic markers during the first trials of the task which would give them a signal orientating their decision-making for subsequent drawings. In contrast, the patients with damage to the ventromedial prefrontal cortex could not establish such associations and did not display any increase in skin conductance. More recent experiments have shown that other areas of the prefrontal cortex, such as the right dorsolateral prefrontal cortex, also participate in the decision-making process [KNO 06]. This area would be particularly involved in behaviors linked to fairness.

2.6.2. *The prefrontal cortex and the top-down regulation of behavior*

Some authors have seen the prefrontal cortex as a region among others whose function is to send "biased messages" to other areas of the brain to guide behavior in the direction of solutions most likely to reach one goal of the individual [DAV 90, DAV 99]. Behavioral choices are often influenced by affective consequences caused by the situation, even to the extent that these do not represent the most adapted responses for the person. In the gambling task, for example, the participants had to be able to inhibit their tendency to act in the direction of immediate gratification (taking cards from the "disadvantageous" pile) and regulate their behavior to choose the more appropriate option for achieving the fixed objectives to complete the task. The prefrontal cortex therefore favors achieving adaptive aims in situations where several behavioral alternatives competing with each other are linked to immediate emotional consequences [OCH 02].

In his functional asymmetry model, Davidson proposes that the left prefrontal cortex is connected to appealing (positive) goals linked to approach behavior, while the right prefrontal cortex is responsible for maintaining behavior (negative) requiring inhibition or avoidance. In a study using the electroencephalography (EEG) technique, Davidson *et al*. [DAV 90] published results corroborating this hypothesis. They demonstrated that energy in the alpha frequency band[7] was reduced in the left prefrontal cortex when presented with faces of positive valence inducing a tendency of approach and lower in the right prefrontal cortex when exposed to facial expressions of disgust eliciting avoidance. The reduction in energy in this frequency band would indicate a specific increase in brain activity. Furthermore, Davidson *et al*. [DAV 00] have proposed that the emotional regulation mechanism for negative effects, such as anger, involves a neural network in which inhibitory connections in certain parts of the prefrontal cortex (probably the orbitofrontal cortex) would extend top-down projections toward the amygdala. More recent studies in neuroimaging have suggested that during attempts to regulate emotion the prefrontal cortex modulates the activity of the amygdala differently according to the adopted strategy (see, for example, [OCH 04]). Dysfunctions in these processes therefore would constitute a prelude to socially unadapted behaviors as well as aggressiveness and violence. Indeed, there are several neuropsychological studies supporting this view [AND 99, BLA 00]. Similarly, while patients with damage to the orbitofrontal cortex are conscious to current social norms, they are not aware of violating these rules [BEE 06].

2.6.3. *The prefrontal cortex and the motivational component of emotion*

Rolls' work [ROL 04] has suggested that within the prefrontal areas of the brain the orbitofrontal cortex plays a role in learning the emotional and motivational value of a stimulus. Due to its numerous connections with other limbic regions [REM 07], the orbitofrontal cortex is part of a neural circuit that underpins the reward process [TRE 99]. In this way, Rolls [ROL 04] has proposed that this region works with the amygdala to establish and represent associations between primary (e.g. taste or touch) and secondary reinforcers (e.g. learnt associations between visual and olfactive stimuli). Beyond this function, the orbitofrontal cortex can flexibly adjust the behavior of an organism, once the reinforcement contingencies change in the environment. Moreover, impairments of this type of reversible learning have been highlighted in animals following damage to the orbital part of the prefrontal cortex [for a review see ROL 04] as well as in humans by neuropsychological investigations [BEC 97, BER 04].

7 The alpha frequency band corresponds to the brain rhythms oscillating between 8.5 and 12 Hz.

2.7. The anterior cingulate cortex

Generally, the anterior cingulate cortex (ACC) can be considered as a region responsible for integrating information from diverse sources and contributing to modulate the processing that occur in other brain areas. Several evidence suggest that the ACC is in fact an interface between cognitive and affective processes and that it participates more specifically in linking emotional and attentional information [DAV 03a]. Lesion and electrophysiological studies, as well as data from cerebral connectivity studies and a limited number of imaging studies, have indicated that the anterior ACC can be further divided in to two parts, processing affective and cognitive information respectively [BUS 00]. The affective part refers to the ventral and rostral territories of the anterior cingulate cortex. It also has connections with limbic and frontal structures (such as the amygdala, the orbitofrontal cortex and the anterior insula) as well as with the visceral, endocrine and visceromotor systems. Although a theoretical framework integrating in a unified view the different processing carried out by the ACC does not exist yet, the symptoms resulting from brain damage to this area have laid the foundation of the anatomo-functional correlates supported by the ACC. Some of these include apathy, inattention, deregulation of functions carried out by the autonomous system as well as emotional instability, reflecting the role of the AAC in emotional feeling, mirroring what Papez had proposed [PAP 37]. Other empirical evidence has further revealed that the affective subdivision of the ACC seems to be involved in evaluating the emotional and motivational value of an event and in controlling emotional responses requiring effortful mental resources [DEV 95, DRE 98, VOG 92, WHA 98]. Finally, some data seem to support a link between dysfunction within the ACC and psychopathological states. For example, hyperactivation in the ACC correlates with anxiety and social phobia [DAV 03a].

The cognitive part is located in the dorsal division of the ACC. Without going into detail, this area of the brain plays a functional role in detecting and controlling errors [DAV 03a]. Some researchers have suggested that it is the centre of the attentional supervision system proposed in Norman and Shallice’s model [NOR 86], which intervenes when automatic action routines are not sufficient to respond to the task’s demands.

2.8. The role of the insula in disgust

The majority of experiments into the neural substrates implementing disgust have highlighted the major role of two brain structures: the insula and the basal ganglia. Neuroimaging data have indicated that these regions are activated when presenting with facial and vocal expressions of disgust [PHI 98]. Furthermore, neuropsychological results have emphasized the existence of a double dissociation in the processes underlying the recognition of fear and disgust in patients respectively suffering from brain damage to the amygdala and the insula. The invert pattern observed in these two patients indicates that lesions to the insula cortex provoke selective impairments in recognizing disgust, which do not reflect more general disruptions in the ability to identify emotions with a negative valence [PHI 98, SPR 98]. Furthermore, while patients display impairments in recognizing disgust, they are capable of naming situations where a person might feel disgust and therefore do not seem to have problems with understanding the concept of disgust [CAL 01]. These results are particularly interesting because the insula receives afferent inputs from the gustatory system [SMA 99] and projects efferent connections in the direction of the majority of regions playing a role in regulating the responses of the autonomous system. Finally, a synthesis of studies examining patients and control participants [DAV 99] argues for the implications of this structure in bodily states.

2.9. Temporal dynamic of brain processes in emotional genesis

As highlighted several times in this chapter, we consider emotion as a complex process taking place across a period of time. The brain processes linked to emotion have often been studied using techniques that cannot account for the complex temporal dynamics of the processes underpinning emotional mechanisms, which is particularly the case with fMRI. There are other methods for studying these mechanisms, however, in particular EEG and magneto encephalography, which can characterize electric and magnetic brain processes in order to elucidate their temporal dynamics. Perception of emotional facial expressions has been a subject of great interest within neuroscience over the past 10 years. Adolphs [ADO 02], as previously mentioned, proposed a temporal dynamic model as well as conscious and unconscious mechanisms that are involved in the perception of emotion in the face. Other works have also studied the temporal dynamics of emotional facial perception using EEG [BAL 08, ESS 04, POU 04] and magneto encephalography [LUO 07, STR 99].

Beyond these works, Grandjean and Scherer [GRA 08a] have recently tested Scherer's 1984 hypothesis about the temporal sequence of appraisal processes. This view is opposed to other theoretical positions, where the appraisal evaluation checks responsible for emotional genesis would occur in parallel or would be determined by the context [ELL 91, SMI 90]. By the systematic experimental manipulation of novelty, intrinsic pleasantness, goal and need relevance, goal conduciveness in two EEG studies, it has been possible to highlight the sequence of brain processes linked to these experimental manipulations, although massive parallel processes remain still present (see [GRA 08a] for further details). Such parallel processes are, for example, linked to the notion of levels of processing. At the first stage, our cognitive system can for example detect the novelty without us being aware that something new has appeared in the environment. The fact that we become aware in an explainable way of a new stimulus involves complex representational mechanisms which work in parallel with other appraisal evaluation checks. Thus affective neuroscience and the tools enabling us to study the brain areas involved in emotional processes as well as methods allowing the investigation of temporal dynamics of such processes, can constrain and refine psychological model of emotion and, as such, improve our knowledge about emotion. In order to tend to a more complete picture of brain mechanisms involved in emotion, the connectivity between various areas of the brain and, in particular, the interactions of regions characterized by a given computational profile, remains a key issue to better account for all the processes participating in emotion phenomena.

2.10. Functional connectivity

Beyond highlighting the areas of the brain involved in a given cognitive or emotional process, cognitive neuroscience today seeks to test models about the brain connectivity contingencies involved during certain cognitive and/or affective processing. There are currently different methods enabling more or less direct investigations concerning the brain connectivity as revealed through different types of indicators.

2.10.1. *Investigations of the connectivity using brain imaging techniques (MRI)*

One of the most favored methods for examining brain connectivity in animals consists of using tracers to highlight the areas of the brain exchanging information. To this end, a given area of the brain, such as a nucleus of the amygdala, is injected with a product that spreads to the regions connected to this nucleus. These injected areas can then be studied in detail *postmortem*. However, this method is not suitable to study temporal dynamic aspects of the connectivity between brain regions.

Another method developed in the 1990s, called diffusion tensor imaging, or DTI, is used to study connectivity in humans using a signal obtained from MRI. This method allows us to see the anatomical connectivity between areas of the brain, but again does not provide any information about temporal dynamics of the connections between different parts of the brain. As an example, this method has been used in humans to study the connectivity between visual and amygdala regions and has revealed the presence of direct connections between these two brain areas via a beam which might be involved in modulation of primary visual sites by the amygdala [CAT 03].

Analysis of the correlations of fMRI signals between one or more areas of the brain is also used to study functional connectivity. Unfortunately, in this situation inferring a direct or even causal relationship between one or more areas is not possible. Instead, this link is highly indirect and does not constitute genuine proof of functional connectivity. Statistical models, such as the Dynamic Causal Modeling, have been developed to systematically study the relations between different areas of the brain using the BOLD signal borrowed from dynamic causal modeling. This technique allows us to study the link between areas of the brain but again does not highlight the temporal dynamic of relations between these same areas. Furthermore, this method requires a set of alternative models of connectivity between several areas and to test which one can best account for covariances observed in fMRI signals between different areas of the brain. Ethofer *et al.* [ETH 06], for example, have highlighted that there is bilateral connectivity between the right temporal region and the right and left frontal regions during the processing of emotional prosody.

2.10.2. *Investigations into connectivity using electroencephalographic (EEG) techniques*

Intracranial local field potential recordings is a direct method allowing to study the functional connectivity dynamically and selectively. It consists of recording in situ the electrical activity of large neuronal assemblies by using macro-electrodes directly located in the brain tissue. This method is used to study interactions between areas of the brain in both animal and human populations. In fact, some disorders of the central nervous system such as pharmaco-resistant forms of epilepsy, can require investigation into the spontaneous electrical activity of brain areas involved in the genesis of an epileptic focus. Following clinical purposes, deep brain electrodes are implanted within the brain with the aim of identifying the areas triggering epileptic seizures whose surgical resection may potentially contribute to reduce or abolish. These patients can take part in simple experimental protocols that allow the joint recording of brain signals within sub-cortical and cortical regions and therefore the study of dynamic relations between these same brain sites. In this clinical context, for ethical reasons implanting electrodes in humans is obviously constrained by merely medical reasons, and as a result, does allow us to study certain brain areas in a restricted number of individuals. The study of the phase[8] of the signals obtained from intracranial recording therefore allows the subsequent exploration of the phase lag (the difference between the phases of two oscillatory signals) between the signals recorded in two more or less distant brain areas by the use of particular designed algorithms developed for this purpose (see, for example, [LAC 99]). This technique can also be applied to surface measures (e.g. on magneto

8 Phase refers to the angular difference between two periodic phenomena.

encephalographic or MEG signals), although this involves a significant reduction in anatomical precision. For example, Guderian and Duzel [GUD 05] have highlighted significant neuronal synchronization in theta frequency band between the temporal and anterior regions when trying to remember contextual elements compared to situations where participants were unable to remember such encoding aspects. This study therefore demonstrates that it is possible to study the relations between areas of the brain non-invasively by the use of surface MEG measures.

Similarly, Pascal Fries [FRI 05] has developed a model of neuronal communication in which the phase coupling represents the *sine qua non* condition allowing areas of the brain discharging jointly to exchange information. According to Fries, brain regions which would not be phase-locked would in contrast not be able to communicate. The physiological constraints due to the so-called refactory period in a neuron or in a group of neurons, the fact that these neurons must have a period of "rest" before being able to discharge again, mean they would not or be less able to integrate incoming information from other groups of neurons [FRI 05]. It has been suggested that the emergence of the subjective feeling could result from neuronal synchronization occurring at different levels which in turn allow us to put together in an integrated representation different emotional components [GRA 08b].

2.10.3. *Benefits of brain connectivity studies*

This section will detail some examples of empirical studies that have sought to examine the links between the amygdala and the orbitofrontal regions. A wide range of studies using various techniques, such as single or multi-unit recordings, brain imaging and brain lesion studies, have shown that these areas of the brain are particularly crucial in emotional processes. As such, these two regions should have the anatomo-functional connections which would play a role in elaborating an integrated representation of stimulus involving emotional mechanisms.

The anatomical description of connectivity highlights the potential links between these structures such as, for example, the connections between amygdala nuclei and different parts of the orbitofrontal cortex. Indeed, several studies using tracers in animals, particularly in Rhesus monkeys, have sought to demonstrate the relations between anterior regions of the brain, such as the orbitofrontal cortex and the amygdala. The results of these studies have indicated the presence of a selective functional connectivity between the medial parts of the orbitofrontal cortex and nuclei within the amygdala since the degree of synchronization between amygdala neuronal population and the anterior dorsolateral regions of the brain were almost non-existent [GHA 02, GHA 07, POR 81]. Moreover, the study of neuronal responses to olfactory conditioning in rats has shown modulation in basolateral amygdaloid nuclei activity in particular contingencies (e.g. the association between positive and negative smells to sweet or bitter liquids) and the involvement of orbitofrontal areas is involved in changes in these contingencies using the reversal learning paradigm [SCH 00, SCH 03, SCH 06]. In this case, a smell previously associated with a positive stimulus is paired with a negative stimulus. The changes observed in rats between the amygdala and orbitofrontal areas in such experimental paradigm highlight the shared role of these two regions in adjusting behavior regarding one stimulus whose associative contingencies have been modified over time. This is reinforced by studies where the damage to orbitofrontal regions in animals has not prevented them from adapting their behavior to new contingencies and has meant that they persist in previous behavior by reverse contingencies [EIC 83, MEU 97]. Observations carried out in human brain-damaged patients go in the same direction [HOR 04, ROL 94].

Affective neuroscience and, more generally, cognitive neuroscience are currently concerned with characterizing the relations between brain areas in order to account for our cognitive system's ability to process and integrate internal and external information that then leads to behavioral contingencies that are more or less plastic and sensitive to learning. The emergence of this area of study into brain connectivity is, without doubt, a key element in the adoption of new perspectives aiming to understand the relations between cognitive and emotional systems. This area of study also allows us to see the cerebral bases of affective systems, not as regions of the brain but as distributed neural networks whose relational contingencies are a determining factor in understanding and modeling the central nervous system and its relations with peripheral and viscero-somatic systems.

2.11. Conclusion

For more than a century, pioneers in the study of the brain mechanisms underlying emotional processes have paved the way to a rich and informative perspective on human behaviors and the means of modeling them. The later and systematic study of such mechanisms has benefited from the recent emergence of affective neuroscience, which has contributed to the enhancement and reinforcement knowledge gained from other, more traditional, techniques used to identify the function of specific brain areas involved in emotional processes. The use of multidisciplinary methods, constitutive elements of the neuroscientist approach has provided a more accurate overview of the different components involved during an emotional episode and has allowed us to identify several levels of study in emotional processes. The study of brain correlates in emotional processes has meant that we have been able to refine and constrain the models set out in emotional psychology and neuropsychology. In the clinical domain, a better understanding of the functional organization of the human brain enables the development of neuro-rehabilitation programs. The areas of application of this knowledge are vast and affect various domains of study as well as diverse emotion-related problems such as marketing or stress management in companies. If we are interested in modeling behavior in virtual machines or robots, on the other hand, we need extensive knowledge of how the human system processes affective information as well as of the way such processing modulates behavior and other cognitive functions such as decision-making.

2.12. Bibliography

[AAL 02] AALTO S., NÄÄTÄNEN P., WALLIUS E., METSÄHONKALA L., STENMAN H., NIEM P.M. *et al.*, "Neuroanatomical substrata of amusement and sadness: A PET arousal study using film stimuli", *Neuro Report*, vol. 13, pp. 67-73, 2002.

[ADA 03] ADAMS R.B., GORDON H.L., BAIRD A.A., AMBADY N., KLECK R.E.,"Effects of gaze on amygdala sensitivity to anger and fear faces", *Science*, vol. 300, p. 1536, 2003.

[ADO 94] ADOLPHS R., TRANEL D., DAMASIO H., DAMASIO A., "Impaired recognition of emotion in facial expressions following bilateral damage to the human amygdale", *Nature*, vol. 372, pp. 669-672, 1994.

[ADO 95] ADOLPHS R., TRANEL D., DAMASIO H., DAMASIO A.R., "Fear and the human amygdala", *J. Neurosci.*, vol. 15, no. 9, pp. 5879-5891, 1995.

[ADO 97] ADOLPHS R., CAHILL L., SCHUL R., BABINSKY R., "Impaired declarative memory for emotional material following bilateral amygdala damage in humans", *Learning and Memory*, vol. 4, no. 3, pp. 291-300, 1997.

[ADO 01] ADOLPHS R., "The neurobiology of social cognition", *Curr. Opin. Neurobiol.*, vol. 11, no. 2, pp. 231-239, 2001.

[ADO 02] ADOLPHS R., "Neural systems for recognizing emotion", *Curr. Opin. Neurobiol.*, vol. 12, no. 2, pp. 169-177, 2002.

[ADO 03a] ADOLPHS R., "Cognitive neuroscience of human social behavior", *Nat. Rev. Neurosci.*, vol. 4, no. 3, pp. 165-178, 2003.

[ADO 03b] ADOLPHS R., "Is the human amygdala specialized for processing social information?", *Annals of the New York Academy of Sciences*, vol. 985, pp. 326-340, 2003.

[ADO 08] ADOLPHS R., SPEZIO M.L., PARLIER M., PIVEN J., "Distinct face-processing strategies in parents of autistic children", *Current Biology*, vol. 18, no. 14, pp. 1090-1093, 2008.

[AGG 92] AGGELTON J.-P., *The Amygdala: Neurobiological Aspects of Emotion, Memory and Mental Dysfunction*, Wiley-Liss, New York, USA, 1992.

[AGG 00] AGGELTON J.-P., *The Amygdala: A Functional Analysis*, Oxford University Press, Oxford, 2000.

[AMA 92] AMARAL D.G., PRICE J.L., PITKÄNEN A., CARMICHAEL S.T., "Anatomical organization of the primate amygdaloid complex", in: AGGLETON J.P. (ed.), *The Amygdala: Neurobiological Aspects of Emotion, Memory and Mental Dysfonction*, pp. 1-66, Wiley-Liss, New York, USA, 1992.

[AND 03a] ANDERSON A.K., SOBEL N., "Dissociating intensity from valence as sensory inputs to emotion", *Neuron*, vol. 39, no. 4, pp. 581-583, 2003.

[AND 03b] ANDERSON A.K., CHRISTOFF K., STAPPEN I., PANITZ D., GHAHREMANI D.G., GLOVER G. *et al.*, "Dissociated neural representations of intensity and valence in human olfaction", *Nat. Neurosci.*, vol. 6, no. 2, pp. 196-202, 2003.

[AND 99] ANDERSON S.W., BECHARA A., DAMASIO H., TRANEL D., DAMASIO A.R., "Impairment of social and moral behavior related to early damage in human prefrontal cortex", *Nat. Neurosci.*, vol. 2, no. 11, pp. 1032-1037, 1999.

[ANG 96] ANGRILLI A., MAURI A., PALOMBA D., FLOR H., BIRBAUMER N., SARTORI G. *et al.*, "Startle reflex and emotion modulation impairment after a right amygdala lesion", *Brain: a Journal of Neurology*, vol. 119, no. 6, pp. 1991-2000, 1996.

[BAL 08] BALCONI M., LUCCHIARI C., "Consciousness and arousal effects on emotional face processing as revealed by brain oscillations", *International Journal of Psychophysiology*, vol. 67, no. 1, pp. 41-46, 2008.

[BAR 07] BARBAS H., "Flow of information for emotions through temporal and orbitofrontal pathways", *Journal of Anatomy*, vol. 211, no. 2, pp. 237-249, 2007.

[BAR 28] BARD P.A., "Diencephalic mechanism for the expression of rage with special reference to the central nervous system", *Am. J. Physiol.*, vol. 84, pp. 490-513, 1928.

[BEC 94] BECHARA A., DAMASIO A.R., DAMASIO H., ANDERSON S.W., "Insensitivity to future consequences following damage to humain prefontal cortex", *Cognition*, vol. 50, pp. 7-15, 1994.

[BEC 95] BECHARA A., TRANEL D., DAMASIO H., ADOLPHS R., ROCKLAND C., DAMASIO A.R., "Double dissociation of conditioning and declarative knowledge relative to the amygdala and hippocampus in humans", *Science*, vol. 269, no. 5227, pp. 1115-1118, 1995.

[BEC 97] BECHARA A., DAMASIO H., TRANEL D., DAMASIO A.R., "Deciding advantageously before knowing the advantageous strategy", *Science*, vol. 275, no. 5304, pp. 1293-1295, 1997.

[BEE 06] BEER J.S., JOHN O.P., SCABINI D., KNIGHT R.T., "Orbitofrontal cortex and social behavior: Integrating self-monitoring and emotion-cognition interactions", *J. Cog. Neurosci.*, vol. 18, no. 6, pp. 871-879, 2006.

[BER 04] BERLIN H.A., ROLLS E.T., KISCHKA U., "Impulsivity, time perception, emotion and reinforcement sensitivity in patients with orbitofrontal cortex lesions", *Brain: a Journal of Neurology*, vol. 127, no. 5, pp. 1108-1126, 2004.

[BLA 00] BLAIR R.J.R., CIPOLOTTI L., "Impaired social response reversal: A case of 'acquired sociopathy'", *Brain: a Journal of Neurology*, vol. 123, no. 6, pp. 1122-1141, 2000.

[BLA 99] BLAIR R.J., MORRIS J.S., FRITH C.D., PERRETT D.I., DOLAN R.J., “Dissociable neural responses to facial expressions of sadness and anger”, *Brain: a Journal of Neurology*, vol. 122, no. 5, pp. 883-893, 1999.

[BRO 08] BROSCH T., SANDER D., POURTOIS G., SCHERER K.R., “Beyond fear: Rapid spatial orienting toward positive emotional stimuli”, *Psychological Science: a Journal of the American Psychological Society/APS*, vol. 19, no. 4, pp. 362-70, 2008.

[BÜC 98] BÜCHEL C., MORRIS J., DOLAN R.J., FRISTON K.J., “Brain systems mediating aversive conditioning: an event-related fMRI study”, *Neuron*, vol. 20, pp. 947-957, 1998.

[BUS 00] BUSH G., LUU P., POSNER M.I., “Cognitive and emotional influences in anterior cingulate cortex”, *Trends in Cognitive Sciences*, vol. 4, no. 6, pp. 215-222, 2000.

[CAH 96] CAHILL L., HAIER R.J., FALLON J., ALKIRE M.T., TANG C., KEATOR D. *et al.*, “Amygdala activity at encoding correlated with long-term, free recall of emotional information”, *Proceedings of the National Academy of Sciences of the United States of America*, vol. 93, no. 15, pp. 8016-8021, 1996.

[CAL 01] CALDER A.J., LAWRENCE A.D., YOUNG A.W., “Neuropsychology of fear and loathing”, *Nature Reviews Neuroscience*, vol. 2, no. 5, pp. 352-363, 2001.

[CAN 00] CANLI T., ZHAO Z., BREWER J., GABRIELI J.D.E., CAHILL L., “Event related arousal in the human amygdala associates with later memory for individual emotional experience”, *J. Neurosci.*, vol. 20, no. 99, pp. 1-5, 2000.

[CAN 02] CANLI T., SIVERS H., WHITFIELD S.L., GOTLIB I.H., GABRIELI J.D., “Amygdala response to happy faces as a function of extraversion”, *Science*, vol. 21, pp. 2191, 2002.

[CAN 29] CANNON W.B., *Bodily Changes in Pain, Hunger, Fear and Rage*, Appleton, New York, USA, 1929.

[CAT 03] CATANI M., JONES D.K., DONATO R., FFYTCHE D.H., “Occipito-temporal connections in the human brain”, *Brain*, vol. 126, no. 9, pp. 2093-2107, 2003.

[DAL 04] DALGLEISH T., “The emotional brain”, *Nat. Rev. Neurosci.*, vol. 5, no. 7, pp. 583-589, 2004.

[DAM 94] DAMASIO H., GRABOWSKI T., FRANK R., GALABURDA A.M., DAMASIO A.R., “The return of Phineas Gage: clues about the brain from the skull of a famous patient”, *Science*, vol. 264, no. 5162, pp. 1102-1105, 1994.

[DAM 96] DAMASIO A.R., “The somatic marker hypothesis and the possible functions of the prefrontal cortex”, *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, vol. 351, no. 1346, pp. 1413-1420, 1996.

[DAM 03] DAMASIO A.R., ADOLPHS A., DAMASIO H., “The contributions of the lesion method to the functional neuroanatomy of emotions”, in: DAVIDSON R.J., SCHERER K.R. and GOLDSMITH H.H. (eds), *Handbook of Affective Sciences*, pp. 66-92, Oxford University Press, Oxford, 2003.

[DAR 72] DARWIN C., *The Expression of the Emotions in Man and Animals*, Chicago University Press, Chicago, USA, 1872/1965.

[DAV 90] DAVIDSON R.J., EKMAN P., SARON C.D., SENULIS J.A., FRIESEN W.V., "Approach-withdrawal and cerebral asymmetry: Emotional expression and brain physiology", *Journal of Personality and Social Psychology*, vol. 58, no. 2, pp. 330-341, 1990.

[DAV 95] DAVIDSON R.J., SUTTON S.K., "Affective neuroscience: The emergence of a discipline", *Curr. Opin. Neurobiol.*, vol. 5, no. 2, pp. 217-224, 1995.

[DAV 99] DAVIDSON R.J., IRWIN W., "The functional neuroanatomy of emotion and affective style", *Trends Cog. Sci.*, vol. 3, no. 1, pp. 11-21, 1999.

[DAV 00] DAVIDSON R.J., PUTNAM K.M., LARSON C.L., "Dysfunction in the neural circuitry of emotion regulation – a possible prelude to violence", *Science*, vol. 289, no. 5479, pp. 591-594, 2000.

[DAV 01] DAVIS M., WHALEN P.J., "The amygdala: vigilance and emotion", *Molecular Psychiatry*, vol. 6, no. 1, pp. 13-34, 2001.

[DAV 03a] DAVIDSON J.R., PIZZAGALLI D., NITSCHKE J.B., KALIN N.K., "Parsing the subcomponents of emotion and disorders of emotion: Perspectives from affective neuroscience", in: DAVIDSON R.J., SCHERER K.R. and GOLDSMITH H.H. (eds), *Handbook of Affective Sciences*, pp. 8-24, Oxford University Press, Oxford, 2003.

[DAV 03b] DAVIDSON R.J., "Seven sins in the study of emotion: correctives from affective neuroscience", *Brain and Cognition*, vol. 52, no. 1, pp. 129-132, 2003.

[DAV 03c] DAVIDSON R.J., SCHERER K.R., GOLDSMITH H.H., "Introduction", in: DAVIDSON R.J., SCHERER K.R. and GOLDSMITH H.H. (eds), *Handbook of Affective Science*, Oxford University Press, Oxford, 2003.

[DEG 99] DE GELDER B., VROOMEN J., POURTOIS G., WEISKRANTZ L., "Nonconscious recognition of affect in the absence of the striate cortex", *Neuro Report*, vol. 10, pp. 3759-3763, 1999.

[DEM 06] DE MARTINO B., KUMARAN D., SEYMOUR B., DOLAN R.J., "Frames, biases, and rational decision-making in the human brain", *Science*, vol. 313, no.5787, pp. 684-687, 2006.

[DES 49] DESCARTES R., *Les Passions de l'Âme*, Napoléon Chaix, Paris, 1649.

[DEV 95] DEVINSKY O., MORRELL M.J., VOGT B.A., "Contributions of anterior cingulate cortex to behavior", *Brain: a Journal of Neurology*, vol. 118, no. 1, pp. 279-306, 1995.

[DOL 04] DOLCOS F., LA BAR K.S., CABEZA R., "Interaction between the amygdala and the medial temporal lobe memory system predicts better memory for emotional events", *Neuron*, vol. 42, no. 5, pp. 855-863, 2004.

[DRE 98] DREVETS W.C., RAICHLE M.E., "Reciprocal suppression of regional cerebral blood flow during emotional versus higher cognitive processes: implications for interactions between emotion and cognition", *Cognition Emotion*, vol. 12, pp. 353-385, 1998.

[DUB 99] DUBOIS S., ROSSION B., SCHILTZ C., BODART J.M., MICHEL C., BRUYER R. *et al.*, "Effect of familiarity on the processing of human faces", *Neuro Image*, vol. 9, no. 3, pp. 278-289, 1999.

[EIC 83] EICHENBAUM H., CLEGG R.A., FEELEY A., "Reexamination of functional subdivisions of the rodent prefrontal cortex", *Exp. Neurol.*, vol. 79, no. 2, pp. 434-451, 1983.

[EKM 73] EKMAN P., *Darwin and Facial Expression: a Century of Research in Review*, Academic, New York, USA, 1973.

[EKM 99] EKMAN P., "Core emotions", in: DALGEISH T. and POWER M. (eds), *Handbook of Cognition and Emotion*, pp. 45-60, Wiley, Chichester, 1999.

[ELL 91] ELLSWORTH P.C., "Some implications of cognitive appraisal theories of emotion", in: STRONGMAN K.T. (ed.), *International Review of Studies of Emotion*, vol. 1, pp. 143-161, Wiley, New York, USA, 1991.

[ESS 04] ESSLEN M., PASCUAL-MARQUI R.D., HELL D., KOCHI K., LEHMANN D., "Brain areas and time course of emotional processing", *Neuro Image*, vol. 21, no. 4, pp. 1189-1203, 2004.

[ETH 06] ETHOFER T., ANDERS S., ERB M., HERBERT C., WIETHOFF S., KISSLER J. *et al.*, "Cerebral pathways in processing of affective prosody: a dynamic causal modeling study", *Neuro Image*, vol. 30, no. 2, pp. 580-587, 2006.

[FAN 07] FAN J., BYRNE J., WORDEN M.S., GUISE K.G., MCCANDLISS B.D., FOSSELLA J., POSNER M.I., "The relation of brain oscillations to attentional networks", *The Journal of Neuroscience: the Official Journal of the Society for Neuroscience*, vol. 27, no. 23, pp. 6197-206, 2007.

[FEL 99] FELDMAN BARRETT L., RUSSELL J.A., "The structure of current affect: Controversies and emerging consensus", *Current Directions in Psychological Science*, vol. 8, no. 1, pp. 10-14, 1999.

[FIN 00] FINE C., BLAIR R.J.R., "The cognitive and emotional effects of amygdala damage", *Neuro Report*, vol. 6, pp. 435-440, 2000.

[FRE 00] FREY S., KOSTOPOULOS P., PETRIDES M., "Orbitofrontal involvement in the processing of unpleasant auditory information", *Eur. J. Neurosci.*, vol. 12, no. 10, pp. 3709-3712, 2000.

[FRI 05] FRIES P., "A mechanism for cognitive dynamics: neuronal communication through neuronal coherence", *Trends Cog. Sci.*, vol. 9, no. 10, pp. 474-480, 2005.

[GEO 01] GEORGE N., DRIVER J., DOLAN R.J., "Seen gaze-direction modulates fusiform activity and its coupling with other brain areas during face processing", *Neuro Image*, vol. 13, no. 6, pp. 1102-1112, 2001.

[GHA 02] GHASHGHAEI H.T., BARBAS H., "Pathways for emotion: Interactions of prefrontal and anterior temporal pathways in the amygdala of the rhesus monkey", *Neurosci.*, vol. 115, no. 4, pp. 1261-1279, 2002.

[GHA 07] GHASHGHAEI H.T., HILGETAG C.C., BARBAS H., "Sequence of information processing for emotions based on the anatomic dialogue between prefrontal cortex and amygdale", *Neuro Image*, vol. 34, no. 3, pp. 905-923, 2007.

[GRA 05] GRANDJEAN D., SANDER D., POURTOIS G., SCHWARTZ S., SEGHIER M.L., SCHERER K.R. *et al.*, "The voices of wrath: brain responses to angry prosody in meaningless speech", *Nat. Neurosci.*, vol. 8, no. 2, pp. 145-146, 2005.

[GRA 08a] GRANDJEAN D., SCHERER K.R., "Unpacking the cognitive architecture of emotion processes", *Emotion*, vol. 8, no. 3, pp. 341-351, 2008.

[GRA 08b] GRANDJEAN D., SANDER D., SCHERER K.R., "Conscious emotional experience emerges as a function of multilevel, appraisal-driven response synchronization", *Conscious Cogn.*, vol. 17, no. 2, pp. 484-495, 2008.

[GRA 08c] GRANDJEAN D., SANDER D., LUCAS N., SCHERER K.R., VUILLEUMIER P., "Effects of emotional prosody on auditory extinction for voices in patients with spatial neglect", *Neuropsychologia*, vol. 46, no. 2, pp. 487-496, 2008.

[GUD 05] GUDERIAN S., DÜZEL E., "Induced theta oscillations mediate large-scale synchrony with mediotemporal areas during recollection in humans", *Hippocampus*, vol. 15, no. 7, pp. 901-912, 2005.

[HAM 99] HAMANN S.B., ELY T.D., GRAFTON S.T., KILTS C.D., "Amygdala activity related to enhanced memory for pleasant and aversive stimuli", *Nat. Neurosci.*, vol. 2, pp. 289-293, 1999.

[HAM 03] HAMANN S., "Nosing in on the emotional brain", *Nat. Neurosci.*, vol. 6, no. 2, pp. 106-8, 2003.

[HAR 68] HARLOW J.M., "Recovery of the passage of an iron bar through the head", *Publ. Mass. Med. Soc*, vol. 2, pp. 327-334, 1868.

[HAR 00] HART A.J., WHALEN P.J., SHIN L.M., MC INERNEY S.C., FISCHER H., RAUCH S.L., "Differential response in the human amygdala to racial outgroup vs ingroup face stimuli", *Neuro Report*, vol. 11, no. 11, pp. 2351-2355, 2000.

[HIN 04] HINTON E.C., PARKINSON J.A., HOLLAND A.J., ARANA F.S., ROBERTS A.C., OWEN A.M., "Neural contributions to the motivational control of appetite in humans", *Eur. J. Neurosci.*, vol. 20, no. 5, pp. 1411-1418, 2004.

[HOR 04] HORNAK J., O'DOHERTY J., BRAMHAM J., ROLLS E.T., MORRIS R.G., BULLOCK P.R. *et al.*, "Reward-related reversal learning after surgical excisions in orbito-frontal or dorsolateral prefrontal cortex in humans", *J. Cogn. Neurosci.*, vol. 16, no. 3, pp. 463-478, 2004.

[JAM 90] JAMES W., *The Principles of Psychology*, Holt, New York, USA, 1980.

[KAR 02] KARAMA S., LECOURS A.R., LEROUX J.M., BOURGOUIN P., BEAUDOIN G., JOUBERT S. *et al.*, "Areas of brain arousal in males and females during viewing of erotic film excerpts", *Human Brain Mapping*, vol. 16, no. 1, pp. 1-13, 2002.

[KNO 06] KNOCH D., PASCUAL-LEONE A., MEYER K., TREYER V., FEHR E., "Diminishing reciprocal fairness by disrupting the right prefrontal cortex", *Science*, vol. 314, no. 5800, pp. 829-832, 2006.

[KOS 05] KOSFELD M., HEINRICHS M., ZAK P.J., FISCHBACHER U., FEHR E., "Oxytocin increases trust in humans", *Nature*, vol. 435, no. 7042, pp. 673-676, 2005.

[KOS 92] KOSSLYN S.M., SHIN L.M., "The status of cognitive neuroscience", *Curr. Opin. Neurobiol.*, vol. 2, pp. 146-149, 1992.

[KOS 95] KOSSLYN S.M., BEHRMANN M., JEANNEROD M., "The cognitive neuroscience of mental imagery", *Neuropsychologia*, vol. 33, no. 11, pp. 1335-1344, 1995.

[LAB 98] LA BAR K.S., GATENBY J.C., GORE J.C., LEDOUX J.E., PHELPS E.A., "Human amygdala arousal during conditioned fear acquisition and extinction: A mixed-trial fMRI study", *Neuron*, vol. 20, no. 5, pp. 937-945, 1998.

[LAB 01] LA BAR K.S., GITELMAN D.R., PARRISH T.B., KIM Y.H., NOBRE A.C., MESULAM M.M., "Hunger selectively modulates corticolimbic arousal to food stimuli in humans", *Behav. Neurosci.*, vol. 115, no. 2, pp. 493-500, 2001.

[LAB 03] LA BAR K.S., LEDOUX J.E., "Emotional learning circuits in animals and humans", in: DAVIDSON R.J., SCHERER K.R. and GOLDSMITH H.H. (eds), *Handbook of Affective Sciences*, pp. 52-65, Oxford University Press, Oxford, 2003.

[LAB 06] LA BAR K.S., CABEZA R., "Cognitive neuroscience of emotional memory", *Nat. Rev. Neurosci.*, vol. 7, no. 1, pp. 54-64, 2006.

[LAC 99] LACHAUX J.P., RODRIGUEZ E., MARTINERIE J., VARELA F.J., "Measuring phase synchrony in brain signals", *Human Brain Mapping*, vol. 8, no. 4, pp. 194-208, 1999.

[LED 00] LEDOUX J.E., "Emotion circuits in the brain", *Ann Rev Neurosci*, vol. 23, no. 1, pp. 155-184, 2000.

[LED 87] LEDOUX J.E., "Emotion", in: PLUM F. (ed.), *Handbook of Physiology*: *The Nervous System*, *Higher Functions of the Brain*, vol. 5, pp. 419-460, Simon & Schuster, New York, USA, 1987.

[LED 96] LEDOUX J.E., *The Emotional Brain*, Simon & Schuster, New York, USA, 1996.

[LEV 03] LEVESQUE J., EUGENE F., JOANETTE Y., PAQUETTE V., MENSOUR B., BEAUDOIN G., LEROUX J.-M., BOURGOUIN P., BEAUREGARD M., "Neural circuitry underlying voluntary supression of sadness", *Biol. Psychiatry*, vol. 53, pp. 502-510, 2003.

[LUO 07] LUO Q., HOLROYD T., JONES M., HENDLER T., BLAIR J., "Neural dynamics for facial threat processing as revealed by gamma band synchronization using MEG", *Neuro Image*, vol. 34, no. 2, pp. 839-847, 2007.

[MAC 49] MACLEAN P.D., "Psychosmatic disease and the visceral brain: recent developments bearing on the Papez theory of emotion", *Psychosomatic Medicine*, vol. 11, no. 6, pp. 338-353, 1949.

[MAR 01a] MARATOS E.J., RUGG M.D., "Electrophysiological correlates of the retrieval of emotional and non-emotional context", *J. Cogn. Neurosci.*, vol. 13, no. 7, pp. 877-891, 2001.

[MAR 01b] MARATOS E.J., DOLAN R.J., MORRIS J.S., HENSON R.N.A., RUGG M.D., "Neural activity associated with episodic memory for emotional context", *Neuropsychologia*, vol. 39, no. 9, pp. 910-920, 2001.

[MEU 97] MEUNIER M., BACHEVALIER J., MISHKIN M., "Effects of orbital frontal and anterior cingulate lesions on object and spatial memory in rhesus monkeys", *Neuropsychologia*, vol. 35, no. 7, pp. 999-1015, 1997.

[MIN 02] MINEKA S., ÖHMAN A., "Phobias and preparedness: the selective, automatic and encapsulated nature of fear", *Biology of Psychiatry,* vol. 52, pp. 927-937, 2002.

[MOG 99] MOGG K., BRADLEY B.P., "Orienting of attention to threatening facial expressions presented under conditions of restricted awareness", *Cognition & Emotion*, vol. 13, no. 6, pp. 713-740, 1999.

[MOR 96] MORRIS J.S., FRITH C.D., PERRET D.I., ROWLAND D., YOUNG A.W., CALDER A.J., DOLAN R.J., "A differential neural response in the human amygdala to fearful and happy facial expressions", *Nature*, vol. 383, pp. 812-815, 1996.

[MOR 98] MORRIS J.S., ÖHMAN A., DOLAN R.J., "Conscious and unconscious emotional learning in the human amygdale", *Nature (London)*, vol. 393, no. 6684, pp. 467-470, 1998.

[MOR 99] MORRIS J.S., ÖHMAN A., DOLAN R.J., "A subcortical pathway to the right amygdala mediating "unseen" fear", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 96, no. 4, pp. 1680-1685, 1999.

[MOR 01] MORRIS J.S., DEGELDER B., WEISKRANTZ L., DOLAN R.J., "Differential extrageniculostriate and amygdala responses to presentation of emotional faces in a cortically blind field", *Brain*, vol. 124, pp. 1241-1252, 2001.

[MUR 07] MURRAY E.A., "The amygdala, reward and emotion", *Trends Cogn. Sci.*, vol. 11, no.11, pp. 489-497, 2007.

[NAD 00] NADER K., SCHAFE G.E., LEDOUX J.E., "Fear memories require protein synthesis in the amygdala for reconsolidation after retrieval", *Nature*, vol. 406, pp. 722-725, 2000.

[NOR 86] NORMAN D.A., SHALLICE T., "Attention to action: willed and automatic control of behavior", in: DAVIDSON R.J. *et al*. (eds), *Consciousness and Self-Regulation*, pp. 1-18, Plenum Press, New York, USA, 1986.

[OCH 99] OCHSNER K.N., KOSSLYN S.M., "The cognitive neuroscience approach", in: BLY B.M. and RUMELHART D. (eds), *Cognitive Science: Handbook of Perception and Cognition* (2nd edition), Academic Press, San Diego, USA, 1999.

[OCH 02] OCHSNER K.N., BUNGE S.A., GROSS J.J., GABRIELI J.D., "Rethinking feelings: An FMRI study of the cognitive regulation of emotion", *J. Cogn. Neurosci.*, vol. 14, no. 8, pp. 1215-1229, 2002.

[OCH 04] OCHSNER K.N., RAY R.D., COOPER J.C., ROBERTSON E.R., CHOPRA S., GABRIELI J.D.E., "For better or for worse: Neural systems supporting the cognitive down-and up-regulation of negative emotion", *Neuro Image*, vol. 23, no. 2, pp. 483-499, 2004.

[ODO 01] O'DOHERTY J., ROLLS E.T., FRANCIS S., BOWTELL R., MCGLONE F., "Representation of pleasant and aversive taste in the human brain", *J. Neurophysiol.*, vol. 85, no. 3, pp. 1315-1321, 2001.

[ÖHM 01] ÖHMAN A., FLYKT A., ESTEVES F., "Emotion drives attention: Detecting the snake in the grass", *J. Exp. Psychol.: General*, vol. 130, no. 3, pp. 466-478, 2001.

[ÖHM 03] ÖHMAN A., MINEKA S., "The malicious serpent: Snakes as a prototypical stimulus for an evolved module of fear", *Current Directions in Psychological Science*, vol. 12, no. 1, pp. 5-9, 2003.

[ÖHM 04] ÖHMAN A., WIENS S., "The concept of an evolved fear module and cognitive theories of anxiety", *Feelings and Emotions: The Amsterdam Symposium*, 2004.

[ÖHM 05] ÖHMAN A., "The role of the amygdala in human fear: automatic detection of threat", *Psychoneuroendocrinology*, vol. 30, no. 10, pp. 953-958, 2005.

[PAN 98] PANKSEPP J., *Affective Neuroscience: The Foundations of Human and Animal Emotions*, Oxford University Press, Oxford, 1998.

[PAP 37] PAPEZ J.W., "A proposed mechanism of emotion", *Arch. Neurol. Psych.*, vol. 38, pp. 725-743, 1937.

[PAV 27] PAVLOV I.P., *Conditioned Reflexes*, Dover, New York, USA, 1927.

[PES 02] PESSOA L., MCKENNA M., GUTIERREZ E., UNGERLEIDER L.G., "Neural processing of emotional faces requires attention", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 17, pp. 11458-11463, 2002.

[PES 08] PESSOA L., "On the relationship between emotion and cognition", *Nat. Rev. Neurosci.*, vol. 9, no. 2, pp. 148-158, 2008.

[PHE 00] PHELPS E.A., O'CONNOR K.J., CUNNINGHAM W.A., FUNAYAMA E.S., GATENBY J.C., GORE J.C. *et al.*, "Performance on indirect measures of race evaluation predicts amygdala arousal", *J. Cogn. Neurosci.*, vol. 12, no. 5, pp. 729-738, 2000.

[PHE 06] PHELPS E.A., LING S., CARRASCO M., "Emotion facilitates perception and potentiates the perceptual benefits of attention", *Psychological Science: a Journal of the American Psychological Society*, vol.17, no. 4, pp. 292-299, 2006.

[PHI 98] PHILLIPS M.L., YOUNG A.W., SCOTT S.K., CALDER A.J., ANDREW C., GIAMPIETRO V. *et al.*, "Neural responses to facial and vocal expressions of fear and disgust", *Proceedings. Biological Sciences/the Royal Society*, vol. 265, no. 1408, pp. 1809-1817, 1998.

[PIT 97] PITKANEN A., SAVANDER V., LEDOUX J.E., "Organization of intra-amygdaloid circuitries in the rat: An emerging framework for understanding function of the amygdale", *Trends Neurosci.*, vol. 20, no. 11, pp. 517-523, 1997.

[POR 81] PORRINO L.J., CRANE A.M., GOLDMAN-RAKIC P.S., "Direct and indirect pathways from the amygdala to the frontal lobe in rhesus monkeys", *Journal of Computational Neuroscience*, vol. 198, no. 1, pp. 121-136, 1981.

[POU 04] POURTOIS G., GRANDJEAN D., SANDER D., VUILLEUMIER P., "Electrophysiological correlates of rapid spatial orienting towards fearful faces", *Cereb. Cortex*, vol. 14, no. 6, pp. 619-633, 2004.

[PRI 07] PRICE J.L., "Definition of the orbital cortex in relation to specific connections with limbic and visceral structures and other cortical regions", *Annals of the New York Academy of Sciences*, vol. 1121, pp. 54-71, 2007.

[RAP 00] RAPCSAK S.Z., GALPER S.R., COMER J.F., REMINGER S.L., NIELSEN L., KASZNIAK A.W. *et al.*, "Fear recognition deficits after focal brain damage: A cautionary note", *Neurology*, vol. 54, no. 3, pp. 575-81, 2000.

[REM 07] REMPEL-CLOWER N.L., "Role of orbitofrontal cortex connections in emotion", *Annals of the New York Academy of Sciences*, vol. 1121, pp. 72-86, 2007.

[ROL 94] ROLLS E.T., HORNAK J., WADE D., MCGRATH J., "Emotion-related learning in patients with social and emotional changes associated with frontal lobe damage", *J. Neurol. Neurosur. Psych.*, vol. 57, no. 12, pp. 1518-1524, 1994.

[ROL 04] ROLLS E.T., "Convergence of sensory systems in the orbitofrontal cortex in primates and brain design for emotion", *The Anatomical Record. Part A, Discoveries in Molecular, Cellular, and Evolutionary Biology*, vol. 281, no. 1, pp. 1212-1225, 2004.

[RUL 02] RULE R.R., SHIMAMURA A.P., KNIGHT R.T., "Orbitofrontal cortex and dynamic filtering of emotional stimuli", *Cognitive, Affective and Behavioral Neuroscience*, vol. 2, no. 3, pp. 264-70, 2002.

[RUS 99] RUSSELL J.A., BARRETT L.F., "Core affect, prototypical emotional episodes, and other things called emotion: dissecting the elephant", *Journal of Personality and Social Psychology*, vol. 76, no. 5, pp. 805-819, 1999.

[SAN 03] SANDER D., GRAFMAN J., ZALLA T., "The human amygdala: An evolved system for relevance detection", *Rev. Neurosci.*, vol. 14, no. 4, pp. 303-316, 2003.

[SAN 05a] SANDER D., GRANDJEAN D., POURTOIS G., SCHWARTZ S., SEGHIER M.L., SCHERER K.R., VUILLEUMIER P., "Emotion and attention interactions in social cognition: Brain regions involved in processing anger prosody", *Neuro Image*, vol. 28, pp. 848-858, 2005.

[SAN 05b] SANDER D., GRANDJEAN D., SCHERER K.R., "A systems approach to appraisal mechanisms in emotion", *Neural Networks: the Official Journal of the International Neural Network Society*, vol. 18, no. 4, pp. 317-352, 2005.

[SAV 91] SAVER J.L., DAMASIO A.R., "Preserved access and processing of social knowledge in a patient with acquired sociopathy due to ventromedial frontal damage", *Neuropsychologia*, vol. 29, no. 12, pp. 1241-1249, 1991.

[SCH 00] SCHOENBAUM G., CHIBA A.A., GALLAGHER M., "Changes in functional connectivity in orbitofrontal cortex and basolateral amygdala during learning and reversal training", *J. Neurosci.*, vol. 20, no. 13, pp. 5179-5189, 2000.

[SCH 01] SCHERER K.R., "Appraisal considered as a process of multilevel sequential checking", *Appraisal Processes in Emotion: Theory, Methods, Research*, pp. 92-120, 2001.

[SCH 02] SCHIENLE A., STARK R., WALTER B., BLECKER C., OTT U., KIRSCH P., SAMMER G., VAITL D., "The insula is not specifically involved in disgust processing: an fMRI study", *Neuro Report*, vol. 13, no. 16,pp. 2023-2026, 2002.

[SCH 03] SCHOENBAUM G., SETLOW B., SADDORIS M.P., GALLAGHER M., "Encoding predicted outcome and acquired value in orbitofrontal cortex during cue sampling depends upon input from basolateral amygdale", *Neuron*, vol. 39, no.5, pp. 855-867, 2003.

[SCH 06] SCHOENBAUM G., SETLOW B., SADDORIS M.P., GALLAGHER M., "Encoding changes in orbitofrontal cortex in reversal-impaired aged rats", *J. Neurophysiol.*, vol. 95, no. 3, pp. 1509-1517, 2003.

[SCH 62] SCHACHTER S., SINGER J.E., "Cognitive, social, and physiological determinants of emotional state", *Psychological Review*, vol. 69, pp. 379-399, 1962.

[SCO 97] SCOTT S.K., YOUNG A.W., CALDER A.J., HELLAWELL D.J., AGGLETON J.P., JOHNSON M., "Impaired auditory recognition of fear and anger following bilateral amygdala lesions", *Nature*, vol. 385, no. 6613, pp. 254-257, 1997.

[SER 08] SERGERIE K., CHOCHOL C., ARMONY J.L., "The role of the amygdala in emotional processing: A quantitative meta-analysis of functional neuroimaging studies", *Neurosci. Biobehav. Rev.*, vol. 32, no. 4, pp. 811-830, 2008.

[SHA 04] SHAROT T., DELGADO M.R., PHELPS E.A., "How emotion enhances the feeling of remembering", *Nat. Neurosci.*, vol. 7, pp. 1376-1380, 2004.

[SMA 99] SMALL D.M., ZALD D.H., JONES-GOTMAN M., ZATORRE R.J., PARDO J.V., FREY S. *et al.*, "Human cortical gustatory areas: A review of functional neuroimaging data", *Neuro Report*, vol. 10, no. 1, pp. 7-14, 1999.

[SMA 03] SMALL D.M., GREGORY M.D., MAK Y.E., GITELMAN D., MESULAM M.M., PARRISH T., "Dissociation of neural representation of intensity and affective valuation in human gestation", *Neuron*, vol. 39, no. 4, pp. 701-711, 2003.

[SMI 04a] SMITH A.P.R., DOLAN R.J., RUGG M.D., "Event-related potential correlates of the retrieval of emotional and nonemotional context", *J. Cogn. Neurosci.*, vol. 16, no. 5, pp. 760-775, 2004.

[SMI 04b] SMITH A.P.R., HENSON R.N.A., DOLAN R.J., RUGG M.D., "FMRI correlates of the episodic retrieval of emotional contexts", *Neuro Image*, vol. 22, no. 2, pp. 868-878, 2004.

[SMI 90] SMITH C.A., LAZARUS R.S., "Emotion and adaptation", in: PERVIN L.A. (ed.), *Handbook of Personality: Theory and Research*, pp. 609-637, Guilford Press, New York, USA, 1990.

[SOM 06] SOMERVILLE LEAH H., WIG GAGAN S., WHALEN PAUL J., KELLEY WILLIAM M., "Dissociable medial temporal lobe contributions to social memory", *J. Cogn. Neurosci.*, vol. 18, pp. 1253-1265, 2006.

[SPR 98] SPRENGELMEYER R., RAUSCH M., EYSEL U.T., PRZUNTEK H., "Neural structures associated with recognition of facial expressions of core emotions", *Proceedings Biological Sciences/the Royal Society*, vol. 265, no. 1409, pp. 1927-1931, 1998.

[STR 99] STREIT M., IOANNIDES A.A., LIU L., WÖLWER W., DAMMERS J., GROSS J., GAEBEL W., MÜLLER-GÄRTNER H.W., "Neurophysiological correlates of the recognition of facial expressions of emotion as revealed by magneto-encephalography", *Cognitive Brain Research*, vol. 7, no. 4, pp. 481-491, 1999.

[SUG 05] SUGASE-MIYAMOTO Y., RICHMOND B.J., "Neuronal signals in the monkey basolateral amygdala during reward schedules", *J. Neurosci.*, vol. 25, pp. 11071-11083, 2005.

[SWA 98] SWANSON L.W., PETROVICH G.D., "What is the amygdala?", *Trends Neurosci.*, vol. 21, no. 8, pp. 323-331, 1998.

[TOD 08] TODOROV A., ENGELL A.D., "The role of the amygdala in implicit evaluation of emotionally neutral faces", *Social Cognitive and Affective Neuroscience*, vol. 3, no. 4, pp. 303-312, 2008.

[TRE 99] TREMBLAY L., SCHULTZ W., "Relative reward preference in primate orbitofrontal cortex", *Nature*, vol. 398, no. 6729, pp. 704-708, 1999.

[VOG 92] VOGT B.A., FINCH D.M., OLSON C.R., "Functional heterogeneity in cingulate cortex: The anterior executive and posterior evaluative regions", *Cerebral Cortex*, vol. 2, no. 6, pp. 435-443, New York, 1992.

[VUI 05] VUILLEUMIER P., "How brains beware: Neural mechanisms of emotional attention", *Trends Cogn. Sci.*, vol. 9, no. 12, pp. 585-594, 2005.

[WEI 56] WEISKRANTZ L., "Behavioral changes associated with ablation of the amygdaloid complex in monkeys", *Journal of Comparative and Physiological Psychology*, vol. 49, no. 4, pp. 381-391, 1956.

[WHA 01] WHALEN P.J., SHIN L.M., MCINERNEY S.C., FISCHER H., WRIGHT C.I., RAUCH S.L., "A functional MRI study of human amygdala responses to facial expressions of fear versus anger", *Emotion*, vol. 1, no. 1, pp. 70-83, 2001.

[WHA 98] WHALEN P.J., BUSH G., MCNALLY R.J., WILHELM S., MCINERNEY S.C., JENIKE M.A. *et al.*, "The emotional counting stroop paradigm: A functional magnetic resonance imaging probe of the anterior cingulate affective division", *Biological Psychiatry*, vol. 44, no. 12, pp. 1219-1228, 1998.

[WIN 03] WINSTON J.S., VUILLEUMIER P., DOLAN R.J., "Effects of low-spatial frequency components of fearful faces on fusiform cortex activity", *Curr. Biol.*, vol. 13, no. 20, pp. 1824-1829, 2003.

[WIN 05] WINSTON J.S., GOTTFRIED J.A., KILNER J.M., DOLAN R.J., "Integrated neural representations of odor intensity and affective valence in human amygdale", *Journal of Neuroscience: the Official Journal of the Society for Neuroscience*, vol. 25, no. 39, pp. 8903-8907, 2005.

[WRI 02] WRIGHT C.I., MARTIS B., SHIN L.M., FISCHER H., RAUCH S.L., "Enhanced amygdala responses to emotional versus neutral schemetic facial expressions", *Neuro. Report*, vol. 13, no. 6, pp. 785-790, 2002.

[YAN 02] YANG T.T., MENON V., ELIEZ S., BLASEY C., WHITE C.D., REID A.J. *et al.*, "Amygdalar arousal associated with positive and negative facial expressions", *Neuro Report*, vol. 13, no. 14, pp. 1737-1741, 2002.

[YON 05] YONELINAS A.P., OTTEN L.J., SHAW K.N., RUGG M.D., "Separating the brain regions involved in recollection and familiarity in recognition memory", *Journal of Neuroscience: the Official Journal of the Society for Neuroscience*, vol. 25, no. 11, pp. 3002-3008, 2005.

PART 2

Non-verbal Behavior

Chapter 3

Emotional Corpora: from Acquisition to Modeling

3.1. Introduction

The purpose of affective systems is to incorporate emotions into computational systems. Central to this is the creation of emotional corpora. They are essential both for creating human–machine interaction systems that can manage and regulate emotional responses and for carrying out perceptive studies into emotion. Corpora are a body of consistent and significant recorded data that we can use to create and design models. The composition and interpretation of corpora obviously depends on the objective for modeling. For example, generating emotional expressions for a conversational agent could require an extensive corpus of emotional behavior from a single locutor chosen for their emotional expressiveness. In contrast, an application designed to detect emotion needs to be highly sensitive to different interlocutors and therefore requires a large multi-speaker corpus.

Emotions are a complex dynamic process [COP 09] and manifest themselves differently from one person to another. The brain is the source of multiple emotions that evolve dynamically mixing pleasant and unpleasant, conscious and unconscious emotions that are stimulated by external or internal events. An emotion is a state of consciousness that can be pleasant or upsetting, which is accompanied by a bodily reaction when faced with an external or internal event (e.g. noise, light, memory or concept). For example, anger makes blood rush to the hands in preparation for self-defense while fear makes blood rush to the muscles to control the bodily movement,

---

Chapter written by Laurence DEVILLERS and Jean-Claude MARTIN.

Emotion-Oriented Systems Edited by Catherine Pelachaud

such as preparing the muscles of the leg to flee. The term emotion should be interpreted in its broadest sense to include emotions, attitudes, moods and other emotional dispositions. Scherer [SCH 81], for example, has described emotions as "the body's interfaces with the outside world". The most important aspects of the emotional process are:

– the evaluation of a situation in relation to an individual's needs and aims;

– physiological and psychological preparation for context specific action;

– communicating the individual's state and intentions to his/her social environment.

Scherer consequently defines an emotion as "episodes of massive, synchronized recruitment of mental and somatic resources allowing [*sic*] to adapt or cope with a stimulus event subjectively appraised as being highly pertinent to the needs, goals and values of the individuals" [SCH 04]. This synchronization, which comes in different forms, justifies the creation of multimodal corpora. Perception studies, for example, evaluate the effect produced, whether this is a mixture of background emotions or emotions that are more obvious. The emotion generated by an individual can be perceived differently by the receiver and depends on his/her personal sensibilities. We therefore only recognize a fraction of an individual's emotional states on the basis of voluntarily controlled or totally uncontrolled signals that the individual conveys and we then interpret. Even if a large majority of emotions felt during an interaction are unnoticeable, they can often be identified through the verbal content of an utterance and/or paralinguistic and bodily signals.

Building a corpus is therefore fundamental for studying these complex processes both for conceptual analysis and for modeling "affective systems" that may occur for a number of reasons. This may include the desire to recognize users' emotional states, to express appropriate emotional states in a virtual agent or even for predicting how a user will respond to the machine's reaction in a certain state. It therefore seems important to carry out studies on this collected emotional data in as natural a context as possible if we are to understand these complex emotional processes and model them. By natural contexts, we mean situations where Scherer's three stages are present: the evaluation of the situation by the individual, physiological and psychological preparation to context appropriate action and, finally, communication of the individual's state and intentions to his/her social environment. However, the majority of current research into building affective systems is based on a sub-set of primary emotions [EKM 75], using a corpus of prototypical emotional data from actors that often consider emotions as a static process. There are fundamental differences between the emotional states observed in artificial data, where there is no real emotional feeling and no induced action or interaction, and affective behavior observed in spontaneous data. If the machines of tomorrow need to be sensitive to subtle emotions in natural interactions, it is important to model emotional behavior in as authentic a way as possible.

This chapter aims to present the different acquisition strategies of mono- and multimodal emotional corpora and to describe the challenges of interpreting these corpora. Section 3.2 will focus on the different acquisition strategies for corpora and will present a summary of the different emotional corpora in existence. Following this, we will examine emotional coding schemes, emotional events and current challenges along with a typology of the complex emotions found in spontaneous data. Building a corpus is of paramount importance for constructing models based on stochastic approaches. The chapter will conclude by illustrating some examples of different applications for corpora in addition to the major scientific advantages and constraints in this area.

3.2. Building corpora: "acted", "induced" and *real-life* emotions

The first issue when studying emotions is to find or create a corpus of data. A corpus is a body of data collected for a specific scientific question. A corpus is generally a collection of annotated data. There are three types of corpora used in research into emotions:

– out-of-context corpora (with real/non actors) and corpora from fictional contexts (e.g. film, theater);

– corpora from artificial interfaces that can display emotions (e.g. *The Wizard of Oz*: interaction with a system simulated by a human);

– *real-life* corpora: human interaction, interaction with human–machine dialog systems.

The majority of experiments on emotion have been carried out on "acted" data recorded by actors, often with few different types of emotions among the "primary emotions" defined by Ekman [EKM 99]. Following Juslin and Laukka's study [JUS 03] reported in [BAN 04] on 104 experimental studies on emotions, 7% of studies used emotion induction, 12% used recordings from real situations and 87% used expressions produced by study subjects. The results in the scientific community therefore relate to acted data, as reflected in the "Corpora for Research on Emotion and Affect" series of workshops [DEV 06a, DEV 08a]. We can generate a table demonstrating the authenticity of these corpora (Table 3.1) according to their experimental controls and context (recording, experimental manipulation, type of interaction).

Type of corpus	Natural/ artificial context	Protocol	Recording	HH interaction, HHM, via telephone
Acted (amateur or professional)	Laboratory	Non-linguistic phrases Acted emotions	Visible	No interaction
Fiction	Theater, cinema	Given context	Visible	HH Controlled
Induction Wizard of Oz HH	Laboratory	Little context, some emotions impossible	+/- visible	HM controlled by H
Induction HM	Laboratory	Example: game	+/- visible	HH controlled by H
HMC	Laboratory or real users	Context: task	Slightly visible	HM without control
HHC	Natural	-Meeting -TV/radio/Web, -Call centers	- Visible - Visible - Slightly visible	HH without control

Table 3.1. *Different types of corpora (adapted from [DEV 06b]): H –human, M–machine, HMC –human–machine communication, HHC –human-to-human communication*

3.2.1. *Acted data*

The majority of previous and current research has been carried out on "acted" data. The context and sometimes the linguistic context have been ignored to focus solely on body language and paralinguistic content. This data therefore have several advantages: they raise fewer ethical problems, are easier to collect and we can gather a large amount of data for each class of emotion (although this is not necessarily always the case in practice). Furthermore, it means that two segments with identical linguistic content can be compared that can attribute differences in perception to single acoustic or multimodal indicators. These data do not really reflect reality, however, due to their lack of context and reduced number of actors (often less than 10), meaning that these corpora have less variability than those composed of spontaneous data. This is because actors often use stereotypes that are characteristic of an emotion and is therefore often very different to a real expression of this emotion. There is also a tendency to exaggerate social codes of communication (pull effect), but without any push effect normally associated with the physiological side of emotion, such as a fast beating heart for instance. Furthermore, these corpora do not always evaluate the quality of emotions expressed by actors. Note, however, that in the case of GEMEP (GENeva Multimodal Emotion Portrayals) data for example

[BAN 07], different scenarios expressing emotions were given to actors who were directed to express emotions less prototypically. This was combined with a perception test to validate any annotations.

One solution for obtaining authentic emotional data in sufficient quantity and without confidentiality constraints is to use fiction by, for example, selecting data from professional actors. Placing an actor in a particular situation could increase the realism of the acted emotions, since some emotions such as fear are difficult to observe in real situations. In their work on fear, Clavel *et al*. [CLA 07, CLA 08] selected film sequences and compared this against one credibility rule: good actors make us believe that their emotions are real. These simulated emotional episodes can, without doubt, allow us to construct good emotional prototypes. The disadvantages of using this type of data are that it is often accompanied by sound effects and rarely reflects real behavior.

3.2.2. *Induced data*

Emotion induction techniques allow us to obtain more controlled "prototypical" behavior, since data recorded in real-life contexts may be rich in emotion but may be uncontrolled. Several techniques have been introduced for inducing emotions:

– hypnosis [GRO 68];

– the *recall technique*;

– showing films [GRO 95], images or games that induce an emotional response [CON 02, DEV 08b, ZAR 07];

– completing a difficult task in a limited amount of time to induce stress;

– *Wizard of Oz* studies [BAT 03];

– the SAL (sensitive artificial listener) technique [DOU 03];

– the *driving simulator technique*, which allows us to study emotions during actions; or even

– the eWiz interface [AUD 04].

However, the emotions induced by these techniques are often weak in intensity. Furthermore, these induction methods do not necessarily induce close emotional states: for example, when tired we can be equally stressed or amused.

3.2.3. *Real-life data*

Different kinds of natural data have been recorded, ranging from data from airline pilots, therapy sessions, call centers, reality television and news reports. The quality of the recording is often poor, the quantity of emotional data low and there is generally little speaking time per interlocutor. Furthermore, the emotion expressed by the locutor is not always obvious. The recording material can also become an obstacle to the natural aspect of data. A criticism of real data is that because the locutor is aware of being recorded, he or she will display and react so as to have an effect on the audience [HES 06], which can compromise the data's validity. The use of call center data provides an interesting alternative, particularly when we are interested in the expression of emotions via the voice only. Inconspicuous recording means that we can acquire spontaneous data. Furthermore, with telephone data emotion is expressed via the voice without interference from actions, gestures or facial expressions. In the case of hotlines, where the motive for calling is often for information or to solve a problem quickly, the interlocutors are not focused on how they will be judged and therefore act spontaneously.

This is not, however, the case with automatic voice systems. For example, in the United States callers often deliberately exaggerate their speech in order to be directed to a human operator more quickly. The content for the most part is emotionally weak, however, often with less than 10% emotional content [DEV 05b]. Depending on the context of the interaction, we can find greater emotional content of up to 30% in a medical corpus [VID 06]. With regard to concerns about the annotation of this type of data (because the locutor cannot be asked about their emotional state), strategies have been developed to account for this that will be further detailed in section 3.4, which focuses on annotation coding. Another criticism is that it is difficult to have enough data produced by interlocutors and even less data that allow you to compare several individuals. This is due to large variations between individuals in how emotions are expressed in terms of both linguistic and acoustic content.

It is possible to collect data of similar reactions, even if these are not identical. The development of human corpora, rich in authentic emotions representing a range of human behaviors that can be annotated at different levels (prosodic, contextual, etc.) is the key to developing statistical models that can detect various emotions. Some recent projects have begun to collect and annotate natural data from sources such as:

– Reading-Leeds Emotional speech corpus (four-and-a-half hours of televised interviews) [STI 00];

– Belfast Naturalistic Emotional Database (298 clips televised in English from 125 speakers, 31 male, 94 female, ranging from 10 to 60 seconds long) [DOU 03a];

– CREST-ESP Project on Expressive Speech Processing at ATR (Advanced Telecommunication Research) [CAM 00] (1,000 hours over five years, of which little is of sufficient quality to be used) where mini recorders were used by subjects to record daily conversations in Japanese, Chinese and English, often at the cost of sound quality) [DOU 03];

– audio from call centers:

- data from a medical emergency call center (20 hours) (CEMO),

- Stock Exchange Customer Service center (BOURSE: four-and-a-half hours of recordings) [DEV 05b];

– audiovisual data from televised clips (Emo TV) [DEV 05a].

Natural data raise ethical problems, however, such as consent and data confidentiality. While conducting physiological tests on fear and anger, Ax [AX 53] created that illusion that there was a serious problem with the equipment using sinister noises and sparks and asking the researchers to pretend to panic. These kinds of experiments that are painful or unpleasant for the participants are now no longer repeated for ethical reasons. The Human Excellence Network is particularly concerned with ethics-related issues, both in terms of means of gathering and processing data as well as in how these data will be used. As such, it is recommended that participants have passed a perception test based on expressions of negative emotions and that they are shown positive videos so that they can leave in an emotional state close to the one in which they arrived before the test.

3.2.4. *Comparison of different types of data*

Lazarus, having studied different types of induced and acted data, decided to focus purely on natural data [LAZ 98]. Similarly, Scherer has concluded that we cannot mix the results from acted data with that of natural data [SCH 81]. Batliner, for example, has compared the experiments carried out on acted, induced (*Wizard of Oz*) and real (human–machine interaction) data and has shown that the best models for acted data did not apply to real data: the good detection scores were inversely proportionate when applied to natural data [BAT 03]. Similar results have been found by [VID 08, VOG 05], with again better performance noted in acted rather than natural data. Similarly, the models based on acted data did not perform well when applied to real data [BAT 03] and vice versa [VID 08].This shows that that the same indicators are not relevant for all types of data. Despite this view that results from artificial data cannot be applied to real data, however, there are still very few studies on natural corpora of any significant size carried out on real data. This awareness is just beginning to emerge, as shown in the "Corpora for Research on

Emotion and Affect" workshop at the Language Resources and Evaluation Conferences of 2006 [DEV 06a] and 2008 [DEV 08a].

3.3. Current emotional corpora

The HUMAINE Association (http://emotion-research.net/), created following the HUMAINE Network of Excellence, keeps registers of the main, natural and unnatural databases used by different members. This is part of an effort to collect natural and induced data and annotation protocols.

All the corpora described in Table 3.2 are human-to-human interaction corpora. They have been annotated with different codes for emotion and context and can play two extremely different roles. A corpus can be "prototypical" or "provocative", where "provocative" refers to a collection of extracts from corpora than to an actual corpus in itself.

The complex emotions found in natural data corpora and the annotation codes for emotions and context are described in the following sections.

Multimodal signals have also been annotated for a small section of the audio-visual corpora.

In order to be able to compare the results of the studies and to account for differences between locutors, it is important to share corpora. At this current time, there are no real-life corpora that are shared across the scientific community, nor any benchmarks for evaluating different systems. In the HUMAINE Association, the CEICES (Combining Efforts for Improving Automatic Classification of Emotional User States) experiment on the AIBO corpus [BAT 06] is currently the only collaborative test on one corpus (seven sites have participated in this experiment).

3.4. Coding schemes

The coding schemes developed at LIMSI (the Computer Sciences Laboratory for Mechanics and Engineering Sciences), and subsequently by the HUMAINE Network of Excellence, include coding of emotions, emotional events and their specific contexts. One of the most important challenges when studying emotional behavior on the basis of natural data is the choice and reliability of the representation and annotation of emotions. The HUMAINE Network has collected information on terminology, the choice of emotion representation, lists of markup languages and has continued to contribute to the development of annotation schemes and language for describing emotions [SCH 07]. Developing a standard language to describe and annotate emotions for computer applications is therefore one of the current areas of research that the W3C working group has focused on and is now part of the W3C working group on Multimodal Interaction (http://www.w3.org/TR/emotionml//).

Corpus	Type of interaction	Audio video	Length	No. of locutors	No. of emotions	Application
Belfast QUB	Interview	Audio Visual (AV) TV	4.5 hours	125	>20	Analysis
SAL QUB	Wizard of Oz	AV	10 hours	20	>20	Detection
CREST ATR	Japanese–English–Chinese dialog	Audio (A)	1,000 hours	3-5	Three levels: state, style, and vocal quality	Synthesis
Geneva Airport Lost Luggage	Interview	AV		112	Anger, cheerfulness, stress, indifference, sadness	Analysis
Sympafly Erlangen	Interaction with the German Vermobil information service	A	110 dialogs	110	Stress/ no stress	Detection
E-WIZ ICP	French WOZ dialog Language learning	AV Physiological	5 hours	17	Labels selected by patients	Analysis/ synthesis
Bourse LIMSI	French dialog Stock Exchange call center	A	4. 5 hours	100 clients 4 agents	Anger, fear, satisfaction, apology, neutral	Detection
CEMO LIMSI	French dialog Medical emergency call center	A	20 hours 688 dialogs	784 callers 7 agents	20 classes 7 macroclasses	Detection
EmoTV LIMSI	Interview in French, TV	AV	1 hour	100 locutors	20 classes 7 macroclasses	Generation
EmoTABOO LIMSI	Human–human dialog in French, taboo games, laboratories	AV	7 hours	10 locutors	20	Generation
SAFE THALES ENST LIMSI	British/American dialog and monologs, realistic fiction	AV	7 hours Extracts from 30 films	400 locutors	Fear positive/ negative Neutral	Detection

Table 3.2. *Extract from the collection of corpora compiled by the HUMAINE Network of Excellence*

Modeling emotional behavior requires the ability to represent the multiple levels of abstraction and temporality involved in emotional processes: the emotion itself, a context describing emotional events, dialog acts, communicative aims and corresponding multimodal behavior. In recent studies, various emotion representation theories have been used. Annotation schemes can combine several aspects of these theories, such as verbal categories and abstract dimensions [COW 00, CLA 08, CRA 04] and verbal categories and evaluation dimensions [DEV 06b]. Representing emotions using verbal categories may seem fairly subjective but an annotation guide specifying the definitions of each term must appear in order to ensure consistency in annotations.

Representing emotions using descriptors of abstract dimensions is the approach that provides the highest level of genericity and seems to be the most reliable approach. Three dimensions are traditionally used [OSG 75] including arousal, evaluation and control. These three dimensions cannot differentiate all emotions. Other studies have proposed including other dimensions, such as intensity or even "stance", which remains an open area of study. The advantage of mixing several representation levels is that we can make up for any lack of nuance between each of these representations and can control the continuity of annotations.

The emotion representation scheme proposed in [DEV 05b, MAR 06] combines various theories and representations. One of the original aspects of this scheme is to propose that several emotions can be felt at the same time, described as "blended" or mixed emotions. Each annotator can choose a main emotion and several minor emotions to describe emotional elements (an emotional segment corresponds to one turn in dialog but can be shorter or longer than this). The reasoning behind this approach is that there are not negative perceptions of emotions, but different social and psychological sensitivities linked to belief systems that are internalized us all. The principle aim, therefore, remains to preserve the nuances of each of these annotations.

Annotations from several authors (at least two or three for a majority vote) are therefore combined in a soft vector. This type of representation means we can select sub-sets of corpora for different tasks. For example, the highest scoring emotion will be considered as a segment label for programming detection systems [DEV 05b]. This involves using complex systems to generate mixed emotional behavior for conversational agents [MAR 06]. Cognitive evaluation dimensions and contextual information should enable us to specify the context of production for emotional behavior. One of the features of this multi-level coding scheme is the possibility of annotating the corpus as a whole as well as segments of it, such as:

– emotional labels as well as abstract dimensions, with the aim of studying their repetition and how they complement one another;

– defining combinations of non-core emotions by using two labels to annotate a single type of emotional behavior;

– different kinds of complex emotions: hidden, stacked, etc.;

– the emotional context can include the role, theme, etc., of emotions on a broad level and, at a segmental level, emotional events, communicative aims etc.

In summary, the originality of this scheme developed in LIMSI-CNRS in collaboration with the University of Belfast [DEV 06c] lies in the integration of context, using three representation theories and annotating several emotions on the same emotional segment.

3.4.1. *Emotional annotation protocols*

An annotation protocol includes a sequence of three phases that are repeatedly carried out at the start of annotation to refine the different phases. This protocol is applicable to annotations other than emotion:

– defining annotated features and the units to which they relate;

– the annotation itself; and

– validating annotations.

This generic scheme can be then adapted for each application. The protocol can also be used to make annotations more reliable.

3.4.1.1. *Definitions of annotated features*

In order to annotate our data, it is first necessary to define the emotional unit in accordance with the given application. Emotions are, by their very nature, dynamic. Emotional segments represent a stable sequence where there is an emotion (or combination of emotions). An emotional segment can be a word, group of words or speaker turn. Emotional units can be smaller than a turn, although one turn is the unit found in the majority of studies.

Segments are obtained manually or automatically using speech-recognition systems with automatic alignment of words on the utterance. The annotated entities also need to be defined.

Dimensions or lists of emotions should be described and illustrated in the annotation guide. It is important to be able to cross reference the annotations with their labels and abstract dimensions so that they can be corroborated.

Some labels are directly linked to valence and intensity. For example, fright is negative and has a strong intensity. However, this is not the case with all emotions. For example, excitement or surprise can be either positive or negative. The simplest way of constructing a list of emotional labels for a given task is to refer to standard lists of emotions from other experiments conducted with a large population sample (lists provided by HUMAINE, for example) and to select around 20 applicable labels selected by majority vote. It is then necessary to regroup the labels according to seven or eight macro-classes.

During the annotation phase, the choice of labels is further refined, where some may disappear and others may appear if sufficiently frequent. Experiments using free annotation of emotional labels have highlighted the challenge of differing ways of producing the same utterance. More than 170 labels have been proposed [ABR 05], and annotators have subsequently refined this set to around 20 more general classes.

3.4.1.2. *Annotation*

Annotation protocol requires a phase of learning and calibration for the evaluation benchmark. It is also crucial to make several annotations in parallel in order to make them reliable. There should be a minimum of two or three annotators in order to make a majority vote. Batliner has used five to annotate data from automatic dialog systems [BAT 03]. In one experiment, where the number of annotators varied from three to 40, it was noted that the curve of annotations stabilizes at around 10 annotators [ABR 06]. This study also showed different annotation results between men and women. It is therefore advisable to use both female and male annotators to have as diverse an overview as possible. In the scheme proposing the annotation of several emotional labels, these markers are then combined in a soft vector. Table 3.3 provides an example of a soft vector for two annotators with different opinions on the initial annotation.

Annotator 1	(wM) Annoyance, (wm) Interest
Annotator 2	(wM) Stress, (wm) Annoyance
Resulting vector *soft-vector*	(wM/W Annoyance, wm/W Stress, wm/Interest)
Example	For wM=2, wm = 1, and W=6 (0.5 Annoyance, 0.33 Stress, 0.17 Interest)

Table 3.3. *Example of a vector of emotions considered where wM and wm are the weight allocated to emotions [DEV 05b]*

3.4.2. *Annotating context*

Contextual information comes in two temporal forms, either being the length of the clip or dialog or in terms of emotional segments. Information is annotated on an overall scale corresponding to broad classes of information that cover several fields such as:

– semantic context (subject of the call, theme of the interview);

– relationship between participants (victim/aggressor [CLA 07], agent/caller [DEV 05b]);

– audience;

– recording conditions;

– physical context;

– participants (age, sex).

Contextual information about an emotional segment (often corresponding to one speaker turn) consists of:

– communicative aims;

– appraisal variables;

– events;

– timing of events (recent, immediate, past, future);

– degree of involvement;

– impact of recording.

The overall textual information highlights the areas of the corpus for closer study. Relations between participants have been used to specify the most effective models between, for example, agent and caller [DEV 05b] and between victim and aggressor [CLA 08]. Contextual information at the level of a speaker turn has been used for different purposes such as:

– communicative acts for designing the artificial agent [ABR 07];

– the timing of events in studying fear to develop models for different types of fear [CLA 07].

An experiment annotating appraisal variables [DEV 06b] has shown a rate of agreement of around four annotators for five variables:

– 4: inherent pleasure (high–low);

– 8: relationship with expectations (in agreement–disagreement);

– 9: favorable to personal aims (favorable–unfavorable);

– 11: control of the event (high–low);

– 12: control of consequences (high–low).

Variables 4, 8 and 9 are positive or negative judgment and variables 11 and 12 are linked to controlling causes and consequences that are deduced from context (communicative act, event). Conversely, introspective variables such as “the person’s adaptation capacity” and agreement with one’s “face” are also impossible to annotate. Another difficulty is linked to the fact that there may be several emotional events and the whole set of appraisal SEC (Sequential Evaluation Checks) variables should be considered for each event, which further complicates their classification. In natural data, there may be multiple events triggering an emotional state, such previous trauma, recent illness, conversations, etc. This model therefore does not easily apply to complex emotions.

3.4.2.1. *Validation*

The main difficulty with this area of study is finding the right level of representation and making the subjective judgments of the annotators as accurate as possible. Since human annotation is subjective, there needs to be a rigorous annotation protocol. Once the annotation scheme and labels have been decided, we then need to select the rules for segmentation, the number of annotators and set validation procedures. We also need to consider the coherence and consistency of annotations that is achieved using accuracy measures. Several measures exist, such as kappa coefficient [COH 60] for discrete variables or Cronbach’s alpha coefficient for continual variables. The formula for kappa coefficient is:

$$k = \frac{P_{obs} - P_{ch}}{1 - P_{ch}}$$

Where P_{obs} is the level of agreement observed and P_{ch} is the level of agreement obtained by chance. P_{obs} is obtained by calculating the number of markers representing agreement and dividing this by the total number of markers. If the annotators always agree then kappa = 1. Table 3.4 provides an interpretation of the different possible values of kappa.

The measures of agreement between annotators are generally used on segments with a single label. In the case of multiple labels for a single segment, one solution with increased kappa is developed in [DEV 06b].

Value of kappa	Interpretation
<0	No agreement
0.0–0.2	Very little agreement
0.2–0.4	Little agreement
0.4–0.6	Medium agreement
0.6–0.8	Correct agreement
0.8–1.0	Perfect agreement

Table 3.4. *Interpretation of the values of kappa*

A good annotator will use the same annotation strategy on the same data with several months between annotations. One study on emotions in call center data [DEV 05b] shows an 85% level of agreement with a single annotator after two months of annotations. These validation methods can be used for all levels of scheme annotation. Perception tests are also used to validate annotation schemes [VAS 09].

3.5. Complex emotions in spontaneous data

The study of complex emotions has several aims:

– to understand complex emotional behavior so that we can learn how to model it;

– to improve the strength of automatic detection systems by selecting parts of the corpus where emotions are not complex and indicators are more reliable for learning;

– to show the importance of using real-life data to construct emotionally "intelligent" systems;

– to move towards expressive virtual agents that display "realistic" emotional states.

Emotions are phenomena that are essential to survival, learning and social integration. They are also a reflection of the overall physical and moral state of an individual. They can be immediate and epidermal, as with fear, or can be highly controlled, as with contempt. Rules of politeness and social norms have taught us in part to hide our emotions. The more formal a situation, the more applicable these rules are and will lead to the modification of our emotional signals. Social context and individual personality are therefore key to deciphering emotions. Showing emotional states goes against self-control and social customs.

There are various types of emotions;

– those that are displayed (intentional): either real or acted;

– those that are hidden and not obvious to others, but known by the subject themselves;

– those that are unconscious and not perceived by the subject.

Numerous theories support the existence of complex emotions as a mixture of several emotions [EKM 92, PLU 82] or as the result of an appraisal process [SCH 99], taking into account the physiological effects of each individual and the effects of the sociocultural context. The main research into expressing mixed emotions has been carried out by Ekman [EKM 75]. According to his research, the simultaneous appearance of several emotions leads to a specific facial expression that can result from certain combinations of expressions from individual emotions. Emotions are revealed by different areas of the face. Ekman has also shown that, depending on whether the emotions are real (feigned or actually felt), mixed feelings are different and some expressions cannot be hidden voluntary or are at very difficult to hide. According to some studies [BAS 79, OCH 06], negative emotions are mainly expressed via the upper part of the face whereas positive emotions are shown by the lower part of the face. For emotions that are both positive and negative, production is more complex and a disgusted grimace may appear with smiling eyes for an emotion such as embarrassment or irony. Castelfranchi and Poggi have proposed four categories for classifying complex emotions [CAS 98]:

– *Omission*: emotion E is felt but not acted. This information can be deduced from the context or small signals.

– *Concealment*: emotion E is felt but does not trigger the predicted action.

– *Falsification*: emotion E is not felt but is expressed.

– *Masking*: emotion E is hidden by emotion F, emotion E is felt but F is deliberately expressed; *masking* is therefore a combination of *concealment* and *falsification*.

There are currently few empirical studies on the mixture of emotional behavior based on real data corpora. In [DEV 05b], conflicting mixed emotions (positive and negative valence), such as annoyance, compassion or relief and concern, have been noted in real data recorded in a medical emergency call center (20 hours of recordings). These improbable mixtures have been validated by perception tests. Similarly, in a multimodal data corpus (using interviews from news bulletins)

[ABR 05], conflicts of emotion such as relief/sadness with contradictory multimodal indicators (crying with relief) have also been noted. These mixtures of emotions (emotion, mood, attitude) are frequent in real data. These results have been obtained by analyzing annotations in corpora and perception tests.

In these approaches, emotional labels are organized hierarchically so that they can be regrouped into macro-classes (or families of emotion). Generally these families include labels of different intensities and different emotional states (emotion, attitude, humor).

For example, the two macro-classes *fear* and *anger* contain:

– fear: worry, stress, panic, fright, etc.;

– anger: annoyance, cold anger, hot anger, etc.

In [DEV 05b], for example, there are three types of mix:

– ambiguous mixed emotions: two emotions from the same family (macro-class), such as anger and annoyance, showing that the intensity of the emotion felt probably lies somewhere between the two;

– non-conflicting mixed emotions: two emotions from two different families but of the same valence, e.g. fear and anger;

– conflicting mixed emotions: two emotions from two different families of emotion but of differing valence, e.g. compassion and annoyance.

In this research using an oral corpus, it is very difficult to identify masking. The typology proposed by Castelfranchi and Poggi [CAS 98] cannot be easily applied. In the EmoTV audiovisual corpus, only masking can be identified since omission, concealment and falsification are impossible to annotate in a reliable way. In this audiovisual corpus, four types of mix are annotated [DEV 05a]:

– mixed: two emotions are present at the same time;

– masked: one emotion is used to mask another;

– sequential: two emotions are present in succession (with or without overlap);

– conflict between cause/effect: indicates another class of emotion e.g. crying with happiness.

The complex emotions found with positive and negative aspects in the two types of corpora demonstrate a mental conflict on the part of the locutors. It is therefore necessary to look at this feature in further detail and attempt to find the answers to the following questions:

– What is the cause of these complex emotions of conflicting valence?

– Do they occur frequently?

– How can we convey several “causal” events in an appraisal model?

Mixed positive/negative emotions can be the result of one or several events triggering the emotional state. For example, in Scherer’s appraisal model the emotional process is triggered by an event or situation that is called the *focus* of the emotion. In several natural data examples, the cause is different from the emotional trigger. These two events (cause and trigger) can be responsible for the complexity of mixed emotions with conflicting valence (both positive and negative at the same time). If there is only a single focus event, then we can say that trigger = cause. Even with a single event triggering the emotion, we can find situations where there is conflict between belief and emotion, between intention and emotion and between intention and moral values.

Some examples of complex emotions in real-life situations might be:

– Masking one emotion with another controlled emotion: the desire to control (hide or mask) an emotion leads to a combination of emotions where an emotion is “controlled” by masking it with another. This situation is the result of social rules and is fairly frequent in real life, where extreme cases would involve lying or manipulation. [ABR 05] cites the following example: a politician hides his disappointment about his election results with a positive emotional expression during an interview in order to not lose face. In this example, the cause event is different from the trigger.

– Mixed emotions of conflicting valence: in the following examples, collected from real-life corpora, we see the following mixed emotions of conflicting valence:

- a caller to a medical emergency hotline (CEMO corpus) [DEV 05b] is relieved because someone has come to their aid while they are suffering and are afraid. The two emotions are felt simultaneously and have conflicting valence;

- a woman leaving court has been cleared of the charges against her and wants to find her child who has been removed from her care (EmoTV corpus) [DEV 05a]. The woman displays strong emotions, crying with relief while portraying a behavior that might be perceived as negative (shame/sadness). In a first annotation experiment with a single emotional label, this example was judged to be negative from a purely visual perspective, and as positive or negative from an audiovisual perspective.

It is also possible to imagine other complex contexts in different instances of everyday life. For example, a mother may scold her child for having done something stupid while laughing inside. Some contexts seem to be characteristic of complex conflicting emotions. These examples therefore show the importance of accounting for context in modeling and the need to orientate emotional studies towards real-life data.

3.6. Applications for corpora

Emotional corpora can have many aims and applications. Exploratory, or *provocative* corpora, allow us to propose and verify annotation schemes on a limited amount of data. This can be particularly useful for exploring annotation protocols for highly complex data. We will now examine some other types of corpora and their applications, whether for automatically detecting or generating emotions.

3.6.1. *Detecting and deciphering emotions in speech*

Corpora are used to develop computational models for detecting emotions and can carry different levels of coding. Once the annotations are validated, we extract a subset of data from the corpus, including the segmented signal and its emotional and contextual annotations. However, there are several questions we need to ask before developing emotional detection systems for data retrieval or for human–machine interaction. For example, which algorithm should we use? How many classes of emotions do we think we will have to identify? How can we get the best results (data pre-processing, choice of learning algorithm, normalization, selection or combination of parameters)? Are there differences between the participants' roles (agent versus caller in a call center application) or sex (e.g. men compared to women)? These questions are further explored in Chapter 5, which focuses on classifying emotions [CLA 09]. Similarly, how will the classifiers applied to one corpus behave with regards to different data? Can one methodology developed for a specific set of data be applied to other data and can the "models" built on the basis of one corpus be tested on other data? At LIMSI, we have pursued several paths of research on detecting emotions from audio indicators, examining:

– variation in the number of emotional classes;

– additional context: differences between agent/caller, male/female, context of the emotion, interactional and semantic context;

– the importance of different types of indicators; mixture of paralinguistic and linguistic indicators.

At LIMSI, we have worked on several oral corpora including AIBO in collaboration with CEICES (Combining Efforts for Improving Automatic Classification of Emotional User States) and the HUMAINE network [BAT 06], SAVE [CLA 07], GEMEP (UNIGE) [VID 08], BOURSE and CEMO (LIMSI/call centers) [DEV 05b]. All these corpora have been collected using different protocols: AIBO (realistic induced data), SAVE (realistic fiction), GEMEP (corpus of emotions acted by professionals), BOURSE and CEMO (data collected from a real-life context). These corpora have then been annotated using different strategies and annotation controls. For example, annotations using emotion vectors [DEV 05b] allow a refined selection of data from the corpus. The annotation of context, such as the role of dialog in call centers or the type of threat, can mean we can select subsets of the corpus to develop better performing models. The learning algorithms used are mostly discriminative algorithms, i.e. they can learn to detect differences between forms. The discriminative performances are greater on acted data than on natural data, which is noisier and more variable. Performance also depends enormously on the size of the corpus used for learning, the number of locutors, type of locutor (e.g. adult or child), the number of emotions to be identified and their proximity in terms of expressive behavior. Overall, the performance of detection systems (discriminative being the most accurate word) is highly linked to the corpus, i.e. to the definitions given to emotions and the data used.

The reasoning behind multi-corpora is that we can study the portability of models created from specific data on other data collected in the same acoustic conditions (e.g. call center for another purpose) or in other conditions. It is essential to understand the importance of context in annotation if we want to build emotion-predicting models. In this emerging field, there is currently a significant lack of reference corpora acting as a benchmark for evaluating different systems and, more specifically, a lack of real data corpora of any significant size where the interactional and semantic context is annotated.

3.6.2. *Designing an expressive agent from corpora*

One of the potential applications of emotional corpora is aiding the design of expressive virtual agents. Virtual agents use a wide range of modalities, such as speech, gesture and facial expressions. These modalities provide the human–machine interface with the possibility of expressing emotions to the user, potentially with different behavior strategies by combining different modalities. The emotional behavior and expressivity of these animated agents play a central role in the use, e.g. by developing interactive narration systems.

How do we define the dynamic of each modality and their combination during emotional behavior? What level of temporality and abstraction should there be? How can we be sure that the agent's emotional behavior will be properly received by the user, especially with regard to natural emotional behavior, which is more complex than core acted emotions? The display of non-verbal behavior plays an important role in our perception of emotions. In addition to the large amount of data that we can find in literature on non-verbal communication, a multimodal emotional corpus allows us to collect precise information in a given context.

Several models have been proposed for selecting and animating agents' behavior. Results from psychological studies are highly useful for designing embodied conversational agents (ECAs) but to date have provided few details and only consider variations in contextual factors of multimodal emotional behavior. Few researchers have used context-specific multimodal corpora for designing ECAs. In a video corpus, the multimodal behavior of subjects has been annotated and used to specify REA (Real Estate Agent) agents' behavioral rules [CAS 01]. In the domain of emotional virtual agents, the majority of work carried out has until now used data performed by actors. The majority of study in this area of research uses either motion capture or videos. For example, the animations of the Greta agent [PEL 05] have been designed using real natural data corpora that is rich in emotional behavior [MAR 06]. The approach is made up of two phases:

– First, annotating the perception of the emotion on multiple levels in videos of televised interviews and then re-performing the emotions detected by the animated agent using annotations. In the first phases, the EmoTV video corpus of televised interviews has been collected and manually annotated by proposing dimensions of relevant representations for annotating the perception of the emotion. The annotation of communicative behavior in social environments is very complex due to the large number of variables involved in the communicative process.

– The second phase of the study consists of animating the agent.

Two simulations have therefore been explored: one using a manual copy of the indicators, the other using indicators taken from and calculated on annotations from the corpus. In these two tests, Greta takes the performed annotation as input during the first phase and calculates the facial animations and gestures of the agent. The annotated emotional labels and the description of the type of movement are used and programmed into the virtual agent called Greta. Transcribing the speech serves as the point of departure. Transcription is carried out by tags that allow us to control the agent's animation. This approach consists of an analysis–synthesis loop that allows us to refine the animation. The annotation of the video segment is rewritten in Greta's specification language, APML (Affective Presentation Markup Language [CAR 04]). The annotation scheme allows us to specify two emotions for the same segment. One segment corresponds to one emotional unit, a measure that is perceptive. Three annotators have validated the segmentations and have annotated each with two emotional labels: a major and minor emotion if necessary. All the annotations are then combined to give a soft vector. In one of the video extracts, the emotion corresponds to three successive segments, the first being annotated as "anger" only, the second being anger (0.67), despair (0.11), deception (0.11), sadness (0.11) and finally, the third by desperation (0.56), anger (0.33) and sadness

(0.11). From the APML text and the behavioral profile of the agent, the system calculates the animation.

3.7. Conclusion

The aim of this chapter was to present some of the core purposes for creating emotional corpora and to provide an overview of this area of research. Collecting spontaneous emotional data from the real world is vital for developing scientific knowledge that is adapted to the richness of emotional behavior in different contexts. There is currently awareness in the scientific community that in affective computing we need to work increasingly with real emotions, which are far more complex than emotions performed by actors – whether in terms of emotion detection or generation.

The sophistication of emotion induction systems is also growing. When we work on natural or induced data, the context (emotional event, locutor's role in the environment, e.g. professor/student or agent/caller) needs to be accounted for. The emotion annotation schemes and their contexts of production proposed by the HUMAINE network or the W3C working group are schemes that attempt to provide an exhaustive description of the different levels of emotional annotation. In the context of a given application, the scheme should clearly identify the necessary annotation levels in advance.

Choosing a learning corpus, i.e. a corpus of data and their representations, is closely linked with the performance of the affective system that we want to create. The evaluation of emotional systems will certainly be the most important focus in coming years and requires the development of reference emotional corpora. Research into emotions has been largely developed in recent years, but too often studies still focus on small, artificially collected corpora that are annotated following a scheme designed by other teams. It is therefore difficult to accept all the results from these studies.

The main challenges currently are:

– to acquire natural data, to identify the useful levels of annotation for use in different applications (identifying or generating emotional behavior);

– to annotate these corpora in a "reliable way" using control protocols;

– to annotate their specific interactional and semantic contexts to establish means (corpora, measures) of evaluating emotional systems.

Natural data can be obtained by recording in natural environments, in mono- or multimodal contexts, such as call centers, or by constructing emotion induction interfaces that generate natural behaviors, as in gaming contexts [DEV 08b]. Currently there is neither a standard annotation nor an evaluation corpus shared by the community. It is therefore essential to build reference corpora that can be shared by different laboratories and to define standard evaluation measures so that we can evaluate systems and improve models as well as our knowledge of emotional processes.

Finally, emotional corpora are essential to the development of emotional applications. They also allow us to raise a number of theoretical issues linked to better understanding of emotional states and their uses in human–machine interactions. The ethical dimension of this research is also highly important, whether it involves building corpora, using the collected data or developing emotional interaction systems in our daily lives.

3.8. Bibliography

[ABR 05] ABRILIAN S., DEVILLERS L., BUISINE S., MARTIN J.-C., "EmoTV1: Annotation of real-life emotions for the specification of multimodal affective interfaces", *11th Int. Conf. Human-Computer Interaction (HCII'2005)*, Electronic proceedings, LEA, Las Vegas, 2005.

[ABR 06] ABRILIAN S., DEVILLERS L., MARTIN J.-C., "Annotation of emotions in real-life video interviews: Variability between coders", *5th Int. Conf. Language Resources and Evaluation (LREC'2006)*, Genoa, Italy, 2006.

[ABR 07] ABRILIAN S., MARTIN J.-C., DEVILLERS L., MARIANI J., Représentation de comportements émotionnels multimodaux spontanés: perception, annotation et synthèse, University of Paris XI, France, September 7, 2007.

[AUD 04] AUDIBERT N., AUBERGE V., RILLIARD A., "EWiz: contrôle d'émotions authentiques", *Journées d'Etude sur la Parole, 25es Journées d'Etude sur la Parole*, Fès, Morocco, 2004.

[AX 53] AX A.F.,"The physiological differentiation between fear and anger in humans", *Psychosomatic Medicine*, 15:433-442, 1953.

[B N 04] BÄNZIGER T., Communication vocale des émotions: perception de l'expression vocale et attributions émotionnelles, UNIGE thesis, no. 325, 2004.

[B N 07] BÄNZIGER T., SCHERER K., "Using actor portrayals to systematically study multimodal emotion expression: the GEMEP corpus", *2nd International Conference on Affective Computing and Intelligent Interaction (ACII 2007)*, vol. 4738, pp. 476-487, LNCS, Springer, Lisbon, Portugal, September 12-14, 2007.

[BAS 79] BASSILI J.N., "Emotion recognition: The role of facial movement and the relative importance of upper and lower areas of the face", *Journal of Personality and Social Psychology*, vol. 37, pp. 2049-2058, 1979.

[BAT 03] BATLINER A., FISHER K., HUBER R., SPILKER J., NOTH E., "How to find trouble in communication", *Speech Communication*, vol. 40, pp. 117-143, 2003.

[BAT 06] BATLINER A., STEIDL S., SCHULLER B., SEPPI, D., LASKOWKSI K., VOGT T., DEVILLERS L., VIDRASCU L., AMIR N., KESSOUS L., AHARONSON V., "CEICES: Combining Efforts for Improving automatic Classification of Emotional user States: "a forced co-operation" initiative", *IS-LTC, Fifth Slovenian and First International Language Technologies Conference*, October, 2006.

[CAM 00] CAMPBELL N., "Databases of emotional speech", *Speech Emotion*, *1st ISCA Workshop on Speech and Emotions*, Newcastle, pp. 34-38, 2000.

[CAR 04] DE CAROLIS B., PELACHAUD C., POGGI I., STEEDMAN M., "APML, a Mark-up Language for Believable Behavior Generation", in: PRENDINGER H. and ISHIZUKA M. (ed.), *Life-like Characters. Tools, Affective Functions and Applications*, pp. 65-77, Springer, Berlin, 2004.

[CAS 01] CASSELL J., NAKANI Y.I., BICKMORE T.W., SIDNER C.L., RICH C. "Annotating and generating posture from discourse structure in embodied conversational agents", *Multimodal Communication and Context in Embodied Agents*, *5th International Conference on Autonomous Agents*, workshop, May 28th June 1st, Montreal, Canada, 2001.

[CAS 98] CASTELFRANCHI C., "Modelling social action for AI agents", *Journal Artificial Intelligence - Special Issue: Artificial Intelligence*, vol. 103, Issue 1-2, pp. 157-182, 1998.

[CLA 07] CLAVEL C., Analyse et reconnaissance des manifestations acoustiques des émotions de type peur en situations anormales, PhD thesis, TSI Traitement du Signal et des Images, ENST, 2007.

[CLA 08] CLAVEL C., VASILESCU I., DEVILLERS L., RICHARD G., EHRETTE T., "Fear-type emotion recognition for future audio-based surveillance systems", *Journal Speech Communication*, vol. 50, no. 6, pp. 487-503, 2008.

[COH 60] COHEN J., "A coefficient agreement for nominal scales", *Educ. Psychol. Meas.*, vol. 20, pp. 27-46, 1960.

[CON 02] CONATI C., "Probabilistic assessment of user's emotions in educational games", *Journal of Applied Artificial Intelligence*, special issue on Merging Cognition and Affect in HCI, vol. 16, pp. 555-575, 2002.

[COW 00] COWIE R., DOUGLAS-COWIE E., SAVVIDOU S., MCMAHON E., SAWEY M., SCHROËDER M., "Feeltrace: an instrument of recording perceived emotion in real-time", *Proc ISCA ITRW on Speech and Emotion: Developing a Conceptual Framework for Research*, pp. 224-231, 2000.

[CRA 04] CRAGGS R., WOOD M.M., "A 2-dimensional annotation scheme for emotion in dialogue", *Proceedings of AAAI Spring Symposium on Exploring Attitude and Affect in Text: Theories and Applications*, Stanford University, USA, 2004.

[DEV 05a] DEVILLERS L., ABRILIAN S., MARTIN J.-C., "Representing real-life emotions in audiovisual data with non core emotional patterns and context Features ", *1st Int. Conf. Affective Computing and Intelligent Interaction (ACII'2005)*, Beijing, China, Spinger-Verlag, Berlin, Germany, pp. 519-526, October 22-24, 2005.

[DEV 05b] DEVILLERS L., VIDRASCU L., LAMEL L., "Challenges in real-life emotion annotation and machine learning-based detection", *Neural Networks*, special issue on Emotion and Brain, vol. 18, no. 4, pp. 407-422, 2005.

[DEV 06a] DEVILLERS L., MARTIN J.-C., COWIE R., DOUGLAS-COWIE E., BATLINER A., "Workshop on Corpora for Research on Emotion and Affect", *5th International Conference on Language Resources and Evaluation (LREC'2006)*, Genoa, Italy, May 23, 2006.

[DEV 06b] DEVILLERS L., Les émotions dans les interactions homme-machine: perception, détection et génération, Habilitation thesis, University of Paris-Sud/LIMSI, France, 2006.

[DEV 06c] DEVILLERS L., COWIE R., MARTIN J.-C., DOUGLAS-COWIE E., ABRILIAN S., MC RORIE M., "Real-life emotions in French and English TV video corpus clips: an integrated annotation protocol combining continuous and discrete approaches", *5th international Conference on Language Resources and Evaluation (LREC 2006)*, Genoa, Italy, May 24-27, 2006.

[DEV 08a] DEVILLERS L., MARTIN J.-C., COWIE R., DOUGLAS-COWIE E., BATLINER A., "Workshop on Corpora for Research on Emotion and Affect", *6th International Conference on Language Resources and Evaluation (LREC'2008)*, Marrakech, Morocco, May 26 2008.

[DEV 08b] DEVILLERS L., MARTIN J.-C., "Coding emotional events in audiovisual corpora", *6th Int. Conf on Language Resources and Evaluation (LREC'2008)*, Marrakech, Morocco, May 2008.

[DOU 03a] DOUGLAS-COWIE E., COWIE R., SCHROEDER M., "The distribution of naturally occuring emotional speech", *15th ICPhS*, Barcelona, 2003.

[DOU 03b] DOUGLAS-COWIE E., CAMPBELL N., COWIE R., ROACH P., "Emotional speech: Towards a new generation of databases", *Speech Communication*, vol. 40, pp. 33-66, 2003.

[EKM 75] EKMAN P., FRIESEN W.V., *Unmasking the face – A Guide to Recognizing Emotions from Facial Clues*, Prentice-Hall, Englewood Cliffs, 1975.

[EKM 92] EKMAN P., "An argument for core emotions", *Cognition and Emotion*, vol. 6, pp. 169-200, 1992.

[EKM 99] EKMAN P., "Core emotions", in: DALGLEISH T. and POWER M.J. (eds), *Handbook of Cognition and Emotion*, John Wiley, New York, USA, pp. 301-320, 1999.

[GRA 04] GRATCH J., MARSELLA S., "Domain-independent framework for modeling emotion", *Journal of Cognitive Systems Research*, vol. 5, no. 4, pp. 269-306, 2004.

[GRO 68] GROSSBERG J.M., WILSON H.K., "Physiological changes accompanying the visualization of fearful and neutral situations", *Journal of Personality and Social Psychology*, vol. 10, pp. 124-133, 1968.

[GRO 95] GROSS J.J., LEVENSON R.W., "Emotion elicitation using films", *Cognition and Emotion*, vol. 9, pp. 87-108, 1995.

[HES 06] HESS U., "Emotion ressentie et simulée", in KIROUAC G. (ed.) *Cognition et Emotions*, pp. 115-128, 2006.

[JUS 03] JUSLIN P.N., LAUKKA P., "Communication of emotions in vocal expression and music performance: Different channels, same code?", *Psychological Bulletin*, vol. 129, no. 5, pp. 770-814, 2003.

[LAZ 98] LAZARUS R.S., *Fifty Years of the Research and Theory of R.S. Lazarus*, Psychology Press, Hove, 1998.

[MAR 06] MARTIN J.-C., NIEWIADOMSKI R., DEVILLERS L., BUISINE S., PELACHAUD C., "Multimodal complex emotions: Gesture expressivity and blended facial expressions", *Special issue of the Journal of Humanoid Robotics on Achieving Human-like Qualities in Interactive Virtual and Physical Humanoids*, vol. 3, no. 3, pp. 269-291, 2006.

[OCH 06] OCHS M., NIEWIADOMSKI R., PELACHAUD C., SADEK D., "Expressions intelligentes des émotions", *Revue Des Sciences et Technologies de l'Information, serie Revue d'Intelligence Artificielle*, vol. 20, no. 4-5, pp. 607-620, 2006.

[OSG 75] OSGOOD C., MAI W.H., MIRON M.S., *Cross-cultural Universals of Affective Meaning*, University of Illinois Press, Urbana, 1975.

[PLU 84] PLUTCHICK R., "Emotions: A general psychoevolutionary theory", in: SCHERER K.R. and EKMAN P. (eds), *Approaches to Emotion*, Erlbaum, Hillsdale, NJ, USA, pp. 293-317, 1984.

[SCH 81] SCHERER K., CHAPTER 10, "Speech and emotional states", in DARBY J. (ed.) *Speech Evaluation in Psychiatry*, pp. 189-220, Grune and Stratton, New York, USA, 1981.

[SCH 99] SCHERER K., "Appraisal theory", in: DALGLEISH T. and POWER M. (eds), *Handbook of Cognition and Emotion*, pp. 637-663, Wiley, Chichester, 1999.

[SCH 04] SCHERER K.R., WRANIK T., SANGSUE J., TRAN V. and SCHERER U., Emotions in everyday life: probability of occurence, risk factors, appraisal and reaction pattern", *Social Science Information*, 43, pp. 499-570, 2004.

[SCH 07] SCHROËDER M., DEVILLERS L., KARPOUZIS K., MARTIN J.-C., PELACHAUD C., PETER C., PIRKER H., SCHULLER B., TAO J., WILSON I., "What should a generic emotion markup language be able to represent?", *Proceedings of the 2nd International Conference on Affective Computing and Intelligent Interaction (ACII 2007)*, vol. 4738, pp. 440-451, Springer, Lisbon, Portugal, September 12-14,2007.

[STI 00] STIBBARD R., "Automated extraction of ToBi annotation data from the Reading/Leeds emotional speech corpus", in *Proc. ISCA Workshop Speech Emotion*, pp. 60-65, 2000.

[VID 06] VIDRASCU L., DEVILLERS L., "Représentation et détection des émotions dans des données issues de dialogues enregistrés dans des centres d'appels: des émotions mixtes dans des données réelles", *Revue des Sciences et Technologies de l'Information*, série "Revue d'Intelligence Artificielle", vol. 20, No. 4-5, pp. 447-476, 2006.

[VID 08] VIDRASCU L., DEVILLERS L. "Emotions detection performances on two corpora with felt and protrayed emotions", *Proceedings of the workshop on Corpora for Research on Emotion and Affect, 6th International Conference on Language Resources and Evaluation (LREC'2008)*, Marrakech, Morocco, May 26 2008.

[VOG 05] VOGT T., ANDRÉ E., "Comparing feature sets for acted and spontaneous speech in view of automatic emotion recognition", *Proceedings of the 2005 IEEE International Conference on Multimedia and Expo*, ICME, Amsterdam, The Netherlands, July 6-9, 2005.

[ZAR 07] ZARA A., MAFFIOLO V., MARTIN J.-C., DEVILLERS L., "Collection and annotation of a corpus of human-human multimodal interactions: emotion and others anthropomorphic characteristics", *Proceedings of the 2nd International Conference on Affective Computing and Intelligent Interaction (ACII 2007)*, vol. 4738, pp. 464-475, Springer, Lisbon, Portugal, September 12-14, 2007.

Chapter 4

Visual Emotion Recognition: Status and Key Issues

4.1. Introduction

In this chapter we will focus on the automatic visual recognition of facial emotional expression. This kind of visual recognition requires the presence of visible signs indicating an emotion. However, the term emotion in itself can be misleading. For example, the *Oxford English Dictionary* defines emotion as a strong feeling, such as joy or anger, which is an instinctive feeling as distinguished from reasoning or knowledge. A facial expression, however, is a change in the face due to a voluntary or involuntary movement of several muscles demonstrating an emotional state and acting as a reflection or regulator of discourse [KAI 01]. An emotion can therefore cause (or not) a facial expression and vice versa. Since these two terms are not equivalent, we will use the term facial expression for these semantic reasons.

Interest in automatic facial recognition has grown sharply in recent years due to the multitude of potential applications for emotion recognition. Numerous conferences have held special sessions on the subject: International Conference on Image Processing (ICIP), International Conference on Automatic Face and Gestures Recognition, International Conference on Multimedia Interaction (ICMI), Human Computer Interaction (HCI) and a network of excellence [NOE 04] has also been set up to focus on this area of research.

Chapter written by Alice CAPLIER.

Emotion-Oriented Systems Edited by Catherine Pelachaud

The appearance of a facial expression is a manifestation of an emotional state or intention. Facial expressions are therefore an essential means of nonverbal communication. This information represents highly significant data for applications such as human–machine interfaces because we can make them behave more naturally by enabling them to adapt their behavior in accordance with the user's detected emotional state (see Chapter 8 for further details). In practice, automatic recognition of facial expressions can also be used for *e-learning* systems [LOH 05], for creating realistic conversational agents [NIE 07], and developing interactive games [ZHA 06], etc.

Despite ongoing research, there is as yet no solution to the problem of automatic recognition of facial expressions. Until now, researchers have focused on automatic recognition of the six universal expressions, which include happiness, sadness, fear, anger, disgust and surprise, by working on caricature data (facial expressions deliberately produced by actors). There have also been a number of highly detailed and comprehensive reviews of research into this area [CHI 03, FAS 03, PAN 00a]. For example, Tian *et al*. [TIA 03] have examined advances in the field since the 2000s.

This chapter also focuses on the emerging challenges facing us in visual facial expression recognition including:

– Automatic recognition of spontaneous facial expressions, i.e. facial expressions that have been naturally created and not intentionally reproduced.

– Accounting for dynamic modality in the recognition process (analyzing sequences of images and not merely photographs).

– Estimating expression intensity as well as just recognizing the emotion.

– Introducing several modalities and accompanying features, such as head and body movement and voice analysis.

In section 4.2 we examine what we mean by the term "facial expression" and how the face communicates this. In section 4.3 we provide a brief overview of previous research into facial expression recognition with reference to review articles that can give a more in-depth perspective on the subject. In the rest of the chapter, we will examine the current challenges in the field, such as analyzing spontaneous expressions, evaluating the intensity of facial expressions, analyzing the temporal dimension of a facial expression and, finally, the issue of accounting for several modalities for recognizing an expression.

4.2. What is a facial expression?

4.2.1. *Definition*

Physically, a facial expression is communicated by a non-rigid and transient flexing of facial features (mouth, eyes, eyebrows, wrinkles, etc.) This flexing results in the contraction of one or more of the 44 muscles that make up the face. Figure 4.1 illustrates some examples of facial expressions. The website www.artnatomia.net/uk/ [ART 04] similarly allows us to view the effect of different muscular contractions interactively.

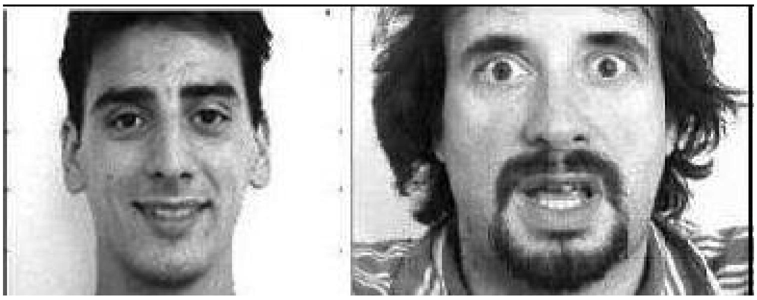


Figure 4.1. *Facial expressions and distortions*

4.2.2. *Description*

A facial expression is characterized by three components:

– the type of facial movement;

– the intensity of this movement;

– the speed and duration of this movement.

These first two components are spatial in the sense that we can analyze them on the basis of one image, such as a photograph. In contrast, the last component is temporal and can only be obtained after a complete study of the image sequence linked to the facial expression being considered.

Ekman [EKM 78] has developed the FACS (Facial Action Coding System) for analyzing facial movements. This system identifies 44 action units (AUs) that show the visual effects of activating different facial muscles. Figure 4.2 provides some examples of these action units and Table 4.1 details the majority of action units.

Figure 4.2. *Action units from left to right:AU4– frowning; AU9–wrinkled nose; AU23–pouting; AU28– pursed lips; AU43–closed eyes*

AU1	Rasing the inner eyebrow	AU2	Raising the outer eyebrow
AU4	Furrowed brow	AU5	Raising the upper eyebrow
AU6	Lifting the cheeks	AU7	Narrowing the eyes
AU9	Furrowed nose	AU10	Raising the upper lip
AU11	Nasolabial furrow	AU12	Raising the corners of the mouth
AU13	Puffed up cheeks	AU14	Lowering the corners of the mouth
AU15	Lowering the corner of the mouth	AU16	Lowering the lower lip
AU17	Lifting the chin	AU18	Rounding the lips
AU20	Stretching the lips	AU22	Pouting lips
AU23	Pursing the lips	AU24	Pursed lips
AU25	Opening the lips	AU26	Lowered jaw
AU27	Opened lips	AU38	Flared nostrils
AU39	Restricting the nostrils	AU41	Closing the eyelids
AU43	Closing the eyes		

Table 4.1. *List of the main action units in the FACS system*

Using this FACS system, the observer can describe any facial expression as a combination of a set number of action units. For example, the facial expression associated with surprise results in the activation of four action units including AU1, AU2, AU5 and AU25.

The intensity of a facial expression is characterized by the amount of facial movement and by the density of expression lines appearing in specific areas of the face, such as the forehead or the nasolabial plane. For example, for a single facial expression the mouth or eyes may be more or less open and there may be lines on the forehead. Figure 4.3 shows several examples of facial expressions associated with sadness. We can see that the expression of sadness is more or less pronounced and this is accompanied by different facial movements.

In the updated version of the FACS system from 2000, Ekman included additional action units to account for differences in facial expressions' intensity. As such, the action units AU24, AU25 and AU26 relating to opening the mouth are coded according to five degrees of intensity ranging from A to E. This also applies to action units AU41, AU42 and AU43, which refer to opening the eyes. Since a facial expression is a transitory transformation of the face, this process is therefore restricted in time and is composed of three phases:

– the initiation phase (*onset parameter*);

– a period of maintenance (*apex parameter*);

– a relaxation phase (*offset parameter*).

The process of creating a facial expression lasts between 250 milliseconds and 5 seconds and some facial expressions are characterized by subtle changes in the face that can only be seen through a dynamic study of the process.

Figure 4.3. *Expression of sadness with three different intensities [DAT 04]*

4.2.3. *Ekman's universal expressions*

The more or less simultaneous contraction of facial muscles can lead to the creation of a large number of different facial expressions. Among this group of expressions, Ekman has highlighted the existence of six universal emotions [EKM 72, EKM 82, EKM 99], since they are innate and found in all cultures. These six emotions are, in general, associated with specific facial movements and therefore generate facial expressions. These six emotions include happiness, anger, surprise, fear, sadness and disgust (further details can be found in Chapter 1). Research into automatic recognition of facial expressions is therefore naturally focused on the analysis of these six universal expressions.

4.2.4. *An ideal system*

Tian *et al.* [TIA 03] have proposed that the ideal facial expression recognition system should be able to:

– cope with variation between speakers, movements of the head, partial occlusion of the face and to changes in lights and image resolution;

– recognize all kinds of expressions, different intensities of expressions and asymmetric and spontaneous expressions;

– be entirely automatic (i.e. without any human intervention) from acquiring the image of the face to recognizing its expression;

– function in real time.

Unfortunately, no such a system exists at this current time.

4.3. Overview of facial expression recognition methods

A facial expression recognition system is composed of three main phases (see Figure 4.4):

– a phase locating the face (*face extraction*);

– a phase extracting the characteristic indicators of potential deformations of the face (*features extraction*);

– a recognition or classification stage (*facial expression classification*).

Before looking at these three phases in closer detail, we will first examine some of the basics of facial expression data.

Figure 4.4. *Structure of a facial expression recognition system*

4.3.1. *Databases*

The first difficulty when attempting to build a facial expression recognition system is the construction of a database. Work on facial expression recognition has therefore run parallel with the difficult task of creating databases. Several databases are now available for research purposes and are generally made up of images of induced and stereotypical facial expressions produced by actors. The facial expressions considered are generally the six universal expressions and the images are front-on views of the face. Several databases have been specifically designed for facial expression recognition, of which the following are available to the public (other non-published and restricted access databases are detailed in [PAN 05]):

– The JAFFE database [DAT 98, LYO 98] (see the left-hand side of Figure 4.5) contains 213 images of seven facial expressions in gray scale (the six universal expressions and a neutral expression) performed by 10 Japanese women. The expression in each of these images was then evaluated by 60 subject judges.

– The CMU database [DAT 00, KAN 00] (see right-hand side of Figure 4.5) contains facial expressions produced by 100 different subjects. Precise instructions were given to different subjects so that they could recognize the use of specific muscles as well as the six universal expressions. The images in this database have been annotated by a FACS expert according to whether they activated specific action units. The images are in gray scale and are 640 x 480 pixels. It should be noted that labeling images according to the FACS system is a very delicate process: teaching an expert requires 100 hours of training and it can take up to 10 hours to annotate each minute of video.

Figure 4.5. *Left: three images from the JAFFE database [DAT 98, LYO 98]. Right: two images from the CMU database [DAT 00, KAN 00]*

– The HCE database [DAT 05a] is made up of video sequences relating to the three facial expressions of happiness, disgust and surprise along with a neutral expression. The expressions are simulated by 21 different subjects. Each sequence lasts approximately 5 seconds and begins with a neutral expression followed by the creation of an expression and then a return to the neutral expression. These images are in color in BMP format and are publicly accessible. The images have been labeled in terms of facial expressions by five subject judges.

– The MMI [DAT 05b] has recently been developed [PAN 05] with the aim of serving as a reference point for comparing the performance of several facial expression recognition systems. This is therefore the most complete database since it contains both fixed images (250) and video sequences (1,000), simple action units and combinations of AUs leading to the six universal expressions, including both front-on and profile views (see Figure 4.6) against either a textured or plain background. The 20 subjects producing the emotional expressions have been trained beforehand by a professional and the ground truth is available for some of the images, since they have been labeled in terms of AUs by experts. The video sequences are a neutral–expression–neutral sequence.

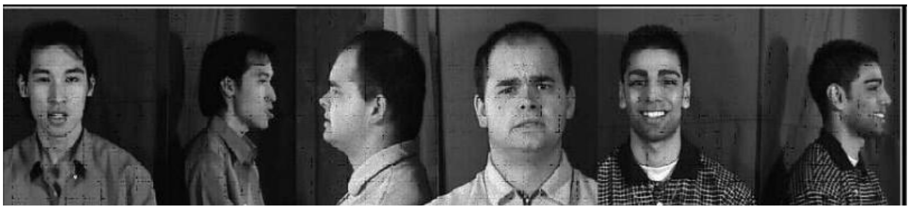


Figure 4.6. *Example images from the MMI database [DAT 05b]*

Other interesting data can be obtained from databases developed for recognizing people rather than for expression recognition. There are three standard databases for evaluating facial biometrics systems. Since facial identification algorithms are sensitive to changes in expression, the standard base images for testing facial recognition algorithms contain images with different facial expressions. These images could potentially be used for facial expression recognition. Other databases include:

– The FERET database [DAT 01, PHI 00] containing more than 14,000 images of faces in grayscale. Only a portion of these images relate to facial expressions. There is no "official" classification of the facial expressions in the FERET database.

– The Yale B database [DAT 97] is made up of 165 gray scale images from 15 different individuals. Some of these images relate to facial expressions of sadness, happiness and surprise.

4.3.2. *Preprocessing: extracting the face*

The first stage of preprocessing involves extracting the area of the face from the image to be processed. This is an area in which for a number of years researchers have attempted to develop algorithms for extracting faces. For further details on this field, refer to [HJE 01, YAN 02]. There is also the algorithm proposed by Viola and Jones in [VIO 04] that has the advantage of proposing an implementation in C. This algorithm is available to the public [MAC 04]. Similarly, Garcia and Delakis [GAR 04] have proposed a second algorithm that performs well but there is no associated code and the complete development of the algorithm from the article's specifications is highly delicate [GAR 04]. Figure 4.7 provides an example of face location obtained by the algorithm described in [VIO 04].

Figure 4.7. *Extraction of a bounding box around each face*

There is a current trend in the extraction phase to focus on assessing the position of the face. Some recent literature has proposed expression-recognition systems based on different views of the face and profiles [PAN 06]. For example, Kleck and Mendolia's [KLE 90] study has shown that humans recognize negative facial expressions (fear, sadness, etc.) more easily by viewing the face front-on and vice versa for positive expressions (e.g. happiness). [MUR 08] has undertaken a review of recent methods developed to analyze the position of the face.

4.3.3. *Extracting facial characteristics*

Facial characteristics provide information that allows us to detect and recognize facial expressions. The problem here consists of reducing the quantity of information conveyed by the face while preserving all the information needed for recognizing an expression. Facial characteristics are processed by classification algorithms, which can be split into two categories:

– The local geometric characteristics that describe the face using a set of discrete points and relations that link them in terms of movements and their respective configurations. Experiments on emotion perception have shown that humans can recognize a facial expression by analyzing the evolution of the expression over time according to a set of discrete points on the face (points linked to changes in permanent features of the face between the eyes, mouth and eyebrows) [BAS 78]. These geometric characteristics are generally linked with extracting points or facial contours. Figure 4.8 provides an example of the face where the specific dimensions are calculated using a contour segmentation phase. These dimensions demonstrate the opening of the mouth, which can vary according to facial expressions (e.g. a smile, etc.).

– Overall facial characteristics highlight the changes occurring in the face, such as the appearance of wrinkles, opening of the mouth, etc. The features are overall characteristics (we consider the face as a whole). One method for studying changes in facial appearance is the use of Gabor wavelets (see Figure 4.9).

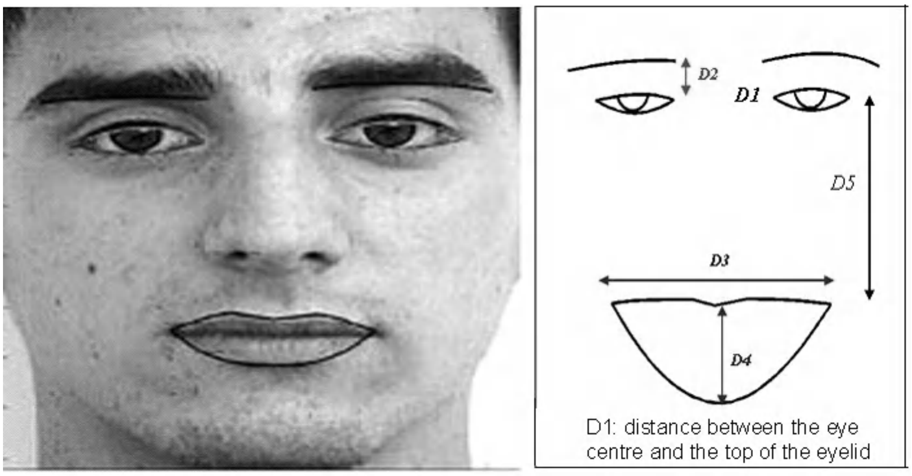


Figure 4.8. *Left: a segmentation of permanent facial features; right: extraction of geometric distances*

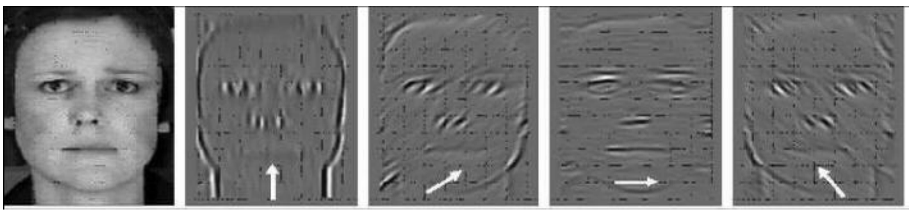


Figure 4.9. *Left: a facial expression; right: the same image following Gabor filtering according to four different orientations (taken from [FAS 02])*

In most cases, whatever the extracted facial characteristics, a normalization phase is necessary to reduce the influence of variations in scale and position. It is these variations in facial characteristics in relation to their values on a blank (or neutral) expression that are considered and not their absolute values. Among the various methods for recognizing facial expressions, some rely on local geometric characteristics or overall appearance or a combination of the two.

4.3.4. *Classification*

The classification stage is the last phase of the system. We first need to identify all the possible classes. There are two types of classification: one that consists of classifying in terms of action units and one that involves recognizing the facial expressions in relation to the six universal emotions. The classification methods used for the two types of classification (in AUs or universal expressions) can be regrouped in terms of static recognition (where we use information from a single image) and a dynamic recognition method (using the information from a series of images). The algorithms used for classification are classic algorithms, such as neural-based networks, discriminative linear analysis, SVM (Support Vector Machine), hidden Markov model (HMM), Bayesian networks or expert systems. For example, Hammal *et al.* [HAM 07] have classified facial expressions according to a new approach using evidence theory. This is justified by that fact that a facial expression is never "pure". For example, there can be a mixture of facial expressions and this approach means that we do not have to make a systematic choice but can consider several expressions (see Figure 4.10).

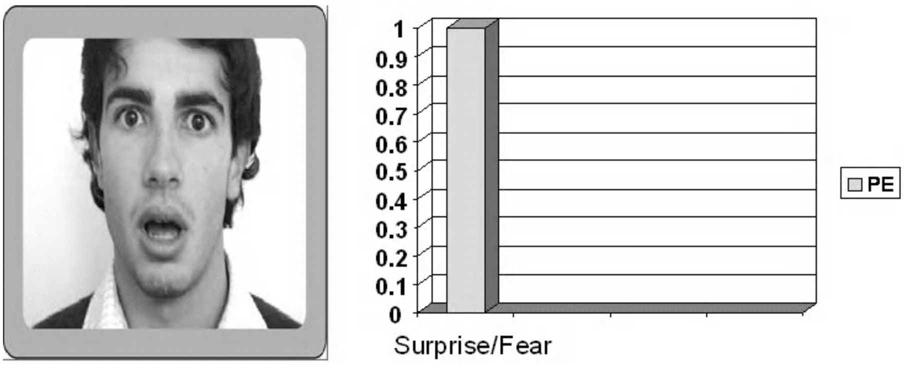


Figure 4.10. *Classification using evidence theory: the system is sure that the facial expression is either fear or surprise but cannot choose between the two*

4.3.5. *Performance*

As always in image analysis, there is the challenge of knowing how to evaluate an algorithm's performance in absolute terms and as well as knowing how to compare the performance of several algorithms. One possible way of evaluating the absolute performance of various facial expression-recognition algorithms is to use classification rates from an image database. According to the figures proposed by [FAS 03, PAN 00a, TIA 03], the rate of recognition varies between 70 and 100%. The validity of these rates is questionable, however, since there are few databases where the method for obtaining ground truth is fully described. For this reason, if we hypothesize that the ideal recognition system is the human system, then each of these images in the database needs to be labeled by human operators. The difficulty here is that operators vary hugely in the way they perceive emotion (influence of sex, age, social status, etc.). The labeling process should therefore be carried out by a group of several subject judges, which is rarely the case.

Comparing the performance of several different algorithms is also difficult. This is because we need to have previously identified a shared test database and while Cohn-Kanade [KAN 00] is often used, there is no absolute test database. As such, it is very difficult to know which system performs the best in terms of recognition rate. The previously cited review articles have focused on a type of classification that does not rate performance in terms of recognition rate but in terms of the similarity of the system in question to the ideal classification system. As such, the systems proposed by [ESS 97, PAN 00b, TIA 01] are those that satisfy the ideal system's constraints the most, although they have not all been verified.

More recently, Bartlett *et al.* have compared several *machine learning* algorithms [BAR 05]. All the algorithms considered have been evaluated using the Cohn-Kanade database to classify the universal expression and to recognize 17 AUs from the FACS system. The best results were found with an algorithm selecting a subset of Gabor filters using the Adaboost algorithm to extract facial characteristics, which are then used by a SVM-based classification system.

While there is still no ideal recognition system, fresh advances have been made in recent years. It is these new developments that we will examine in the rest of this chapter.

4.4. Spontaneous facial expressions

4.4.1. *Position of the problem*

Recent and ongoing research is focused on analyzing "natural" emotions. Unlike the work that has been carried out in the past, attention has now turned to spontaneous facial expressions, i.e. those produced as a result of appropriate induction and not caricatured emotions produced by actors. Psychological studies [COH 04, ROI 04, SCH 03] have found that spontaneous expressions can differ from deliberate facial expressions in appearance and temporal sequencing. Figure 4.11 shows both a spontaneous and forced smile. In terms of appearance, the difference between the spontaneous and forced smiles lies in whether the muscles around the eyes are activated. It has also been shown in [COH 04] that there is a correlation between a spontaneous smile and its strength and duration that is not found with forced smiles. The strength of a spontaneous smile is also weaker than that of a forced smile.

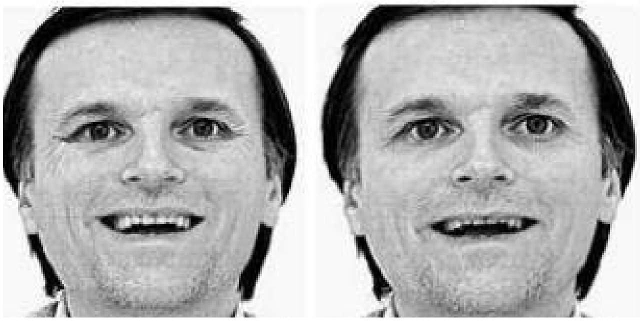


Figure 4.11. *A spontaneous expression versus a forced expression: left – a natural (or Duchenne) smile; right – a forced (or social) smile*

The differences between deliberate and spontaneous facial expressions mean that the systems designed around deliberate expressions cannot be applied to contexts such as developing intelligent human–machine interfaces. Research has therefore turned to analyzing spontaneous expressions. This raises several challenges, however, such as that of building databases of spontaneous expressions and developing adapted recognition algorithms.

4.4.2. *Databases*

Creating a database of spontaneous expressions is a research problem in itself that requires a large amount of time and is composed of two phases:

– Collecting video data in natural conditions: this consists of requesting a facial expression from a subject using an *ad hoc* protocol. This is then filmed without the subject's awareness so that this does not affect his/her reaction. However, it could be suggested that the mere fact of bringing participants into a lab for an experiment introduces a certain amount of bias.

– Labeling the collected data: since the data will be used for developing and testing automatic expression recognition systems, we need to have ground truth. How can we obtain this?

Picard *et al*. have set out five factors that can influence the relevancy of data collected [PIC 01]:

– Is the subject stimulated by a situation out of his/her control?

– Have the data been collected in a laboratory or in the subject's everyday environment?

– Was the subject aware of being recorded?

– Was the subject aware of participating in an experiment about facial expressions?

– Does the expression produced correspond to a tested feeling?

There is currently no database of spontaneous expressions acting as a reference point in natural expression recognition. We will now examine some initial attempts at this in terms of both protocol and labeling ground truth.

In their 2002 study, for example, Moriyama *et al*. [MOR 02] needed a database for the study and recognition of spontaneous eye blinking. They used videos from [FRA 97] and based their protocol around deception. The participants had to admit to or deny having stolen a significant amount of money. In order to create as natural conditions as possible, a reward or punishment system was put in place. The subjects were filmed using a color 640 × 480 camera. Following filming, the videos were labeled manually as AUs by two FACS-qualified coders.

Similarly, Bartlet *et al*. [BAR 05] used a data-collection protocol based on a false opinion model. They first asked participants to fill out a questionnaire about their opinions on a range of sensitive issues where they could either lie or tell the truth. They were then interviewed and had to convince the interviewer that they had told the truth throughout the questions. This model induced a wide range of facial expressions in the participants. It is not indicated, however, whether the subjects were aware of being filmed. Following recording, the video sequences were labeled by two FACS-qualified coders to establish ground truth.

The protocol used by Zeng *et al.* [ZEN 06] to induce emotions was the AAI (*adult attachment interview*) protocol, which consists of a semi-structured interview around participants' feelings, such as previous parent/child experiences. The subjects were not aware of being filmed. Those leading the interview referred to a standard questionnaire and the data from the two subjects (a man and a woman) were recorded resulting in two videos of 39 and 42 minutes respectively. These videos were then analyzed by two FACS experts who encoded the sequences where a facial expression appeared on the participants' faces. No audio data were included in this analysis.

In their study, Sebe *et al*. [SEB 06] presented a series of different films (comedy, horror, thriller, etc.) to participants. The camera was hidden and the 60 participants were not aware that their emotional reactions were the focus of research. After having seen these different films, the subjects were asked to complete a questionnaire about their impressions of each film. The means of establishing ground truth was not clarified. Labeling focused on three facial expressions (happiness, surprise and disgust) and on neutral expression.

As yet, none of the databases used in these four experiments are accessible to the public.

Before concluding our discussion on databases, we should first examine the Dynemo project carried out by the *Agence Nationale de la Recherche* (National Research Agency) between 2006 and 2008. The project aimed to create a database of facial expressions obtained in natural conditions. The building of the database involved several phases:

– Methods of inducing the emotions to be studied were first identified. Eleven desired emotions were selected: amusement, shame, annoyance, satisfaction, boredom, disgust, surprise, interest, being emotional, humiliation and neutral. The induction material was composed of slide shows where the subject did or did not react with the desired emotion.

– Video data were collected from 400 subjects using a hidden camera. The subjects were unaware that the focus of the experiment was emotions (see Figure 4.12).

– Video data were labeled by the subjects themselves via questionnaires during debriefing.

– Subject judges then labeled the video data a second time, indicating the facial expressions recognized throughout the video sequence. This demonstrated that the potentially large difference between the emotion actually felt and its manifestation as a facial expression. Figure 4.12 shows an example result for a video sequence where the subject was believed to have felt disgust. From image number 450, more than 60% of subject judges considered the subject to have expressed disgust.

At the end of the project, the Dynemo database will provide a series of videos of spontaneous emotions where the ground truth will consist of the emotion genuinely felt by the subject producing the video. It will also provide a temporal cross-section of the various emotional sequences recognized by around 20 subject judges. This cross-section will also include the recognition rate by the subject judges of secondary facial expressions if necessary (i.e. facial expressions other than those previously induced). This database will be freely accessible for all non-commercial use [DAT 08].

4.4.3. *Recognizing spontaneous expressions*

The findings from these spontaneous data collection projects are as follows:

– not all facial expressions correspond to Ekman's six universal expressions;

– there is large interpersonal variation in the way we express emotion. For example, women are generally more expressive than men;

– the reaction of several individuals to the same stimulus can be very different for family or cultural reasons;

– the face can be subject to highly subtle changes. For one given expression, these changes can be less obvious in spontaneous expressions as opposed to deliberate expressions (see Figure 4.11);

– a facial expression is rarely absolute. It is often a mixture of several expressions;

– the subject is not guaranteed to return to a neutral expression when changing from one facial expression to another. We can therefore encounter co-articulation effects that will modify the way an expression is produced.

As a result, the first problem facing us is that of defining the classes of emotional expressions to be recognized. There are several ways of approaching this. With regard to the complexity of the classes of possible spontaneous facial expressions, Zeng *et al*. have proposed a minimalist categorization of facial expression/no facial expression [ZEN 06]. Sebe *et al*., on the other hand, have found that the four expressions they examined (happiness, disgust, surprise and neutral) corresponded to the four expressions displayed during the data-acquisition process [SEB 06]. In their study, however, Moriyama *et al*. [MOR 02] viewed this problem in terms of identifying three situations: a single blink, multiple blinks or no blink corresponding to whether AU45 (complete closing of the eye) and AU42 (partial closure) were carried out. In the study by Bartlett *et al*. [BAR 05], classification was based on 16AUs (based primarily on the eyebrows, lips, chin and mouth).

While describing facial expressions using a series of discrete labels is instinctive, this does not provide a full overview of these expressions. An alternative might therefore be to look at representing different emotional states on a 2D continuum of valence/arousal (see Chapter 6 for further details). The valence axis indicates whether we have a positive (e.g. happiness) or negative (e.g. disgust) expression and the arousal axis is linked to the individual's propensity in a particular emotional state to undertake (or not) a particular action. At this moment in time, the problem of classifying facial expressions on a valence/arousal continuum (or even on a continuum of several dimensions) remains largely unresolved.

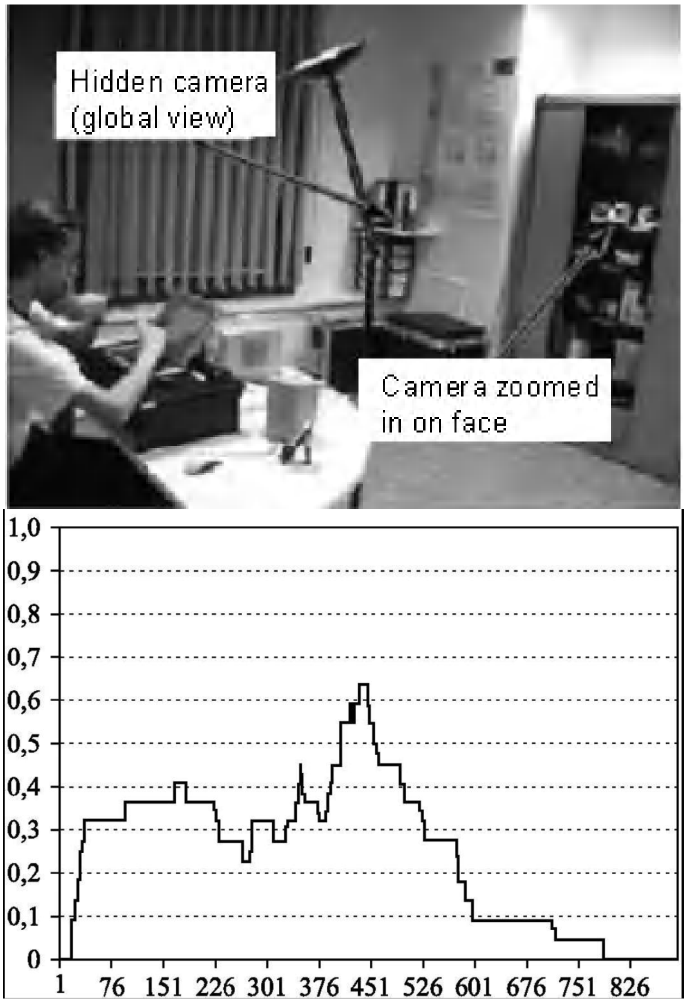


Figure 4.12. *Top: experimentation room for data collection. Bottom: ground truth established by 20 subjects for an expression of disgust*

With regard to the analytical system employed, the flow diagram proposed in Figure 4.4 remains valid. A preliminary stage consisting of locating the face is still necessary; however this is generally more difficult for spontaneous emotions. This because when recording this type of facial expression it is impossible to demand that the subject stands exactly in front of the camera and remains immobile, even if the specified experimental protocol requires the participant to be subject to these conditions. The mathematical tools for the classification phase remain the same, although the first tests will preferably be oriented toward classification methods requiring fewer data for learning. For example, Zeng *et al.* [ZEN 06] have found that the expected classification is based on expression/no expression and that classifying an expression into a single class is preferable in order to avoid modeling a non-facial expression that is, by definition, a broad group and therefore difficult to learn with limited data.

It is when extracting characteristics that the most significant differences appear. For example, motion parameters extracted from a 3D face modeling and tracking have been used in [SEB 06, ZEN 06]. Using a 3D model of the face is preferred because this allows us to account for head movement. Bartlett *et al.*, for example, have used a battery of Gabor filters to account for subtle facial changes [BAR 05].

4.5. Expression intensity

Human beings are capable of producing a multitude of different facial expressions. Even among Ekman's six universal emotions there are several other expressions that appear according to the intensity of facial movement. Fear, for example, can be expressed as slight fear or high anxiety, if not pure terror. FACS enables us to view an emotion on an activation scale of different AUs. Some AUs are split into several possible intensities represented as five grades ranging from A to E (see Figure 4.13). There is generally not an equal breakdown of the intensity scale, as most cases lie somewhere between C and D.

A	B	C	D	E
Trace	Slight	Marked/Pronounced	Severe/Extreme	Max

Figure 4.13. *The different intensity levels as defined in FACS*

Research into automatic analysis of the intensity of facial expressions is currently in its early stages. One of the main problems with this is finding a precise definition for this notion of intensity.

FACS grades intensity according to five levels that are linked to the 12 AUs that Fasel and Luettin have attempted to identify in their system [FAS 00]. Classification is achieved by analyzing the differences between the image in question and a reference image of a neutral expression. A reduction of the problem's complexity is achieved by independent component analysis and using a k-nearest-neighbor classifier. The proposed system is tested on a database where the ground truth has been obtained by analyzing the movement of specific markers. According to the series of images tested, the recognition rate of AU intensity varies between 30 and 40%. Fasel and Luettin concluded that the most difficult aspect is recognizing the level of arousal. For example, Ghanem *et al.* [GHA 08] have reduced these five arousal levels to three by grouping levels A and B as "weak intensity" label, D and E as "strong intensity" and C as "medium intensity". Instead of studying the AU, they proposed basing their evaluation of intensity on analyzing facial distances (mouth open, eyes open, distance from each eye to each eyebrow),which vary when a facial expression appears. There is not always a clear transition between the three classes of intensity, but a recognition system using the formulism of evidence theory allows a classification that can differentiate between the weak/medium and medium/strong classes. The proposed results showed that the intensity recognition rate varies between 50 and 100%, with the biggest errors occurring when evaluating very weak intensities that can be difficult to distinguish from a neutral expression.

Another way of measuring facial expression intensity is to measure the relative movement of the face when expressing an emotion (using displacement of points or speed vector analysis) in relation to a neutral face. This can be characterized by a continual scale of intensity ranging from minimum to maximum intensity [LEE 03, LIE 98b, LIS 98].

Whatever the definition chosen to characterize facial expression intensity, little research has been carried out in this area. One of the main reasons for this is that there is a lack of labeled databases. There is one publicly available database, the DAFEX database [BAT 04, DAT 04], that contains 1,000 video sequences of deliberate facial expressions (including Ekman's six universal expressions and a neutral expression). These facial expressions were produced by eight professional actors with three levels of intensity (weak, medium and strong). The ground truth therefore comes from the actors in videos where there was more or less intensity.

4.6. Dynamic analysis

Numerous psychological experiments have highlighted the existence of a dynamic dimension in the production of a facial expression and this dynamic component is involved in the human recognition process [BAS 78, SCH 99]. For example, Bassili has found that an observer can recognize a facial expression by examining a visualization of the movement of certain points on the face only. Temporal information also helps the (albeit weaker) recognition of facial expressions, such as shame or amusement. While this has been established, little research has focused on this temporal dimension. Since we are interested in spontaneous facial expressions, we cannot ignore interpersonal variability with regard to the production of facial expressions, since there is little variation with deliberate facial expressions where all the participants are supposed to react in a predefined way. The dynamic component is part of information that differs according to different expressions produced by different people.

Research relating to the dynamic analysis of facial expressions has focused on two main problems:

– Automatic segmentation of a continual sequence of emotions into different emotional groups. Pantic and Patras, for example, have hypothesized that we can differentiate emotional and non-emotional segments because when there is a facial expression there is movement on the face and vice versa for a non (i.e. neutral) expression [PAN 06]. This model therefore presents the distinction as being a question of activation or non-activation of AUs corresponding to an exact division of facial expressions into different segments if the subject maintains a neutral expression before producing a new facial expression. Similarly, Otsuka and Ohya [OTS 97] have used a method where the peak of a facial expression is detected as being associated with a peak of a facial movement and the expression is supposed to begin and finish at the lowest point of the preceding and following facial movements.

– Modeling the temporal signature of a given facial expression (potentially for a specific individual). Zhang *et al.*, for example, have attempted to understand the temporal behavior of a facial expression [ZHA 05]. In a preliminary modeling phase, infrared images of the face are analyzed in order to understand the temporal development of a facial expression. A dynamic facial expression-recognition system based on dynamic Bayesian networks has been proposed as being able to account for the temporal aspect of a spontaneous facial expression. The research carried out by Shan *et al*. [SCH 06] has found that the appearance of each face is characterized by the local binary pattern (LBP) indicator and a temporal Bayesian model that provides a compact temporal representation of the face data. This Bayesian model is then combined with a decisional voting-based scheme in order to automatically identify the expression present in a sequence of analyzed images. Shan *et al*. also detail a comparison of recognition performance between a static system (based on a nearest-neighbor classifier) and the proposed dynamic system. At this time, this system can only process a series of images beginning with a neutral expression and finishing at the peak of the facial expression. Figure 4.14 provides an example result on such a video sequence.

Another proposed classification is that of a two-phase dynamic classification [YEA 04]. The first phase consists of the nearest-neighbor classifier generating the temporal signature of an emotion, which is based on optic flow characteristics. HMMs are then used to model the temporal signatures of each of the six basic expressions. Cohen *et al*., for example, have used a HMM multi-level-based system to simultaneously solve the problem of segmenting a continuous video into segments of different facial expressions and the problem of recognizing facial expressions in a dynamic context [COH 03].

Whatever the method employed, it seems that dynamic classifiers are more complex because they require more data for the learning phase and the control of a greater number of parameters. However, dynamic classification is also indispensable in accounting for the challenge of interpersonal variability, which is intrinsic to spontaneous facial expressions.

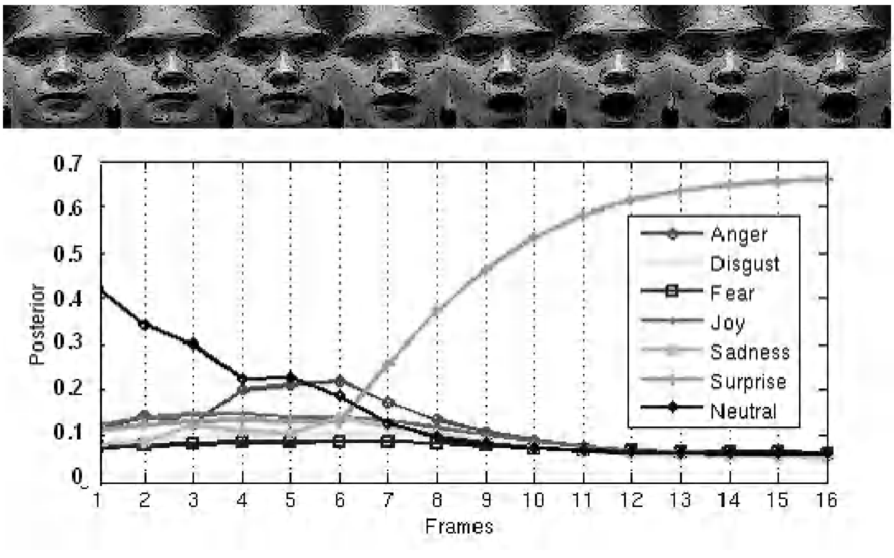


Figure 4.14. *Dynamic recognition of the facial expression of surprise (taken from [SCH 06])*

4.7. Multimodality

All psychological studies agree that an individual's emotional state is mediated by a number of indicators including not only facial expression but also the voice and body language [RUS 03, SCH 07, VAN 07]. Human perception of emotions is therefore multimodal by its very nature. As a result of research into affective human–machine interfaces, the issue of multimodal automatic systems for emotion recognition has grown in importance [COW 01, FRA 05]. These new-generation human–machine interactions must be able to recognize the emotional state of the user. Since there have already been several review articles on the subject, we will not provide an exhaustive description of the research carried out in this area. For further information on the study of audiovisual spontaneous emotion recognition systems, see [PAN 03, SEB 05, ZEN 07a, ZEN 08]. In this section, our aim is to highlight the new challenges accompanying this multimodal approach, such as classifying emotional expressions, i.e. classifying external manifestations of an emotion whether visual, audio, etc.

A person's emotional state is communicated through several different information streams called modalities in which information can be communicated via several features:

– Visual multimodality includes features such as facial expression, posture, gesture, movement of the head and body and direction of gaze. For example, sadness is visually communicated by the face and is also expressed as a general attitude by slowed movements, relaxed muscles and a lowered head position.

– Audio multimodality is more than a mere linguistic feature and includes amplitude, prosody, flow and frequency. For example, stress is expressed by accelerated speech.

– Physiological modality includes skin conductivity, blood pressure, heart and breathing rhythm and EEG. For example, fear can increase heart rate and skin conductivity (i.e. increased sweating).

These audio and visual multimodalities can be combined using information components (e.g. facial expressions and prosody) or can be used separately by several components (e.g. visual emotion recognition using simultaneous analysis of facial expressions and body language). However, our main objective in this chapter remains to highlight the new challenges raised by this multimodal, multi-feature approach. These challenges include:

– Whether we can evaluate the influence of each respective modality. Some academics in psychology have argued that the relative contribution of visual and audio modalities varies according to the environment and context, while others have maintained that visual modality is dominant. Other studies have shown that the influence of each modality in the recognition process depends on the emotion to be recognized. For example, experiments described in [SIL 98] have shown that anger, disgust, happiness and surprise are better recognized via the visual multimodality, while fear and sadness are better recognized via the audio modality.

– If we are able to answer this question about the respective influence of each modality, we need to develop systems that can account for the influence of each modality or feature according to their level of importance.

– There is again the problem of finding new databases that are significant in terms of spontaneity and quantity. This is because there are currently not enough data to be able to consider using *machine learning* techniques.

– We need to determine which fusion strategy to use. Traditionally, there are two strategies [WU 99]: early fusion of characteristic vectors (*feature-level fusion*) and late fusion of decisions (*decision-level fusion*). Numerous pieces of research have focused on recognizing emotions in terms of both audio and visual modalities. It is therefore very tempting to treat each modality separately and consider merging the results. This approach, however, falsely assumes that this information is independent. For example, it has been shown that there is a link between body posture and audio information and between body posture and video information [VAN 07]. This shows that a human observer can recognize an emotion better when the face and body language express the same trend. Furthermore, there are numerous decision-level fusions [KIT 98, KUN 02] and the best strategy has not yet been identified. A fusion of the results provides the advantage of being able to develop the most adapted or efficient classification algorithm for each modality or feature. This strategy is often adopted when we need to fuse two sets of data that are not necessarily synchronous. In contrast, we could consider feature-level fusion vectors, which allow us to overcome the problem of data dependency but lead to the creation of large dimension characteristic vectors. It is therefore necessary to attempt to reduce this dimensionality. The strategy we should use to reduce the dimension of data while preserving the data that will maximize the performance of the classifier has yet to be determined. This brings us back to the issue of how to account for the correlation between different modalities and components in the fusion process.

– The way we fuse different information. Audio and video modalities lead to characteristic vectors that are defined according to different time and dynamic scales that are defined on different temporal structures. For example, audio characteristics are generally evaluated on a longer portion of the signal than video characteristics, which can be evaluated on the basis of each image. How, then, do we construct a single characteristic vector?

– The way we fuse information that may not be synchronous. In other words, when we should make the decision knowing the signifiers of a person’s emotional state using various features or modalities may be delayed in manifesting themselves.

While these problems have not been resolved, there is active research into audiovisual and visual emotion recognition [SCH 09]. These examples illustrate how these problems have been approached:

– Castellano *et al*. [CAS 07] have proposed a recognition system focusing on eight emotions using facial, gestural and audio data. The analysis of the facial component has led to the construction of a vector composed of 19 coordinated points, an analysis of the gestural component using a movement vector and the analysis of the audio component with a vector of 377 characteristics. The characteristics were normalized, although no further details were provided. A Bayesian classifier is used on each of the components and two fusion modes are considered and compared: decision-level fusion or characteristic-level fusion. With characteristic-level fusion, there is a reduction in the dimension of the global vector. There was a recorded increase of 10% in the recognition rate compared with the highest rates obtained when focusing on a single component only. The best results came from classifying using characteristic-level fusion.

– Other studies such as [CHE 00] have proposed using a bimodal (audio and video) system. Video recognition was achieved by analyzing a characteristic vector on the facial movement parameters linked to AUs (six parameters for movement of the mouth, two for the eyebrows and two for the eyelids) and a neuron-based network classifier was used for classification. Audio recognition was achieved using a vector of 18 characteristics based on the energy, flow and basic frequency of the audio signal. A Bayesian classifier was then used for classification. Fusion of modalities was carried out at the decision level of each classifier using a rule set that must account for the fact that these two modalities were not always simultaneously available. When a single modality was available, the decision was taken via this modality. When two modalities were available, it was noted that analysis of the visual modality was no longer weak because movement of the mouth was linked to speech and not to expressing an emotion. As such, a preliminary decision was taken by analyzing the audio modality alone and this was therefore faced and fused with the decision made by analyzing the images taken just after the speaker turn. In this example, there is an optimal classifier for each modality and the fusion of decision-level modalities is justified by the non-simultaneity of emotional expressions via each category. This is the same overall strategy used for similar reasons by [SIL 00, ZEN 07b].

– Gunes and Piccardi [GUN 05] have focused on multicomponent recognition of visual expressions of emotion. For sequences of dialog where the subject is sitting, they analyzed emotions according to facial and gestural expressions. The extracted characteristics are measures of movement and speed linked to the mouth, hands and torso. A monocomponent classification system is therefore proposed. Note that this introduces a relatively significant difference between the two components. Classification based on gestural information is linked to a number of restricted classes because it is considered to be less informative than facial data and therefore can only provide a broader classification. In terms of the fusion of two components, the *feature-level* and *decision-level* strategies are considered and compared, showing an advantage with characteristic level fusion.

These examples demonstrate the abundance of possible strategies as well as the fact that a number of points are still under development. However, all of these studies agree on one important point: using several modalities or several components increases the performance of emotional expression recognition systems in all cases.

4.8. Conclusion

Emotions in their visual form as facial expressions play a vital role in everyday communication. This chapter has sought to provide an overview of the research into automatic visual recognition of facial expressions, an area of research that has seen a huge expansion over the past 20 years. After extensive work on automatic recognition of Ekman's expressions or the recognition of specific deliberate AUs, the current tendency is to focus on analyzing spontaneous facial expressions due to the significant differences between deliberate expressions (whose movements are stereotypical and exaggerated) and spontaneous expressions.

In terms of applications (e.g. affective human–machine interaction), the recognition system is faced with spontaneous facial expressions. Work on spontaneous facial expressions has been accompanied by new challenges, the most significant of which we have examined in this chapter:

– acquiring databases;

– identifying the classes of facial expressions to be recognized;

– estimating the intensity of facial expressions;

– the dynamics of facial expressions; and

– multimodality.

We also need to account for movement and obscuring of the head, variation between individuals in producing a facial expression, recognition of facial expressions from a non-facing perspective and accounting for context in recognition processes. It therefore remains to find an ideal system for automatically recognizing spontaneous visual expressions. Such a system will have to involve a multidisciplinary approach using both technology and social science, such as psychology and neuropsychology.

4.9. Bibliography

[ART 05] ARTANATOMY, Report, http://www.artnatomia.net/uk/index.html, 2005.

[BAR 05] BARTLETT M., LITTLEWORT G., FRANK M., LAINSCSEK C., FASEL I., MOVELLAN J., "Recognizing facial expressions: Machine learning and application to spontaneous behaviour", *Computer Vision and Pattern Recognition*, CVPR, San Diego, USA, June 2005.

[BAS 78] BASSILI J., "Facial motion in the perception of faces and of emotional expression", *Journal of Experimental Psychology*, vol. 4, pp. 373-379, 1978.

[BAT 04] BATTOCCHI A., PIANESI F., "DaFEx: Un database di espressioni facciali dinamiche", *SLI-GSCP Workshop Comunicazione Parlata e Manifestazione delle Emozioni*, November 3-December 1, Padua, Italy, 2004.

[CAS 07] CASTELLANO G., KESSOUS L., CARIDAKIS G., "Multimodal emotion recognition from expressive faces, body gestures and speech", FIORELLA DE ROSSIS R.C., (ed.), *The Doctoral Consortium of 2nd International Conference on Affective Computing and Intelligent Interaction*, Lisbon, Portugal, September 2007.

[CHE 00] CHEN L., HUANG T., "Emotional expressions in audiovisual human computer interaction", *IEEE International Conference on Multimedia and Expo*, ICME, 30 July-2 August, 2000.

[CHI 03] CHIBELUSHI C., BOUREL F., "Facial expression. Recognition: a brief tutorial overview", *On-Line, Compendium of Computer Vision*, 2003.

[COH 03] COHEN I., SEBE N., GARG A., CHEN L., HUANG T., "Facial expression recognition from video sequences: temporal and static modeling", *Computer Vision and Image Understanding*, vol. 91, no. 1, pp. 160-187, July 2003.

[COH 04] COHN J., SCHMIDT K., "The timing of facial motion in posed and spontaneous smiles", *International Journal of Wavelets, Multiresolution and Information Processing*, vol. 2, pp. 1-12, 2004.

[COW 01] COWIE R., DOUGLAS-COWIE E., TSAPATSOULIS N., VOTSIS G., KOLLIAS S., FELLENZ W., TAYLOR J., "Emotion recognition in human computer interaction", *IEEE Signal Processing Magazine*, pp. 32-80, January 2001.

[DAT 97] Yale Database, http://cvc.yale.edu/projects/yalefaces/yalefaces.html, Report, 1997.

[DAT 98] Jaffe Database, http://www.kasrl.org/jaffe.html, Report, 1998.

[DAT 00] CMU/VASC Database, http://vasc.ri.cmu.edu/idb/html/face/facial_expression /index.html, 2000.

[DAT 01] Ferret Database, http://www.itl.nist.gov/iad/humanid/feret/, Report, 2001.

[DAT 04] Dafex Database, http://tcc.itc.it/research/i3p/dafex/, Report, 2004.

[DAT 05a] HCE Database, http://www.lis.inpg.fr/pages_perso/caplier/english/emotionnelle. html.en/emotionnelle_2.html.en.html, Report, 2005.

[DAT 05b] MMI Database, http://www.mmifacedb.com/, Report, 2005.

[DAT 08] Dynemo Database, https://ligforge.imag.fr/, Report, 2008.

[EKM 72] EKMAN P., FRIESEN W., ELLSWORTH P., *Emotion in the Human Face*, Pergamon Press, New York, USA, 1972.

[EKM 78] EKMAN P., FRIESEN W., *Facial Action Coding Syste: A Technique for the Measurement of Facial Movement*, Consulting Psychologists Press, 1978.

[EKM 82] EKMAN P., FRIESEN W., ELLSWORTH P., "Research foundations", EKMAN P. (ed.), *Emotion in the Human Face*, Pergamon Press, New York, USA, 1972.

[EKM 99] EKMAN P., *The Handbook of Cognition and Emotion: Facial Expression*, John Wiley & Sons, New York, USA, 1999.

[ESS 97] ESSA I., PENTLAND A., "Coding, analysis interpretation, recognition of facial expressions", *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 757-763, July 1997.

[FAS 00] FASEL B., LUETTIN J., "Recognition of asymmetric facial action unit activities and intensities", *Int. Conf. Pattern Recognition*, June 2000.

[FAS 02] FASEL B., LUETTIN J., Automatic Facial Expression Analysis: A Survey, Research Report n°RR-9919, IDIAP, Switzerland, November 2002.

[FAS 03] FASEL B., LUETTIN J., "Automatic facial expression analysis: A survey", *Pattern Recognition*, vol. 36, no. 1, pp. 259-275, 2003.

[FRA 97] FRANK M., EKMAN P., "The ability to detect deceit generalizes across different types of high-stake lies", *Journal of Personality & Social Psychology*, vol. 72, pp. 1429-1439, 1997.

[FRA 05] FRAGOPANAGOS F., TAYLOR J., "Emotion recognition in human computer interaction", *Neural Networks*, vol. 18, pp. 389-405, 2005.

[GAR 04] GARCIA C., DELAKIS M., "Convolutional face finder: A neural architecture for fast and robust face detection", *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 26, no. 11, pp. 1408-1423, 2004.

[GHA 08] GHANEM K., CAPLIER A., STILLITTANO S., "Estimation of facial expression intensity based on the belief theory", *VISAPP 2008–International Conference on Computer Vision Theory and Applications*, Madeira, Portugal, January 2008.

[GUN 05] GUNES H., PICCARDI M., "Affect recognition from face and body: early fusion vs. late fusion", *IEEE International Conference on Man and Cybernetics*, vol. 4, pp. 3437-3443, 2005.

[HAM 07] HAMMAL Z., COUVREUR L., CAPLIER A., ROMBAUT M., "Facial expression classification: An approach based on the fusion of facial deformation using the transferable belief model", *Int. J. Approximate Reasoning*, vol. 46, pp. 542-567, 2007.

[HJE 01] HJELMÄS H., LOW B., "Face detection: a survey", *Computer Vision and Image Understanding*, vol. 83, pp. 236-274, 2001.

[KAI 01] KAISER S., WEHRLE T., "The role of facial expression in intra-individual and inter-individual emotion regulation", CAÑAMERO D. (ed.), *Emotional and Intelligent II: The Tangled Knot of Cognition*, Papers from the 2001 AAAI Fall Symposium, pp. 61-66, Menlo Park, CA, AAAI Press, 2001.

[KAN 00] KANADE T., COHN J., TIAN Y., "Comprehensive database for facial expression analysis", *4th IEEE Int. Conf. Automatic Face and Gesture Recognition*, pp. 46-53, Grenoble, France, 2000.

[KIT 98] KITTLER J., HATEF M., DUIN R., MATAS J., "On combining classifiers", *IEEE Trans. Pattern Analysis Machine Intelligence*, vol. 20, no. 3, pp. 226-239, 1998.

[KLE 90] KLECK P., MENDOLIA M., "Decoding of profile versus full-face expressions of affect", *Journal of Non Verbal Behaviour*, vol. 14, no. 1, pp. 35-49, 1990.

[KUN 02] KUNCHEVA L., "Theoretical study on six classifier fusion strategies", *IEEE Trans. Pattern Analysis Machine Intelligence,* vol. 24, no. 2, pp. 281-286, February 2002.

[LEE 03] LEE K., XU Y., "Real time estimation of facial expression intensity", *IEEE Int. Conf. on Robotics and Automation*, Taipei, China, September 2003.

[LIE 98] LIEN J., KANADE T., COHN J., LI C., "Subtly different facial expression recognition and expression intensity estimation", *IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 853-859, Santa Barbara, CA, USA, 1998.

[LIS 98] LISETTI C., RUMELHART D., "Facial expression recognition using a neural network", PRESS A. (ed.), *11th Int. Florida Artificial Intelligence Research Society Conference*, pp. 328-332, 1998.

[LOH 05] LOH M., WONG Y., WONG C., "Facial expression analysis in e-learning systems - The problems and feasibility", *Proceedings of Fifth IEEE International Conference on Advanced Learning Technologies*, Kaohsiung, Taiwan, IEEE Computer Society, pp. 442-446, July 2005.

[LYO 98] LYONS M., AKAMATSU S., KAMACHI M., GYOBA J., "Coding facial expressions with Gabor wavelets", *Third IEEE International Conference on Automatic Face and Gesture Recognition*, Nara, Japan, pp. 200-205, 1998.

[MAC 04] MACHINE PERCEPTION TOOLBOX F. D. A., Report, http://mplab.ucsd.edu/grants/project1/freesoftware/MPTWebSite/introductionframe.html, 2004.

[MOR 02] MORIYAMA T., KANADE T., COHN J., XIAO J., AMBADAR Z., GAO J., IMAMURA H., "Automatic recognition of eye blinking in spontaneously occurring behavior", *16th International Conference on Pattern Recognition*, Quebec City, Canada, 2002.

[MUR 08] MURPHY-CHUTORIAN E., TRIVEDI M., "Head pose estimation in computervision: A survey", *IEEE Trans. Pattern Analysis Machine Intelligence*, vol. 31, no. 4, pp. 1-8, 2008.

[NIE 07] NIEWIADOMSKI R., PELACHAUD C., "Model of facial expressions management for an embodied conversational agent, *Proceedings of the 2nd International Conference on Affective Computing and Intelligent Interaction*, Lisbon, Portugal, pp. 12-23, 2007.

[NOE 04] NOE H., Report, http://emotion-research.net, 2004.

[OTS 97] OTSUKA T., OHYA J., "Recognizing multiple persons' facial expressions using HMM based on automatic extraction of significant frames from image sequences", *IEEE Int. Conf. Image Processing*, vol. 2, pp. 546-549, 1997.

[PAN 00a] PANTIC M., ROTHKRANTZ L., "Automatic analysis of facial expressions: the state of the art", *IEEE Trans. Pattern Analysis Machine Intelligence*, vol. 22, no. 12, pp. 1424-1445, 2000.

[PAN 00b] PANTIC M., ROTHKRANTZ L., "Expert system for automatic analysis of facial expression", *Image and Vision Computing*, vol. 18, no. 11, pp. 881-905, 2000.

[PAN 03] PANTIC M., ROTHKRANTZ L., "Toward an affect sensitive multimodal human computer interaction", *Proceedings of the IEEE*, vol. 91, pp. 1370-1390, September 2003.

[PAN 05] PANTIC M., VALSTAR M., RADEMAKER R., MAAT L., "Web-based database for facial expression analysis", *IEEE Int'l Conf. Multmedia and Expo (ICME'05)*, Amsterdam, The Netherlands, July 2005.

[PAN 06] PANTIC M., PATRAS I., "Dynamics of facial expression: Recognition of facial actions and their temporal segments from face profile image sequences", *IEEE Trans. Systems, Man Cybernetics – Part B: Cybernetics*, vol. 36, no. 2, pp. 443-449, April 2006.

[PHI 00] PHILLIPS P. J., MOON H., RAUSS H. M., RIZVI S., "The FERET evaluation methodology for face recognition algorithms", *IEEE Trans Pattern Analysis Machine Intelligence*, vol. 22, no. 10, pp. 1090-1104, October 2000.

[PIC 01] PICARD R., VYSAS E., HEALY J., "Towards machine emotional intelligence: analysis of affective physiological state", *IEEE Trans. Pattern Analysis Machine Intelligence*, vol. 23, no. 10, pp. 1175-1191, 2001.

[ROI 04] ROISMAN G., TSAI J., CHIANG K., "The emotional integration of childhood experience: Physiological, facial expressive and self-reported emotional response during the adult attachment interview", *Developmental Psychology*, vol. 40, no. 5, pp. 776-789, 2004.

[RUS 03] RUSSELL J., BACHOROWSKI J., DOLS J.F., "Facial and vocal expressions of emotion", *Ann. Rev. Psychol.*, vol. 54, pp. 329-349, 2003.

[SCH 99] SCHERER K., "Appraisal theory", in: DALGLEISH T. and POWER M. (eds), *Handbook of Cognition and Emotion*, Wiley, Chichester, 1999.

[SCH 07] SCHERER K., ELLGRING H., "Multimodal expression of emotion: Affect programs or componential appraisal patterns?", *Emotion*, vol. 7, no. 1, pp. 158-171, 2007.

[SCH 03] SCHMIDT K., COHN J., TIAN Y., "Signal characteristics of spontaneous facial expressions: automatic movement in solitary and social smiles", *Biological Psychology*, vol. 65, no. 1, pp. 49-66, 2003.

[SCH 09] SCHULLER B., "Emotion recognition in the next generation: an overview and recent development", *Proceedings 10th Annual Conference of the International Speech Communication Association*, Brighton, UK, September 2009.

[SEB 06] SEBE N., COHEN I., GEVERS T., HUANG T., "Multimodal approaches for emotion recognition: A survey", *Internet Imaging Conference*, vol. 5670, pp. 56-67, 2006.

[SEB 07] SEBE N., LEW M., SUN Y., COHEN I., GEVERS T., HUANG T., "Authentic facial expression analysis", *Image and Vision Computing*, vol. 25, no. 12, pp. 1856-1863, 2007.

[SHA 06] SHAN C., GONG S., MCOWAN P., "Dynamic facial recognition using a Baesian temporal manifold model", *BMVC*, Edinburgh, UK, September 2006.

[SIL 98] SILVA L.D., MIYASATO T., NAKATSU R., "Use of multimodal information in facial emotion recognition", *IEICE Trans. Inf Syst*, vol. E81-D, no. 1, pp. 397-401, 1998.

[SIL 00] SILVA L.D., CHI N.P., "Bimodal emotion recognition", *Fourth IEEE International Conference on Automatic Face and Gesture Recognition*, 26-30 March, Grenoble, France, 2000.

[STO 07] DEN STOCK J.V., RIGHART R., GELDER B.D., "Body expressions influence recognition of emotions in the face and voice", *Emotion*, vol. 7, pp. 487-494, 2007.

[TIA 01] TIAN Y., KANADE T., J.COHN, "Recognizing action units for facial expression analysis", *IEEE Trans. Pattern Analysis Machine Intelligence*, vol. 23, no. 2, pp. 97-115, 2001.

[TIA 03] TIAN Y., KANADE T., COHN J., "Facial Expression Analysis", in: LI S.Z. and JAINS A.K. (eds), *Handbook of Face Recognition*, http://www.ri.cmu.edu/pub_files/pub4/tian_ying_li_2003_1/tian_ying_li_2003_1.pdf, 2003.

[VIO 04] VIOLA P., JONES J., "Robust real time face detection", *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137-154, May 2004.

[WU 99] WU L., OVIATT S., COHEN P., "Multimodal integration – A statistical view", *IEEE Trans. Multimedia*, vol. 1, no. 4, pp. 334-341, 1999.

[YAN 02] YANG M., KRIEGMAN D., AHUJA N., "Detecting face in images: a survey", *IEEE Trans. Pattern Analysis Machine Intelligence*, vol. 24, no. 1, pp. 38-58, January 2002.

[YEA 04] YEASIN M., BULLOT B., SHARMA R., "From facial expression understanding from image sequences", *IEEE Conf. Computer Vision Pattern Recognition*, Washington DC, July 2004.

[ZEN 06] ZENG Z., FU Y., ROISMAN G., WEN Z., HU Y., HUANG T., "Spontaneous emotional facial expression detection", *Journal of Multimedia*, vol. 12, no. 5, pp. 1856-1863, August 2006.

[ZEN 07] ZENG Z., TU J., LIU M., HUANG T., PIANFETTI B., ROTH D., LEVINSON S., "Audio visual affect recognition", *IEEE Trans. Multimedia*, vol. 9, no. 2, pp. 424-428, 2007.

[ZEN 08] ZENG Z., TU J., PIANFETTI B., HUANG T., "Audio-visual affective expression recognition through multi-stream fused HMM", *IEEE Trans. Multimedia*, vol. 10, no. 4, pp. 570-577, 2008.

[ZEN 09] ZENG Z., PANTIC M., ROISMAN G., HUANG T., "A survey of affect recognition methods: Audio, visual and spontaneous expressions", *IEEE Trans. Pattern Analysis Machine Intelligence*, vol. 31, no. 1, pp. 39-58, 2009.

[ZHA 05] ZHANG T., JI Q., "Active and dynamic information fusion for facial expression understanding from image sequence", *IEEE Trans. Pattern Analysis Machine* Intelligence, vol. 27, no. 5, pp. 699-714, May 2005.

[ZHA 06] ZHAN C., LI W., OGUNBONA P., SAFAEI F., "Facial expression recognition for multiplayer online games", *Proceedings of the 3rd Australasian Conference on Interactive Entertainment*, vol. 207, pp. 52-58, ACM International Conference Proceeding Series, Perth, Australia, 2006.

Chapter 5

Recognition of Acoustic Emotion

5.1. Introduction

In the past 15 years, research into emotions in speech has moved beyond merely analyzing vocal manifestations of emotional states to focusing on developing automatic emotion classification systems. This has arisen as a result of the emergence of affective sciences or *affective computing* [PIC 97] and their potentially numerous practical applications.

Many instances of emotion voice recognition involve human–machine interactions. A good example of this might be dialog systems where being able to identify a user's emotional state will allow us to adapt our dialog strategy. In a call center context, for example, if the user shows signs of irritation or frustration with the automatic system, then a strategy may be triggered, sending the caller to a human operator [DEV 05, LEE 02]. However, emotion recognition is not only applicable to dialog systems. For example, some medical academic research [IST 03] has focused on emotion recognition as a means of helping older or hospitalized patients.

Similarly, this technology can be used for security, with applications such as crisis management or audio surveillance. In crisis management, for example, work has been undertaken to monitor victims' and rescue workers' emotions using collaborative *search and rescue* robots [LOO 07]. These robots are vital for crisis management, particularly in environments considered too dangerous or inaccessible to emergency crews. In the field of security, diagnostic machines have been used to

Chapter written by Chloé CLAVEL and Gaël RICHARD.

Emotion-Oriented Systems Edited by Catherine Pelachaud

aid surveillance. The majority of automatic surveillance systems currently rely on video to detect and analyze abnormal situations [NGH 07]. Some research has begun to focus on integrating audio emotion-related information with video information [CLA 08, HEN 07].

This chapter aims to describe the various stages required in developing emotion-recognition systems based on acoustic indicators[1]. The first stage consists of collecting and annotating data. This is examined in detail in Chapter 3. This chapter focuses on the second stage, which consists of developing emotion recognition "systems" and writing acoustic descriptors that can characterize and differentiate between different emotional manifestations as well as record the classification techniques used (learning and decisions). We will conclude by examining some of the different criteria for evaluating these systems and their performance.

5.2. Principles of automatic emotion-recognition systems

An automatic emotion-recognition system normally relies on four main phases (see Figure 5.1):

1) *Extraction of acoustic descriptors*: this is usually carried out by an analysis module that converts the speech signal into a sequence of acoustic vectors containing the values of the selected descriptors (or parameters). The aim of this stage is to establish a compact representation of the speech signal's main acoustic characteristics that we need to identify different emotional manifestations. The intensity of the speech signal and the formants' central frequency are two examples of these acoustic descriptors.

2) During the *learning* phase, several acoustic vectors corresponding to sounds of the same class are used to create a prototypical model of this class. The representative can, for example, be obtained like barycenters (or centroides) by acoustic vectors that are characteristic of the class in question. A model allows the best characterization of the distributions of the acoustic vectors' values for each class (each class corresponds to an emotional class). This representative or model is normally obtained using a "learning" database that has been annotated previously (see Chapter 3).

3) During the *classification* phase, the vocal signal's acoustic vectors being analyzed are compared to the representatives or models of each class. After this phase, a probability of belonging to each class can be obtained for each acoustic vector.

---

1 Some systems rely on a combination of linguistic and acoustic information [SCH 04], although this chapter will focus on acoustic emotion recognition only.

4) The last phase is the *decision* phase, which involves the allocation of a class to a speech segment using the acoustic vectors' probabilities of belonging.

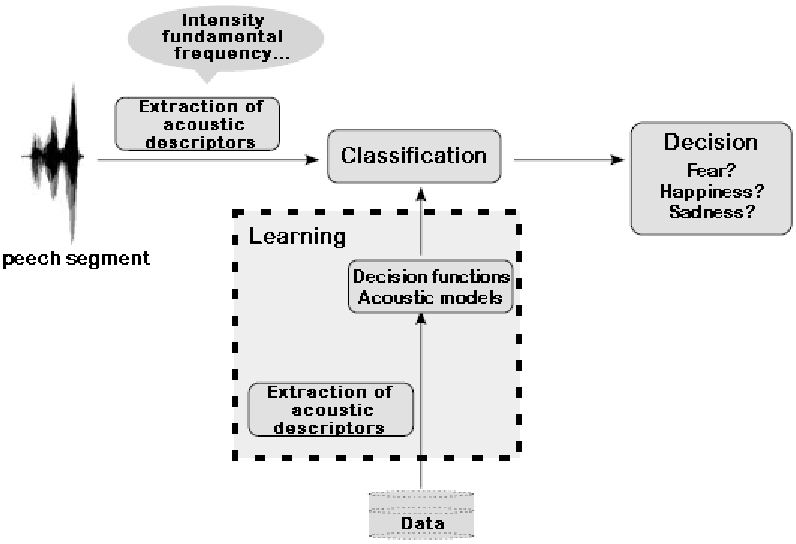


Figure 5.1. *The principle of an emotion-recognition system*

5.3. Acoustic descriptors

The physical phenomenon of producing speech is traditionally represented by a source-filter model [FAN 60]. This simple model remains a reference point for identifying the main acoustic characteristics of a vocal signal. It is therefore in the framework of this production that the acoustic parameters (or descriptors) used in automatic emotion recognition will be described.

A large proportion of acoustic descriptors used for characterizing emotions are intended for modeling alterations in acoustic signals linked to physical changes in the glottis. This is particularly the case with prosodic and voice quality parameters (i.e. breathy, squeaky, harsh, tense). In this section we will provide an overview of how these descriptors are used in emotional speech research. We will also examine some of the parameters (formant and cepstral coefficient) that convey changes in the acoustic signal as a result of modifications in the vocal tract (e.g. part of the source/filter model).

The physical changes accompanying certain emotion states have a strong influence on how a speaker produces an utterance. For example, fear is often accompanied by typical physical changes, such as increased pulse and blood pressure or dryness of the mouth. This is in turn manifested by a deeper, less intense and slower voice. This contrasts with boredom and sadness that generally correlate with a slower heart rhythm and a more serious, less intense and slower voice [PIC 97].

This modification of emotional vocal expression acoustic descriptors resulting from bodily changes has also been studied by Scherer in relation to various evaluations of external stimuli (*appraisal theory* [SCH 03]). According to this theory, it is these various evaluations that determine our emotional behavior based on the following parameters:

– Is the situation new? (*Novelty check*).

– Is the situation pleasant? (*Intrinsic pleasantness check*).

– Does the situation favor the goals of the individual? (*Goal/need significance check*).

– Does the individual have the necessary resources to cope with this situation? (*Coping potential check*).

– Is the situation compatible with the individual's socio-cultural norms? (*Norm/self-compatibility check*).

The intensity and pitch of the voice (or voice quality descriptors) are called *high-level* descriptors of emotional manifestations because they provide an increased level of interpretations. These *high-level* descriptors are commonly used in synthesizing and analyzing emotional speech. However, when attempting to recognize emotions using learning algorithms (as Oudeyer [OUD 03] and Clavel *et al*. [CLA 08] have done), the descriptors used are often more numerous, including more *low-level* descriptors where it is difficult to identify a perceptual correlate (see Chapter 8). If the acoustic information conveyed by such descriptors is less explicit than that of prosodic descriptors, for example, this information is nevertheless still useful and is integrated into acoustic models constructed by learning algorithms.

5.3.1. *Voiced versus unvoiced content*

Studies into acoustic speech emotional descriptors have focused on voice content that has been found, particularly with prosodic descriptors, to carry a large amount of characteristic information about the locutor's emotions. Depending on the emotions studied, however, the unvoiced content of an utterance can also convey important information about emotional manifestations. This is the case with intense emotions that are often accompanied by strong physical changes, such as clenching, trembling and increased heart rate. Physiological and/or physical activity accompanying oral production leads to non-verbal signals, such as heavier breathing and exclamations. These manifestations have effects on both voiced and unvoiced content. Clavel *et al*. [CLA 08] have therefore studied the impact of emotional manifestations on acoustic content using both voiced and unvoiced data.

5.3.2. *A temporal unit for emotional analysis*

More important than the choice of acoustic descriptors for characterizing emotional content is the essential question of which temporal duration we should use to measure these descriptors? Each descriptor's behavior depends on the time window in which it is considered. There are two approaches to this issue. The first, as cited by Vlasenko *et al*. [VID 05], uses statistical descriptors proposing a global model of each description (e.g. average, minimum, etc.) on different temporal durations such as syllables, words or phrases. The second approach relies on extracting acoustic descriptors from each analysis window [AMI 96] or is based on a combination of modeling the window with an overall model of the descriptors ([VLA 07] (windows of 25 milliseconds, one speaker turn and [CLA 08] windows of 40 milliseconds, voiced and unvoiced trajectories[2], identical segments[3]).

Modeling in terms of analysis windows and trajectories means we do not make assumptions about the utterance's structure and do not require knowledge of the linguistic content linked to the speech signifier. In contrast, global modeling relies on a precise segmentation of the speech, which can be obtained using an automatic speech-recognition system. However, this kind of analysis cannot always cope with features such as co-articulation, non-modal voices and non-verbal manifestations (e.g. exclamations, etc.) that arise during spontaneous emotional manifestations and can potentially pose problems depending on the type of data in question.

5.3.3. *Prosodic descriptors*

Prosodic descriptors, originally used for voice recognition and synthesis (voice pitch and intensity, length of syllables) are used for structuring the flow of speech. Prosodic descriptors are often used to identify specific segmental elements and can also provide information on an utterance's syntactic structure. In summary, they make the synthetic signal both more natural and more intelligible by indicating the main themes in the sentence.

2 A voiced trajectory (and unvoiced respectively) is a succession of voiced (or unvoiced) adjacent windows.

3 A segment is defined here as a portion of a speaker turn with uniform emotional content.

Prosodic descriptors allow us to model accents, rhythm, intonation and melody and are highly useful for modeling a locutor's emotional state. Initial studies into emotional speech have focused on analyzing prosody using descriptors such as pitch (or fundamental frequency) and voice intensity [AMI 96, DEL 96] corresponding to a model of the supra-segmental[4] changes in the acoustic signal produced by physical changes in the glottis [SCH 98].

5.3.3.1. *Fundamental frequency (pitch)*

Fundamental frequency relates to the frequency of vibrations in the vocal cords when producing voiced sounds. It is directly linked to voice pitch (high, low). The voiced sections of speech have a pseudo-periodic structure and the signal is generally modeled on these sections as a combination of the periodic signal (T) and white noise. Fundamental frequency is therefore the opposite of the period T, $F_0 = \frac{1}{T}$.

There are several ways of estimating fundamental frequency: temporal methods (autocorrelation and *average square difference function*, ADSF), maximum likelihood, cepstral analysis based methods [HES 84, KLA 08], etc. One of the most popular methods, used most notably by the Praat software [BOE 05], is a temporal approach that consists of examining similarities between different versions of the same recorded signal. These similarities are evaluated using autocorrelation, which can be expressed as follows:

$$r_s(m) = \begin{cases} \frac{\sum_{n=0}^{N-1-m} s(n)s(n+m)}{\sqrt{\sum_{n=0}^{N-1-m} s(n)^2}\sqrt{\sum_{n=0}^{N-1-m} s(n+m)^2}} & \textit{if } m \geq 0 \\ r_s(-m) & \textit{otherwise} \end{cases}$$

The period T is determined by finding the value of m for which $r_s(m)$ is maximal in a previously selected interval $m \in [T_{min}: T_{max}]$. The autocorrelation function also gives an estimation of the speech signal's "strength of voicing"; the closer r_s (T) is to 1 (the maximum possible value), the closer the signal is to a signal of period T. Figure 5.2, adapted from [CLA 07b], illustrates the behavior of fundamental frequency on two examples containing different degrees of emotional intensity[5].

These two segments correspond to the same word "Josh" produced by the same locutor during an ongoing situation. With the first example, the locutor has just realized that her friend Josh has disappeared and calls him for the first time. The acoustic indications of fear in this example are difficult to perceive and there is an

4 This varies according to emphasis and intonation, also called the prosodic level.
5 The intensity dimension is derived from the activation dimension defined by [OSG 75] as a state of excitement ranging from calm to highly excited (see Chapter 3).

increase in the question's characteristic fundamental frequency. In the second example, the locator – after having called a number of times without any response – shouts the name of her friend. In this example, the contour of the fundamental frequency is much less smooth, containing a number of jumps with a higher average fundamental frequency.

5.3.3.2. *Intensity*

Intensity (or energy) is measure of the force of the voice (weak or strong). The intensity (in dB) is generally, as with Praat [BOE 05], calculated on a signal section of N length where ω is an analysis window, as follows:

$$I = 10\log(\sum_{n=1}^{N} s^2(n)w(n))$$

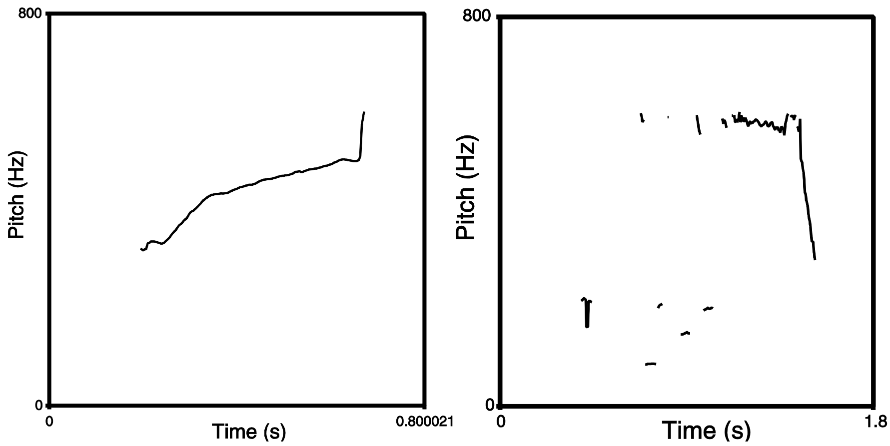


Figure 5.2. *Example of the fundamental frequency estimated using autocorrelation from the Praat software on the example "Josh?". Left: corresponding to fear with weak intensity (worry). Right: The example "Joooosh!" corresponds to fear with strong intensity (panic)*

Figure 5.3 illustrates intensity on these same examples. With panic, the average intensity is higher with more modulations.

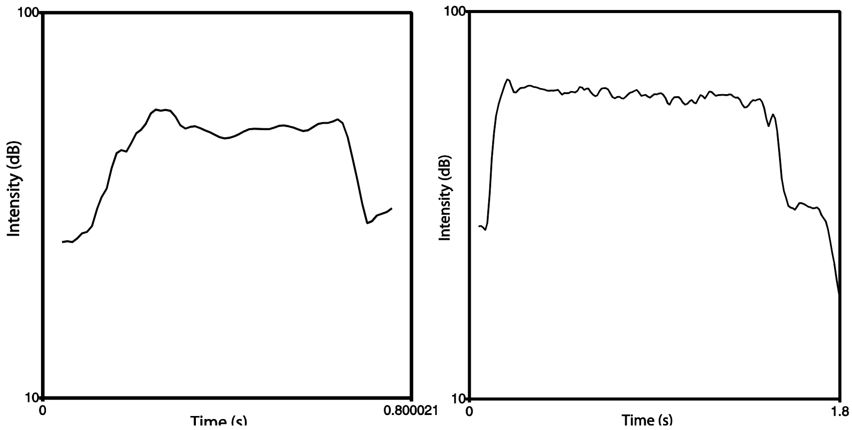


Figure 5.3. *Example of an intensity contour in Praat calculated on the example "Josh" (left) corresponding to fear with weak intensity (worry) and "Jooooosh!" (right) corresponding to fear with strong intensity (panic)*

5.3.3.3. *Rhythm descriptors*

The third classic descriptor of prosody is a rhythmic descriptor linked to speech flow [KOZ 65] and is measured by the number of vocal units in one unit of time, e.g. the number of syllables or phonemes uttered per minute. However, this can only be calculated if we have segmentation of the speech signal into syllables or phonemes (which involves using a speech-recognition system). This segmentation can be complex, depending on the recording conditions. An alternative [CLA 08] might be to calculate the duration of the voiced trajectory (consecutive voiced windows). This descriptor allows us to characterize the differences in the rate of speech flow with a large proportion of long-voiced trajectories being a sign of a slower rate.

5.3.4. *Voice quality descriptors*

Variations in voice quality, while often associated with pathological voices, are also the result of physiological changes caused by emotional changes in the locutor. Voice quality descriptors are increasingly being used to characterize emotional variations.

5.3.4.1. *Normalized amplitude quotient*

Some research into emotions in speech has examined the strong potential interactions between glottal configuration and emotional expression[6]. Campbell and Mokhtari, for example, have developed an algorithm for measuring voice quality descriptors by considering physiological changes in emotional speech [CAM 03]. Voice quality is measured using the normalized amplitude quotient (NAQ), which is an indicator of breath levels; the higher the NAQ, the more breathing there is in the voice.

In this chapter, the role of descriptors modeling the glottal source for automatic emotion recognition is clearly highlighted by data from recordings of a Japanese speaker's daily interactions. In these recordings, the NAQ performance is analyzed in terms of type of speaker, degree of politeness and "effort" made by the speaker to communicate. NAQ has also been used by Audibert *et al*. [AUD 04] to characterize emotions.

By modeling the glottal flow of a triangular pulsation during the open phase and using a zero signal during the closing phase, the amplitude quotient (AQ) corresponding to the glottal closing time can be written as follows:

$$AQ = \frac{fac}{dpeak}$$

In this equation, *fac* corresponds to the maximum amplitude of the glottal flux and *dpeak* to the gradual decrease in glottal flow. The AQ here depends on fundamental frequency. The NAQ corresponds to the amplitude quotient, which is decorrelated from the fundamental frequency (F_0):

$$NAQ = log(AQ) + log(F_0)$$

Further details on the calculation of this coefficient are described in [ALK 02].

5.3.4.2. *Frequency modulation (jitter)*

Jitter is described as a deviation in fundamental frequency (see section 5.3.3) and corresponds to a measure of sounds on the frequency (this is also sometimes referred to as oscillations around the fundamental frequency). In music this, along with shimmer or amplitude modulations, is used to model the striking, holding and release of notes [VER 03]. It is also used to study pathological voices (creaky

6 However, this research requires "ideal" recording conditions, e.g. short distance between mouth and microphone.

voices) and has notably been used by France *et al*. [FRA 03] to characterize depressive states and by Clavel *et al*. [CLA 07a] to model fear-like emotions. Figure 5.2 shows strong modulations in fundamental frequency in the cry of “Joooosh!”. A standard way of estimating jitter is to calculate the following ratio:

$$Jitter = \frac{\sum_{n=2}^{N-1} |2T_n - T_{n-1} - T_{n+1}|}{\sum_{n=1}^{N} T_n}$$

In this equation, T_n is the n^{th} interval of a signal period and N is the number of intervals.

5.3.4.3. *Amplitude modulation (shimmer)*

Often used in combination with jitter, shimmer models the amplitude modulation of vocal signals. As with Jitter, Figure 5.3 demonstrates strong modulations in intensity with the utterance of “Jooosh!”. An evaluation of shimmer, similarly to jitter, can be obtained by calculating the following ratio on a portion of the signal of length N:

$$Shimmer = \frac{\sum_{n=2}^{N-1} |2A_n - A_{n-1} - A_{n+1}|}{\sum_{n=1}^{N} A_n}$$

In this calculation, the average amplitude is calculated on the interval T_n and N is the number of intervals.

5.3.4.4. *Rate of unvoiced windows*

The rate of unvoiced windows corresponds to the proportion of windows considered to be unvoiced in a section of the speech signal. In practice, this parameter can be calculated as follows: a window is considered to be unvoiced if the “voicing force” (autocorrelation score, see section 5.3.3) is lower than a given threshold (typically fixed in the interval between 0.45 and 0.6).

5.3.4.5. *Harmonic to noise ration*

The aim of this descriptor is to find an indicator of breath levels in the voice by measuring the harmonic to noise ratio in the speech signal (HNR). For example, the algorithm developed by Yegnanarayana *et al*. [YEG 98] (also used by Clavel [CLA 07a] to characterize fear-like emotions) relies on estimating the degree of replacements of harmonics with noise, i.e. the ratio between the acoustic energy of harmonic components and the acoustic energy of noise components.

This algorithm relies on splitting the vocal signal into two separate components. The first component is periodic, resulting from the quasi-periodic vibration of vocal chords. The second is aperiodic and regroups the set of noises in speech (fricative pink noise, plosive noise, whistling). The HNR is adapted according to the estimation of the noise contribution by considering both the noise from oscillations in the vocal cords (inharmonicity of the glottal waveform) and additional noise that can be expressed as the following formula:

$$PAP = 10\log\left(\frac{Energy_{periodic}}{Energy_{aperiodic}}\right)$$

5.3.5. *Cepstral and spectral descriptors*

A large number of low-level descriptors rely on analyzing the signal spectrum and including spectral descriptors (formant descriptors, Bark band energy, spectral centroides) and cepstral descriptors (Mel-frequency cepstral coefficients or MFCC).

5.3.5.1. *Formant parameters*

Acoustic analysis of vocal tracts has illustrated resonances in the voice. These resonances – which vary depending on the position of different articulators (tongue, lips, etc.) – are called formants. They are notably characterized by a central frequency (or formant frequency) and a bandwidth of -6dB at their peak. In terms of voiced content, the position of the first two formants is independent of pitch (fundamental frequency) and is characteristic of a specific vowel. Formants and their bandwidths are also significant for unvoiced windows because they provide a model of the vocal tract and are thereby voice quality descriptors. In practice, the initial formants can be estimated similarly as in Linear Predictive Coding (LPC).

Formant parameters are largely used to describe emotions, especially for acoustic characterization of fear-like emotions [CLA 08, FRA 03, KIE 00]. The importance of these parameters can be explained partly by their insight into voice quality descriptors and partly by the relation between the first formant and the degree of vowel opening, which can be said to be greater with extreme emotions.

5.3.5.2. *Mel-frequency cepstral coefficients*

MFCC parameterization is widely used in speech processing, whether for automatic speech recognition, speaker recognition or language recognition. MFCCs are also used for emotion recognition [KWO 03, SHA 03].

MFCCs belong to the cepstral family of descriptors, which are based on a cepstral representation of a signal. The cepstrum has, to a certain extent, the advantage of allowing us to separate the respective contributions of the source and the vocal tract.

MFCCs are obtained using the Mel frequency scale, a non-linear frequency scale accounting for the specificities of the human ear for calculating the cepstrum. The Mel scale corresponds to an approximation of psychological feeling of a sound's pitch and accounts for greater frequency selectivity in the ear with low frequencies. For each frame of the signal, the amplitude spectrum *S(k)* is integrated by Mel bands to obtain a modified amplitude spectrum $\tilde{a}_m$ $m = 1 \ldots M_b$, representing the Mel *m* band amplitude. As such, triangular filters with a constant bandwidth that are evenly spaced on the Mel scale are currently used. Finally, the MFCC coefficients are obtained by carrying out a discrete cosine transform of the coefficient logarithm previously obtained:

$$\tilde{c}(\tau) = \sum_{m=1}^{M_b} \log(\tilde{a}(m)) \cos[\tau(m - \frac{1}{2})\frac{\pi}{M_b}]$$

In this equation, M_b is the number of triangular filters. The contribution of the vocal tract is mainly found in the initial cepstral coefficients, which shows that in practice only the first coefficients are kept (typically between 12 and 15).

5.3.5.3. *Bark band energy*

Bark band energy relies on another currently used perceptive scale – a pitch scale where the unit used is called a Bark. The signal spectrum for this descriptor is split into different frequency bands, which are determined by this scale. The Bark scale is based on critical bands that are sensed by the ear [ZWI 57]. There are several formulas converting Bark frequency scale into Hertz, as proposed by Serkey and Hanson [SER 84]:

$$f_{Bark} = 6 \operatorname{arcsin}(\frac{f_{Hz}}{600})$$

This can be calculated by splitting the signal into different frequency bands with a width equal to 1 or 2 Barks [HER 04]. For each band, i, energy is expressed as follows, with *m* being the number of the band in question:

$$EBB(m) = \sum_{f_{inf}(m)}^{f_{sup}(m)} |S(f)|^2$$

Energy in specific band frequencies is a descriptor that has been used to explain different types of voice (e.g. noisy, aggressive, sensual) [EHR 04] and for studying acoustic manifestations of emotions, such as anger, happiness, fear, boredom and sadness [KIE 00], laughter and cheers [CAI 03].

5.3.5.4. *Spectral centroide*

Spectral centroide, or spectral balance, can be used as a descriptor of voice "timbre" [HER 04]. This descriptor has also been calculated on unvoiced fricatives and allows us to measure the degree of constriction [KIE 00]. More generally used to describe a sensation of auditory brightness, it has also been used for the characterization and automatic recognition of musical instruments [ESS 05, MCA 99]. It corresponds to the spectral moment of the first order and is calculated as follows, with k as the band frequency number and a_k being the energy of this band:

$$C_s = \frac{\sum_k kEBB(k)}{\sum_k EBB(k)}$$

5.4. Automatic emotion classification

Automatic emotion classification systems rely on "learning" methods due to their ability to learn (or characterize) the acoustic properties of each type of emotion (using a sufficient quantity of data). There are two types of classification, both supervised and unsupervised. During a supervised classification, the class of each object (represented by a label) is provided to the learning program at the same time as the data. In unsupervised classification, emotional classes are automatically identified according to the data structure. Automatic emotion-classification systems are mainly supervised where the classes are determined by the type of application desired. There are numerous classification techniques as examined by Duda and Hart [DUD 73]. We will now examine some classification algorithms by highlighting the different problems inherent to emotions.

5.4.1. *Choosing descriptors*

5.4.1.1. *Normalizing descriptors*

The range of values that a descriptor can take varies strongly from one descriptor to another. For example, if the rate of unvoiced windows varies between 0 and 1, the fundamental frequency takes values in the order of several hundred. This heterogeneity of values can have consequences for the behavior of descriptor space-reduction algorithms, such as a selection of the most discriminatory descriptors and learning algorithms. Higher value descriptors may have more weight than those with weaker values. Normalization techniques can therefore be designed to avoid this bias. Some of these have already been successfully used for emotion recognition including:

– the “min–max” normalization technique [CLA 08];

– sigma–mu normalization;

– normalization by genre, speaker, phoneme [DEV 05].

5.4.1.2. *Reduction in data representation space*

In theory, increasing the number of descriptors could improve system performance. In reality, however, the use of too many descriptors, apart from the problem of complexity caused by the high dimension of data-representation space, can lead to decreased performance [DUD 73]. This stage of reducing the dimension of data-representation space prior to the learning stage and selection of classification system is therefore essential if a large number of parameters have been chosen (see [GUY 03] for example).

There are two ways of reducing the space of descriptors:

– Projecting the data representation space on a smaller dimension space (e.g. principal component analysis, discriminant analysis).

– Selecting a subset of the most discriminant descriptors (e.g. the Fisher selection algorithm and genetic algorithm). This option has the advantage of being able to directly extract the descriptors that are the most relevant for the test stage while projection methods require the prior calculation of the set of descriptors from the sample being tested.

The Fisher selection algorithm is a method whose simplicity and efficiency have been demonstrated many times (see [ESS 05] for an example of how this has been used in a musical instrument-recognition system). It is derived from Fisher discriminant analysis (for a more detailed description, see [DUD 73]. This consists of maximizing the Fisher discriminant ratio (FDR) between interclass and intraclass dispersion for each descriptor separately:

$$FDR_{d_i} = \frac{\left(\mu_{i,class1} - \mu_{i,class2}\right)^2}{\sigma^2_{i,class1} + \sigma^2_{i,class2}}$$

In this equation, $\mu_{iclass\ 1}$ and $\mu_{iclass\ 2}$ are the average values corresponding to the d_i descriptors for each class and the corresponding variances $\sigma^2_{i,\ class1}$ and $\sigma^2_{i,\ class2}$.

Multiclass selection by the Fisher algorithm, split into two phases, is also possible[7]:

1) For each descriptor i and for each class k, an intermediary score is allocated using the data from each of these classes according to the following formula:

$$s_i^k = \sum_{l=1}^{K} FDR_{l,k}(i)$$

In this formula, k is the number of classes and $FDR_{\ l,k}$ is the Fisher criterion calculated for the two classes l and k:

$$FDR_{l,k}(i) = \frac{(\mu_{i,l} - \mu_{i,k})^2}{\sigma^2_{i,l} + \sigma^2_{i,k}}$$

Here, $\mu_{i,l}$ and $\mu_{i,k}$ are the averages of the descriptor i on the data linked to the classes l and k and the variances $\sigma^2_{i,l}$ and $\sigma^2_{i,k}$.

2) The scores $s^k_{i1\leq\ i\leq I\ ;1\leq\ k\leq K}$ are then sorted in descending order and the N first distinct descriptors linked to the highest scores are selected.

This algorithm allows us to select the most discriminative descriptors, although this does not account for potential correlations between them. In order to avoid the final set of selected descriptors having too high a redundancy, the Fisher algorithm can be used in two stages as done in Clavel *et al*. [CLA 07a]:

(i) An initial selection is carried out separately for each family of descriptors (prosodic, voice quality and spectral).

(ii) The second selection is carried out by applying the Fisher algorithm again on the set of selected descriptors in the previous stage.

---

7 http://www.kyb.mpg.de/bs/people/spider/. This is also the option chosen by the Spider toolbox.

There are, however, more sophisticated descriptor selection methods that allow us to compare subsets of descriptors by eliminating redundant descriptors. This is especially true of the inertia ratio maximization using feature space projection algorithm [PEE 03] or genetic algorithms that carry out a systematic search of subsets of possible descriptors.

5.4.2. *Learning algorithms*

The use of classification methods in the field of emotions is an emerging area of research. In this section we will examine some of the classification methods used in this area in further detail.

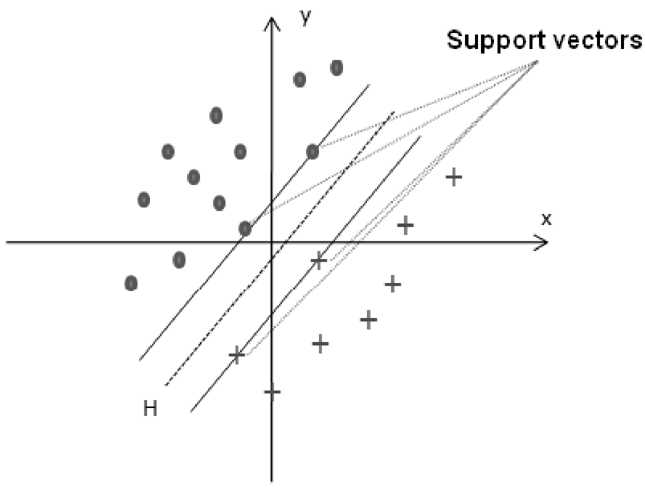


Figure 5.4. *Hyperplane separator and support vector in a two-dimensional space*

5.4.2.1. *Separators with vast margin*

Separators with vast margin (SVM) were introduced by Vapnik in 1995 [VAP 95]. This method is therefore a relatively recent alternative to classification. While this method was initially designed to solve problems of classifying between two classes, it can now make multiclass generalizations [HSU 02]. The principle of SVMs can be described as follows: for two given classes of examples, the examples are separated from each class using a hyperplane by maximizing the distance of the learning examples from the hyperplane. The nearest points of the hyperplane, which are the only ones used to determine the hyperplane, are called support vectors (see Figure 5.4.). The distance that we want to maximize is the minimum distance between the hyperplane and the learning examples. This distance is called the "margin".

There are two types of SVM models: linearly separable and non-linearly separable models. In order to overcome the disadvantages of non-linear separable models, the aim of SVMs is to transform the data space in order to change a non-linear separation problem into a linear separation problem in a larger dimensioned redescription space. This non-linear transformation is carried out using a function called a kernel function. Apart from the linear kernel *(x, y)* = *x.y*, the most commonly used kernels are the polynomial kernels $K(x,y) = (1 + x.y)^d$ (polynomial of degree d) and Gaussian polynomials $k(x,y) = \exp\left(\frac{\|x-y\|^2}{2\sigma^2}\right)$

In this case, the variance σ of the kernel regulates the precision of the separating hyperplane. A small σ signifies that the separating hyperplane is very close to the support vectors and risks overtraining. A very large σ corresponds to a situation where the separating hyperplane becomes linear and can cause undertraining. The classification of a new test example is given by its position in redescription space in relation to the optimum hyperplane defined during learning. SVMs provide a distance to the hyperplane where the sign determines the class of the example being tested.

5.4.2.2. *Gaussian mixture models*

Gaussian mixture models (GMMs) are commonly used for speech recognition. For a number of years, this model has also been the dominant approach for speaker verification systems [BAR 03, FRE 01, SAN 05]. It has also been successfully used for emotion recognition. GMMs consist of modeling the data x_d as a weighted sum of the coefficients $w_{m,q}$ for each class C_q, from the Gaussian probability density functions $P_{m,q}(x)$:

$$p(x/C_q) = \sum_{m=1}^{M} w_{m,q} p_{m,q}(x)$$

Here, $w_{m,q}$ are the weighting coefficients and $\sum_{m=1}^{M} w_{m,q} = 1$ for each of the considered classes q, and M is the number of density components considered for the model. Each component is expressed in terms of its average $\mu_{m,q}$ and its covariance matrix $\Sigma_{m,q}$:

$$p_{m,q} = \frac{1}{(2\pi)^{1/2} |\Sigma_{m,q}|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(x-\mu_{m,q})^T (\Sigma_{m,q})^{-1}(x-\mu_{m,q})\right]$$

The covariance matrix used is diagonal, i.e. the models are learnt by considering that the observations associated with each of the descriptors are independent.

For each class, each one of the mixture's components models a different area of the data space, also known as a *cluster*. The learning phase consists of estimating the parameters of the Gaussian functions of the model of each class using the observations of the corresponding class. For each class C_q, the parameters to be estimated are:

– the weights $(w_{m,q})_{\ m=1,\ldots M}$ linked to each of the M components in the mixture;

– the means and covariance matrices of each of the mixture's components:

$$(\mu_{m,q}, \Sigma_{m,q})_{m=1,\ldots M}$$

The parameters are normally initialized by the *k*-means algorithm, which allows us to obtain the approximate values of the Gaussian parameters for that class. The parameters' evaluations are then carried out using the *expectation–maximization* algorithm [DEM 77].

Classification can be carried out using a decision rule based on the maximum *a posteriori*. For each classifier a score *a posteriori* (*SAP*) is allocated to each class of emotion to be recognized for each speech segment. This score corresponds to the average of the *a posteriori* log probabilities, which is calculated by multiplying the probabilities obtained for each analysis window. As such, the score calculated for each classifier is:

$$SAP(C_q) = \frac{\sum_{n=1}^{N} \log p(C_q/x_n)}{N}$$

Here, X_n is the observation vector corresponding to the analysis window n and $p(C_q/\ X_n) = \frac{1}{q}$ the *a posteriori* probability corresponding to this window. This is expressed according to Bayes formula as follows:

$$p(C_q/x) = \frac{p(C_q)p(x/C_q)}{p(x)}$$

If we consider that the classes C_q are equiprobable[8], we have $p(C_q) = \frac{1}{q}$ and the problem is reduced to maximizing the modified *a posteriori* score as follows (which in turn is reduced to a maximum likelihood score):

$$S\tilde{A}P(C_q) = \frac{\sum_{n=1}^{N} \log(p(x_n/C_q))}{N}$$

5.5. Performance and assessment

It is currently difficult to compare the performance of systems from different research laboratories. As yet there has been no attempt at assessing them as there has been with speech, speaker and language-recognition systems. However, a first step has been made in this direction with a cooperative initiative called CEICES (*Combining Efforts for Improving automatic Classification of Emotional user States*), established in 2005 by FAU Erlangen in Germany and followed up by the HUMAINE Excellence Network[9] [BAT 06]. This initiative relies on the following principle: an annotated database (audio files, phonetic dictionaries, manual segmentation into words and emotional) is given to different partners. The principal aim is to improve recognition rates by collecting acoustic descriptors from different partners and combining the classifiers. It should be stressed, however, that it is difficult to compare classification methods and the efficacy of sets of descriptors that have been used separately while accounting for the variety of descriptor normalization and transformation procedures. Indeed, the performances of classification systems developed by different laboratories are influenced by a number of different factors:

– data and classes of emotions;

– the problem of "ground truths";

– manual preprocessing;

– learning conditions.

5.5.1. *First factor: data and classes of emotions*

System performance is strongly influenced by the database on which the algorithms have been tested. This leads to the following questions: are the databases

8 The probability of each class can be set differently if we know the frequency of classes considered prior to this.
9 http://www5.informatik.uni-erlangen.de/Forschung/Projekte/HUMAINE/?language=en.

composed of acted or spontaneous emotions? What is the recording quality? What is the degree of diversity with regard to different speakers and contexts?

The first difficulty in comparing performances therefore comes from the limited, if not non-existent, sharing of emotional databases. Some recent studies have tackled this problem and have proposed some solutions to it [BAT 08, CUL 08].

In other research, there have been very few studies undertaken with data recorded in real-life contexts and higher classification scores have, until now, only been obtained on acted data. For example, Schuller *et al*. [SCH 04] obtained a score of 93% for classification into seven classes. Natural data have a greater range of emotional manifestations than laboratory-recorded acted data, indicating that using real data will lead to loss in performance, as noted by [BAT 00]. Shami and Verhelst [SHA 07] have also compared performances from two databases (Kismet and BabyEars) they sensitively illustrate the same emotional classes. The performance achieved through using the database of acted emotions (Kismet) returned around 90% performance for discrimination between three classes (approval, attention, prohibition) while the database of natural interactions between parents and children achieved a score of around 65%.

Higher scores are not necessarily related to the type of data being used. The caricatural and stereotypical nature of acted data certainly aids classification. Nevertheless, the variability of emotions is also a complicating factor in acoustic modeling. As such, higher scores are often linked to more restrained demonstrations of emotional manifestations (speakers, situations) found in laboratory data across the same emotional class. A corpus of acted data (the Situation Analysis in a Fictional and Emotional corpus (SAFE), composed of sequences from fiction films) shows a wide diversity of contexts in terms of speakers, sound environments and situations and, consequently, in the types of emotional manifestations (based on 40 films and 400 speakers). This wealth of context also contributes to a more realistic result – performance was around 70% for discrimination into two classes [CLA 08].

Performance was also dependent on the number of emotional classes under consideration, as classification according to 10 classes is more difficult than for two classes. Performance also depends on the choice of emotional classes being studied, e.g. discriminating between happiness and satisfaction is generally more difficult than differentiating between happiness and a neutral expression. Devillers and Vidrascu [DEV 07], for example, have obtained 55% discrimination with paralinguistic indicators for five classes (fear, anger, sadness, relief and neutral) and have reached 80% for classes of valence (negative and neutral) with paralinguistic indicators.

Among the rare studies examining more than four or five emotional states, Batliner *et al.* [BAT 03] have examined seven states to detect emotions in contextual acted data from a human–machine dialog system using a multimodal kiosk. Current performances of emotion-recognition systems are still too low to imagine being able to identify more than seven classes of emotions from real data or provide any real intraclass variability.

5.5.2. *Second factor: the problem of "ground truths"*

Evaluating performance in classification and recognition systems consists of analyzing differences between "ground truth" annotations and the system's decisions. In terms of speaker recognition, for example, annotations serving as a reference for evaluating performance are fixed and reliable: the speaker in the sound recording is known in the test database. It is therefore a question of comparing the results from the speaker recognition system with the annotation. In speech recognition, reference annotations are often subject to human error, which is estimated to be around 2% (see Chapter 8 for further details). In the field of emotions, we cannot speak about actual human error, since the situation is far more complex.

The emotional content annotation phase of sound recordings is subject to subjectivity in emotion perception and it is therefore difficult to merge different annotations. As a result, performance is dependent on reference annotations. The reliability of annotations also depends on the adequacy of the annotation for the task in question (see Chapter 3). Merging different annotations for different classes of emotions considered is currently measured using kappa[10] [CLA 08, COW 03, CRA 04, DEV 05]. Clavel, for example [CLA 07b] has evaluated an emotion classification system examining fear versus a neutral expression by comparing three annotators' responses with that of the system. The performance obtained with equal error rate[11] varies between 30 and 35% and has a difference of around 5%. To overcome this problem of subjectivity, Steidl *et al*. [STE 05] have proposed using an entropy measure for evaluating performance.

10 Kappa corresponds to the rate of agreement between different annotators who have been corrected such that 0 kappa corresponds to a rate of agreement that could only be obtained by chance [CAR 96].

11 For a GMM-based system, the balance between the rate of false detection and the rate of false rejections will depend on a chosen decision threshold. This is a threshold where the value taken by the rate of false detections is equal to false rejections.This value corresponds to an equal error rate.

5.5.3. *Third factor: manual preprocessing*

Different studies into emotion classification may rely on manual preprocesses whose automization is likely to lower performance. First, various strategies are adopted to extract descriptors, the majority of which rely on previous knowledge of linguistic content. The sound signal, once aligned with the transcription of linguistic content, can be segmented into sentences, words and even syllables. This segmentation can then serve as a basis for an acoustic model of the signal. To develop an effective classification system, this approach hypothesizes that a detailed segmentation of a speech signal can be carried out automatically without decreasing performance, which is not the case for all types of corpora.

Finally, for the normalization phase for descriptors, some descriptors such as fundamental frequency vary from one speaker to another and the differences can be particularly significant between a man and a woman. Studying the variations of these descriptors in relation to emotion can be strongly biased if the study uses several speakers without previous knowledge of them. In order to overcome this bias, numerous studies have used speaker normalization for this descriptor, as Devillers and Vidrascu [DEV 06] and Wred and Schriberg [WRE 03] have done. This normalization presupposes prior knowledge of the speaker being recorded. This is the case for some applications where emotion-detection software can be adapted for a specific speaker (detecting stress in a pilot, for example [VAR 06] and in some dialog systems). However, in a number of applicative contexts (such as with surveillance and in call centers), the identity of the speaker is not previously known and the number of speakers to be processed is significantly greater.

5.5.4. *Fourth factor: learning algorithms*

There have been various studies into comparing different learning algorithms for the same classification problem. For example, Schuller *et al*. [SCH 04] have compared two learning methods on acted laboratory data, one belonging to the generative class of algorithms (GMMs) and the other relying on a discriminative approach (SVMs). The two methods achieved similar results, with 75% for GMMs and 76% for SVMs. This result was also found by Clavel [CLA 07b] when comparing these two methods for a fear/neutral classification on a corpus of fiction of diverse data.

Schuller *et al*. [SCH 03b] have compared GMM classification with continuous hidden Markov model classification. The first method uses overall statistics for characteristics derived from the frequency scale and the energy contour of the speech signal. The second method introduces a temporal complexity by applying continuous Markov models and examining high-level instantaneous characteristics rather than overall statistics. In this framework, performances without inter-speaker evaluations have eventually proved to be better than GMMs.

In general, these two learning methods provided similar results and it is therefore difficult to ascertain from these results which is the best method of classification. This was also found by Lee *et al*. [LEE 02] when they tested three different methods, including discriminative analysis, support vector machines and *k* nearest neighbors. In contrast, Petrushin [PET 03] has compared neural networks and *k* nearest neighbors and has found slightly better performance with neural networks (70% instead of 55%).Vidrascu and Devillers [VID 05], however, have compared SVMs and decision trees on real data (call centers) and the results did not show significant differences.

The classification system uses several neural networks on different learning subsets (*bootstrap*). Shami and Verhelst [SHA 07] have carried out a good comparison of performance by testing different algorithms (*k* nearest neighbors, SVMs, Ada-boost decision trees) on several databases. Algorithms achieving the best performances varied according to the database used. However, this variation is trivial with regard to the confidence interval due to the small size of existing databases.

5.5.5. *Fifth factor: learning conditions*

Evaluation of classification system performance is carried out on databases that may have highly variable characteristics. The distribution of the database into a learning set and a test set must be carried out according to the target application. For example, will the system vary according to speaker? If this is not the case, then it is essential to ensure that the speaker of the test sample is present in the learning database. Schuller *et al*. [SCH 04] have compared the two learning conditions (speaker dependent versus speaker independent) on the same database with the same descriptors and the same classification methods. Performance fell from 89 to 93% to 76 to 75% according to the method used when the system became speaker independent.

5.6. Conclusion

Expectations for the potential uses of automatic voice emotion recognition are high and this area of research is still emerging. Current work has begun to account for different scientific problems highlighted by the practical use of such systems. Some progress has been made but there remain numerous challenges to overcome.

We should also review the steps to be considered when designing an automatic emotion-recognition system:

1) Choosing and acquiring study material and evaluating its quality in terms of suitability for the research objective (type of emotions and contexts researched).

2) Identifying an *annotation strategy* that allows the system to use the data (to describe the corpus' emotional content in a language understandable to the machine that has been adapted to the task in question, offering reliable and relevant annotations for understanding the system's behavior).

3) Managing problems of annotations' *subjectivity* (the annotations considered for forming the learning set, the annotations serving as test references, creating a referential in terms of human performance for categorizing emotions (the second factor in section 5.5).

4) Representing data in the form of *effective acoustic descriptors* for the classification problem considered and extracted according to applicative constraints – normalization techniques and choice of temporal analysis unit (the third factor in section 5.5).

5) Learning of a classification problem by choosing and setting different classification methods.

6) Evaluating system performance against protocols defined according to practical constraints – conditions of speaker or context dependence/independence, etc. (the fifth factor of section 5.5.)

Depending on the application, it is not a question of emotion *recognition* systems (dialog systems), but of emotion-detection systems (surveillance, crisis management, medical applications, robotics, etc.). Speaker turns are also not always clearly separated according to the application. An ongoing challenge is that of processing different types of overlaps between speakers. It is therefore a question of accounting for this in the annotation strategy and allowing the system to process them.

In addition to this, acoustic emotion-recognition systems can be used in parallel with video recognition systems (see Chapter 4). Acoustic indicators alone are often insufficient for detecting emotion [CLA 04]. For example, it is difficult for both humans and systems to distinguish between a cry of happiness and one of fear. Video – and more generally context (situation and linguistic content, i.e. what is said) – provide additional, and even essential, information for automatic emotion recognition. The correlation between these different information sources is a significant issue for emotion detection [ZEN 07].

5.7. Bibliography

[ALK 02] ALKU P., BÄCKSTRÖM T., VILKMAN E., "Normalized amplitude quotient for parametrization of the glottal flow", *Journal of the Acoustic Society of America*, vol. 112, no. 2, pp. 701-710, 2002.

[AMI 96] AMIR N., RON S., "Towards an automatic classification of emotions in speech", *Proc. of ICSLP*, pp. 555-558, Philadelphia, USA, 1996.

[AUD 04] AUDIBERT N., ROSSATO S., AUBERGE V., "Paramétrisation de la qualité de voix: EEG vs. filtrage inverse", *Actes des Journées d'Etude sur la Parole*, Fes, Morocco, pp. 53-56, 2004.

[BAR 03] BARRAS C., GAUVAIN J.-L., "Feature and score normalization for speaker verification of cellular data", *Proc. of ICASSP*, pp. 49-52, Hong Kong, China, 2003.

[BAT 00] BATLINER A., FISHER K., HUBER R., SPILKER J., NTH E., "Desperately seeking emotions or: actors, wizards and human beings", *Proceedings of the ISCA Workshop on Speech and Emotion*, Belfast, Northern Ireland, pp. 195-200, 2000.

[BAT 03] BATLINER A., FISCHER K., HUBER R., SPILKER J., NÖTH E., "How to find trouble in communication", *Speech Communication*, vol. 40, no. 1-2, pp. 117-143, 2003.

[BAT 06] BATLINER A., STEIDL S., SCHULLER B., SEPPI D., LASKOWSKI K., VOGT T., DEVILLERS L., VIDRASCU L., AMIR N., KESSOUS L., AHARONSON V., "Combining efforts for improving automatic classification of emotional user states", *Proc. ISLTC*, pp. 240-245, Ljubljana, Slovenia, 2006.

[BAT 08] BATLINER A., STEIDL S., NTH E., "Releasing a thoroughly annotated and processed spontaneous emotional database: the FAU Aibo Emotion Corpus", *Proceedings of the Workshop on Corpora for Research on Emotion and Affect LREC*, Marrakech, Morocco, pp. 28-31, 2008.

[BOE 05] BOERSMA P., WEENINK D., *Praat: doing Phonetics by Computer*, http://www.praat.org/, 2005.

[CAI 03] CAI R., LU L., ZHANG H.-J., CAI L.-H., "Highlight sound effects detection in audio stream", *Proc. ICME*, July, vol. 3, III-37-40, Baltimore, USA, 2003.

[CAM 03] CAMPBELL N., MOKHTARI P., "Voice quality: the 4th prosodic dimension", *Proc. Int. Congr. Phonetic Sciences*, August, vol. 3, pp. 2417-2420, Barcelona, Spain, 2003.

[CAR 96] CARLETTA J., "Assessing agreement on classification tasks: the kappa statistic", *Computational Linguistics*, vol. 22, no. 2, pp. 249-254, 1996.

[CLA 04] CLAVEL C., VASILESCU I., DEVILLERS L., EHRETTE T., "Fiction database for emotion detection in abnormal situations", *Proc. ICSLP*, Jeju, South Korea, pp. 2277-2280, 2004.

[CLA 07a] CLAVEL C., DEVILLERS L., RICHARD G., VASILESCU I., EHRETTE T., "Abnormal situations detection and analysis through fear-type acoustic manifestations", *Proc. ICASSP*, April, vol. IV-21-24, Honolulu, USA, 2007.

[CLA 07b] CLAVEL C., Analyse et reconnaissance des manifestations acoustiques des émotions de type peur en situations anormales, PhD thesis in Signals and Images, ENST, 2007.

[CLA 08] CLAVEL C., VASILESCU I., DEVILLERS L., RICHARD G., EHRETTE T., "Feartype emotions recognition for future audio-based surveillance systems", *Speech Communication*, vol. 50, pp. 487-503, 2008.

[COW 03] COWIE R., CORNELIUS R., "Describing the emotional states that are expressed in speech", *Speech Communication*, vol. 40, no. 1-2, pp. 5-32, 2003.

[CRA 04] CRAGGS R., "Annotating Emotion in dialogue – issues and approaches", *Proc. CLUK Research Colloquium*, January 2004.

[CUL 08] CULLEN C., VAUGHAN B., KOUSIDIS S., MCAULEY J., "Emotional speech corpus construction, annotation and distribution", *Proceedings of the Workshop on Corpora for Research on Emotion and Affect LREC*, Marrakech, Morocco, pp. 32-37, 2008.

[DEL 96] DELLAERT F., POLZIN T., WAIBEL A., "Recognizing emotion in speech", *Proc. ICSLP*, vol. 3, pp. 1970-1973, Philadelphia, USA, 1996.

[DEM 77] DEMPSTER A., LAIRD N., RUBIN D., "Maximum likelihood from incomplete data via the EM algorithm", *Journal of the Royal Statistical Society*, vol. 39, no. 1, pp. 1-38, 1977.

[DEV 05] DEVILLERS L., VIDRASCU L., LAMEL L., "Challenges in real-life emotion annotation and machine learning based detection", *Journal of Neural Networks*, vol. 18, no. 4, pp. 407-422, 2005.

[DEV 06] DEVILLERS L., VIDRASCU L., "Représentation et détection des émotions dans des dialogues enregistrés dans un centre d'appel – Des émotions complexes dans des données réelles", *Interaction Emotionnelle*, vol. 20, no. 4-5, pp. 447-476, 2006.

[DEV 07] DEVILLERS L., VIDRASCU L, "Real-life emotion recognition in speech", *Speaker Characterization II*, Springer-Verlag, Berlin, Germany, 2007.

[DUD 73] DUDA R., HART P.E., *Pattern Classification and Science Analysis*, Wiley-Interscience, New York, USA, 1973.

[EHR 04] EHRETTE T., Les voix des services telecoms, de la perception à la modélisation, PhD thesis, Paris XI University, France, 2004.

[ESS 05] ESSID S., Classification automatique des signaux audio-fréquences: reconnaissance des instruments de musique, PhD thesis, Telecom-Paris, France, 2005.

[FAN 60] FANT G., *Acoustic Theory of Speech Production*, La Hague, Paris, France, 1960.

[FRA 03] FRANCE D., SHIAVI R., SILVERMAN S., SILVERMAN M., WILKES D., "Acoustical properties of speech as indicators of depression and suicidal risks", *IEEE Transactions on Biomedical Engeneering*, vol. 47, no. 7, pp. 829-837, 2003.

[FRE 01] FREDOUILLE C., BONASTRE J., MERLIN T., "Bayesian approach based decision in speaker verification", *Proceedings of a Speaker Odyssey, the Speaker Recognition Workshop*, Greece, pp. 77-81, 2001.

[GUY 03] GUYON I., ELISSEEFF A., "An introduction to feature and variable selection", *Journal of Machine Learning Research*, vol. 3, pp. 1157-1182, 2003.

[HEN 07] VAN HENGEL P.W.J., ANDRINGA T., "Verbal aggression detection in complex social environments", *Proceedings of the IEEE Conference on Advanced Video and Signal Based Surveillance*, London, pp. 15-20, September 2007.

[HES 84] HESS W., O'SHAUGHNESSY D., "Pitch determination of speech signals: Algorithms and devices", *The Journal of the Acoustic Society of America*, vol. 74, no. 4, pp. 1277-1278, October 1984.

[HSU 02] HSU C.-W., LIN C.-J., "A comparison of methods for multi-class support vector machines", *IEEE Transactions on Neural Networks*, vol. 13, no. 2, pp. 415-425, 2002.

[IST 03] ISTRATE D., Détection et reconnaissance des sons pour la surveillance médicale, PhD thesis, University of Grenoble, France, 2003.

[KIE 00] KIENAST M., SENDLMEIER W.-F., "Acoustical analysis of spectral and temporal changes in emotional speech", *Proceedings of the ISCA ITRW on Speech and Emotion*, Belfast, Northern Ireland, pp. 92- 97, 2000.

[KLA 08] KLAPURI A., "Multipitch analysis of polyphonic music and speech signals using an auditory model", *IEEE Trans. Audio, Speech Language Processing*, vol. 16, no. 2, pp. 255-266, February 2008.

[KOZ 65] KOZHEVNIKOV V.A., CHISTOVICH I.A., *Speech Production and Perception*, Joint Publication Research Service, Washington DC, USA, 1965.

[KWO 03] KWON O.-W., CHAN K., HAO J., LEE T.-W., "Emotion recognition by speech signals", *Proc. Eurospeech*, Geneva, Switzerland, pp. 125-128, 2003.

[LEE 02] LEE C., NARAYANAN S., PIERACCINI R., "Classifying emotions in human machine spoken dialogs", *Proc. ICME*, vol. 1, pp. 737-740, 2002.

[LOO 07] LOOIJE R., NEERINCS M., KRUIJFF G.-J., "Affective collaborative robots for safety and crisis management in the field", *Proc. ISCRAM*, pp. 497-505, 2007.

[MCA 99] MC ADAMS S., "Perspectives on the contribution of timbre to musical structures", *Comput. Music J.*, vol. 23, no. 3, pp. 85-102, 1999.

[NGH 07] NGHIEM A., BREMOND F., THONNAT M., VALENTIN. V., "ETISEO, an evaluation project for video surveillance systems", *Proceedings of the IEEE International Conference on Advanced Video and Signal based Surveillance*, London, pp. 476-481, September 2007.

[OSG 75] OSGOOD C., MAI W. H., MIRON M., *Cross-cultural Universals of Affective Meaning*, University of Illinois Press, Urbana, USA, 1975.

[OUD 03] OUDEYER P.-Y., "The production and recognition of emotions in speech: features and algorithms", *International Journal of Human Computer Interaction, special issue on Affective Computing*, vol. 59, no. 1-2, pp. 157-183, 2003.

[PEE 03] PEETERS G., "Automatic classification of large musical instrument databases using hierachical classifiers with inertia ratio maximization", *Proc. AES 115th Convention*, New York, USA, pp. 1-13, October 2003.

[PET 03] PETRUSHIN V., "Emotion in speech: Recognition and application to call center", *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, Hong Kong, China, pp. 7-10, 2003.

[PIC 97] PICARD R., *Affective Computing*, MIT Press, Cambridge, USA, 1997.

[SAN 05] SANCHEZ-SOTO E., Réseaux Bayésiens Dynamiques pour la Vérification du Locuteur, PhD thesis, Télécom-Paris, France, 2005.

[SCH 98] SCHERER K.R., JOHNSTONE T., SANGSUE J., " L'état émotionnel du locuteur: facteur négligé mais non négligeable pour la technologie de la parole", *Actes des Journées d'Etudes sur la Parole*, Matigny, France, pp. 249-257, 1998.

[SCH 03a] SCHERER K., "Vocal communication of emotion: a review of research paradigms", *Speech Communication*, vol. 40, no. 1-2, pp. 227-256, 2003.

[SCH 03b] SCHULLER B., RIGOLL G., LANG M., "Hidden Markov model based speech emotion recognition", *Proc. ICASSP*, Hong Kong, China, pp. 1-4, 2003.

[SCH 04] SCHULLER B., RIGOLL G., LANG M., "Speech emotion recognition combining acoustic features and linguistic information in a hybrid support vector machine-belief network architecture", *Proc. ICASSP*, vol. 1, Montreal, Canada, pp. I- 577-80, 2004.

[SER 84] SERKEY A., HANSON B., " Improved 1-Bark auditory filter", *Journal of the Acoustical Society of America*, vol. 75, no. 6, pp. 1902-1904, 1984.

[SHA 03] SHAFRAN I., RILEY M., MOHRI M., "Voice signatures", *Proc. ASRU Workshop*, St. Thomas, USA, pp. 31-36, 2003.

[SHA 07] SHAMI M., VERHELST W., "An evaluation of the robustness of existing supervised machine learning approaches to the classification of emotions in speech", *Speech Commun*., vol. 49, no. 3, pp. 201 212, 2007.

[STE 05] STEIDL S., LEVIT M., BATLINER A., NÖTH, E., NIEMANN, H., "Of all things the measure is man, Automatic classification of emotions and inter-labeler consistency", *Proc. ICASSP*, Philadelphia, USA, pp. 317-320, 2005.

[VAP 95] VAPNIK V., *The Nature of Statistical Learning Theory*, Springer, Berlin, Germany, 1995.

[VAR 06] VARADARAJAN V., HANSEN J., AYAKO I., "UT-SCOPE – A corpus for speech under cognitive/physical task stress and emotion", *Proceedings of the LREC Workshop en Corpora for Research on Emotion and Affect*, Genoa, Italy, pp. 72-75, 2006.

[VER 03] VERFAILLE V., Effets audionumériques adaptatifs – théorie, mise en oeuvre et usage en création musicale numérique", PhD thesis, University of Aix-Marseille II, 2003.

[VID 05] VIDRASCU L., DEVILLERS L., "Detection of real-life emotions in call centers", *Proc. Eurospeech*, Lisboa, Portugal, pp. 1841-1844, 2005.

[VLA 07] VLASENKO B., SCHULLER B., WENDEMUTH A., RIGOLL G., "Frame vs. turn-level: Emotion recognition from speech considering static and dynamic processing", *Affective Computing and Intelligent Interaction*, Lisbon, Portugal, pp. 139-147, September 2007.

[WRE 03] WRED B., SHRIBERG E., "Spotting 'hot spots' in meetings: Human judgments and prosodic cues", *Proc. Eurospeech*, Genoa, Italy, pp. 2805-2808, 2003.

[YEG 98] YEGNANARAYANA B., D'ALESSANDRO C., DARSINO V., "An iterative algorithm for decomposition of speech signals into periodic and aperiodic components", *IEEE Transactions on Speech and Audio Processing*, vol. 6, no. 1, pp. 1-11, 1998.

[ZEN 07] ZENG Z., PANTIC M., ROISMAN G. I., HUANG T. S., "A survey of affect recognition methods: audio, visual and spontaneous expressions", *Proceedings of the 9th International conference on Multimodal Interfaces*, ACM, New York, USA, pp. 126-133, 2007.

[ZWI 57] ZWICKER E., FLOTTORP G., STEVENS S., "Critical bandwidth in loudness summation", *J. Acoust. Soc. Am.*, vol. 29, pp. 548-557, 1957.

Chapter 6

Modeling Facial Expressions of Emotions

6.1. Expressive conversational agents

As users' expectations grow in terms of friendliness of human-machine interfaces, the ubiquitous and affective computing domains investigate the use of embodied conversational agents (ECAs) that can express emotion. ECAs are virtual agents that can autonomously communicate verbally and non-verbally with a user. Focus on this area of research is driven by a desire to improve human–machine interactions through the use of social and emotional signals. To express emotions, the agent needs to have access to a model that identifies a method of communication that can be understood by humans and also be endowed with non-verbal communication means.

Studies have shown that humans communicate their emotions via a number of modalities and that the face is a primary communication method [DAR 72, DUC 99, EKM 72, KAI 01]. In human–machine interaction, communicating with an emotionally expressive ECA creates a more satisfying interaction than one with an expressively neutral agent [WAL 94]. However, more than merely retaining the interest of the user, non-verbal expressions can also be useful for clarifying verbal texts [ELL 97]. This is evident with situation-specific responses where inappropriate expressions can lead to undesired consequences [BED 05, WAL 94]. Similarly, the

Chapter written by Sylwia Julia HYNIEWSKA, Radosław NIEWIADOMSKI and Catherine PELACHAUD.

Emotion-Oriented Systems Edited by Catherine Pelachaud

same agent, depending on the quality of its facial expressions, may be perceived as more or less believable, which may in turn reduce the user's trust [REH 05]. According to some academics, human interactions with virtual characters can be similar to those developed with real humans [BRA 05, REE 96, SCH 06]. This therefore indicates that a human emotional model could be applied to modeling artificial emotions.

In this chapter we will study some theoretical approaches from affective psychology that have contributed to modeling conversational agents' facial behavior. We will also examine how discrete emotion theories, as well as dimensional and componential emotion theories, view these expressions in the framework of the complex process that is emotion. Following this, we will investigate some of the computational models used to design ECA expressions and the theoretical models on which they are based. The chapter will conclude by discussing some of the research into expressions conveying blends of emotions and by examining their social constraints.

6.2. Expressions and their emotional states

Identifying the mechanisms controlling emotions is a topic of ongoing debate. There are three main approaches to this question: discrete, dimensional and componential emotion theories. We will examine each of these theories to illustrate the debate around these mechanisms

Discrete emotion theories [EKM 84, EKM 89, EKM 92, IZA 71, IZA 93, TOM 62, TOM 84] propose that there are number of universal emotions that are determined by neuromotor mechanisms. These mechanisms trigger a prototypical emotional response with a fixed pattern of facial expressions specific to each emotion.

Dimensional emotion theories [FON 07, MEH 96, RUS 80] define emotions as points in a continuous multidimensional space where each dimension is a fundamental property against which all emotions are evaluated. Dimensional theories propose that there is a link between an emotion, a facial expression (or its components) and these different dimensions [RUS 80].

Component theories of emotion [ELL 03, ROS 01, SCH 84, SCH 87, SCH 01, SMI 85] propose that individual elements of facial expressions are determined by the results of cognitive evaluations (appraisals).

In our next section we will examine these theories and how they interpret facial expressions of emotions in further detail.

6.2.1. *Expressing discrete emotions*

Proponents of discrete emotion theories propose that emotions are triggered by automatic mechanisms, such as neuromotor emotional programs. These programs are independent from cognitive evaluations (see [EKM 72] for further details). For example, Tomkins has described these affect programs as the cause of unique expression patterns that are characteristic of a specific emotion [TOM 63]. Research into this approach has focused on "basic" prototypical emotional patterns. Each basic emotion is characterized by an adaptive function, an expression (i.e. a specific facial behavior), a series of physiological changes and a conscious feeling of this emotion [KEL 97, MAN 05]. According to this theory, both emotions and their facial expressions are considered discrete. As such, every expression is made up of a defined element and not a set of independent and variable elements. All the components of an expression have a common development without temporal variations between sub-expressions. For example, raising the eyebrows and lowering the corners of the mouth may indicate an expression of sadness. However, these actions do no express sadness if taken separately. Lowering the corners of the mouth is not an indicator of sadness in itself and is therefore not an emotional indicator alone. During an expression of sadness, the lips and eyebrows begin and finish moving at the same time. Discrete emotion-based research has mostly relied on photographs of expressions at their apex and has suggested that expressions have a linear development.

6.2.2. *Dimensional approaches to emotional expression*

Other researchers have proposed that emotions are not structured in terms of discrete categories, but as continuous variations over a range of proposed dimensions. The two most widely accepted dimensions today are valence (positive versus negative emotions) and arousal (or general activation – high or low level of arousal) [RUS 80]. Russell's circumplex model supposes that there is a direct link between each muscular movement in the face and its position on the activation and valence dimensions. Russell, however, has argued that the face does not communicate emotions more than the rest of the body (intonation, posture, etc.) but communicates general information on an individual's underlying internal state [GRA 06], which can be viewed in relation to these two dimensions. This facial information does not allow us to categorize an emotion with a concrete label. For example, in one study participants were shown facial expressions but were unable to give a specific name to the emotion demonstrated without being provided the context of production. Perceptions of valence and arousal, on the other hand, appear to be fairly fluid [RUS 91].

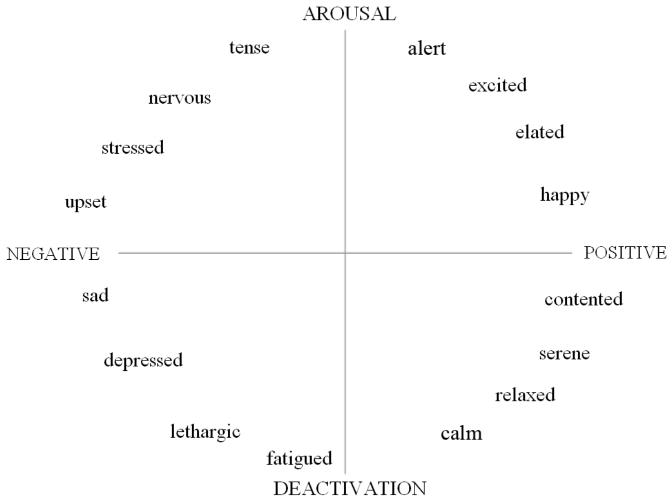


Figure 6.1. *The dimensional representation proposed by Breazeal [BRE 03] and inspired by Russell [RUS 97]*

A different dimensional theory has also been proposed by Schlosberg [SCH 52], who has modeled facial expressions as a circle labeled according to the dimensions of "pleasant–unpleasant" and "attention–rejection". In this model, six expressions including happiness, surprise, fear, sadness, disgust and anger are arranged around a circle according to the extent to which they overlap. The PAD (pleasure–arousal–dominance) emotional model however represents a 3D approach. The three dimensions in this model (based on the theoretical foundations set out by Mehrabian [MEH 80, MEH 95]) are independent and allow us to describe and measure emotional states. These three dimensions are those of pleasure–displeasure, activation–deactivation and dominance–submission, allowing us to identify a greater number of emotions. For example, anger and fear are both negative and active emotions but are differentiated by their level of dominance.

The terms used to define emotions can be visualized as points in a 3D space. When PAD scores are standardized, each emotional term can be described in terms of its values on each of these axes. Evaluations of emotions are represented by scores from -1 to +1. For example, anger is a strongly unpleasant emotional state (-0.51), strongly active (0.59) and, on average, dominant (0.25) [MEH 97]. In the PAD emotional model there are eight emotional groups that are defined by high or low combinations of pleasure, activation and dominance. As such, states of anxiety (low pleasure, high activity and low dominance) include nervousness, insecurity and suffering. The importance of using these three dimensions and not just those of pleasure and activation has largely been proven [MEH 86, MEH 87, SOR 06,].

Fontaine *et al.* [FON 07] have proposed a 4D model for representing emotions. This work was based on theoretical concepts where emotions are considered as inter-linked changes of six components including cognitive evaluations of events, psycho-physiological changes (bodily feelings), motor expressions (e.g. facial expressions), action tendencies, subjective experience and emotional regulation [ELL 03]. In their research, Fontaine *et al*. asked speakers of three different languages to evaluate 24 prototypical emotional labels on a Likert scale (often used in questionnaires to graduate responses) of 144 characteristics representing emotional activity. Breaking these down into their components demonstrates that evaluation–valence is the dimension that best explains variance. This is followed by situation control and activation, with the fourth dimension being the unexpected. This model allows us to understand emotions in terms of dimensions and their defining characteristics. Among these, we find expressive behavior associated with each of these dimensions, e.g. unexpected and new events are often expressed by raising the eyebrows, lowering the jaw, etc.

6.2.3. *Componential expression of emotions*

Cognitive theoreticians [KAI 98, KAI 01, ORT 90] from the componential approach have opposed the concept of emotions as a result of fixed biological programs. In contrast, many have proposed that there are a large number of highly nuanced emotional states that linguistic labels can only convey by grouping them together or through the main tendencies of the most common states. Scherer, a proponent of the componential approach, termed these states “modal emotions” [SCH 84]. This approach also maintains that emotions are the result of cognitive evaluations (appraisals). The notion behind this is that an emotional state is caused by the significance attributed by us to different elements in a given situation. Emotion is not defined and triggered directly by a situation or a stimulus but depends on the relation that an individual establishes with their environment. This relation is created by an appraisal (for more details, see [KAI 01]) and can be perceived via different modalities. Facial expression is closely linked to different stages of appraisal. Scherer has described these facial changes in terms of action units (AUs), as defined by Ekman [EKM 78], each representing a particular activation of muscles during an expression. Scherer has proposed that, for each stage in the appraisal process, there are several combinations of AUs, with the final expression being the accumulation of all the expressions specific to the appraisal stages. As such, with the numerous potential combinations of AUs, it is possible to obtain a large number of facial expressions.

6.3. Computational models for facial expressions of emotions

In this section we will focus on the existing computational models providing ECA expressions and the theoretical models on which they are based. Several facial expression models have been proposed to enhance the facial behavior of conversational agents. These models often rely on the generation of new expressions by averaging the values of the parameters of discrete expressions of basic emotions (see section 6.2.1 [EKM 75, EKM 03a]). There are also several models that use dimensional (see section 6.2.2.) and componential (see section 6.2.3) approaches for modeling facial expressions.

6.3.1. *A discrete representation of facial expressions*

The MPEG-4 format [PAN 03] relies on a discrete representation of emotions. This format introduced the parameterization of a facial expression according to 68 facial action parameters (FAPs). The vast majority (66) of FAPs are linked to movement in specific parts of the face, such as the eyebrows or the corners of the mouth. There are also two global parameters, one specifying visemes (articulatory movements of the lips) and the other (FAP 2) is used to display one of the six predefined facial expressions (anger, happiness, sadness, disgust, surprise and fear as defined by Ekman [EKM 75]). This last parameter also allows mixtures of these six emotions. Mixtures can be generated by specifying the importance of the different emotional components. A new expression can be generated by combining the weights given to the basic expressions [OST 02]. Finally, both identifying the basic facial expressions and the interpolation algorithm between the key frames depend on the direct implementation of the format. As such, based on the definition of expressions in terms of discrete representations, we get a large number of facial expressions. Nevertheless, contrary to the other approaches (sections 6.3.2 and 6.3.3) this model does not allow us to create new facial expressions where the emotion could be verbally identified by a specific label.

6.3.2. *Dimensional representation of facial expressions*

Tsapatsoulis *et al*. [TSA 02] and Albrecht *et al*. [ALB 05] have also developed models for generating expressions. To generate new facial expressions both models exploit the distances between basic and non-basic emotions in a certain dimensional emotion space to generate non-basic facial expressions [EKM 75]. They use different multidimensional spaces in which emotional labels were placed. In these two approaches, new expressions are constructed based on Ekman's six basic expressions (anger, disgust, sadness, happiness, surprise and fear [EKM 75]). Tsapatsoulis *et al*. introduced an algorithm to their model that can construct new expressions using FAPs, defined by the MPEG-4 standard. The originality of this approach lies in using fuzzy definitions of facial expressions.

Each facial expression is defined by a set of intervals. For each animation parameter, Tsapatsoulis *et al*. defined an interval of an expression's possible values. In order to generate the new expression (which is also determined by intervals), mathematical transformations are applied to the intervals of the original expression(s). Tsapatsoulis *et al.* used two approaches [TSA 02]. First, an expression can be derived from a basic expression by "scaling" of all the FAP values of a basic emotion's expression. Second, the expression can be generated by using the definitions of the two closest basic emotions as defined in the dimensional spaces of both Whissell [WHI 89] and Plutchik [PLU 80]. Whissell, for example, has proposed using two values (corresponding to two axes of activation and evaluation) for some emotional labels while Plutchik has defined the angle value on his emotional wheel for each emotion. Thus, Tsapatsoulis *et al*., used 2D coordinates and the angle value to calculate new expressions. The expression parameters (FAPs) are weighted by the values of the researched emotion and the two adjacent basic emotions. Albrecht *et al.* [ALB 05] have proposed an improvement on this model by using an anatomical representation of a face. They use a model of muscular contractions instead of using MPEG-4 FAPs. They also used a 3D space of emotional states defined by activation, evaluation and power, as proposed by Cowie *et al*. [COW 99].

While the EmotionDisc model [RUT 03] is based on the six basic expressions [EKM 75], it can be considered to be dimensional due to its use of bi-linear interpolation between different expressions. In the EmotionDisc, six expressions are positioned equally apart around a circle and the neutral expression is represented by the middle of the circle (see Figure 6.2). The expression synthesized is the result of the interpolation between two basic expressions that are closest together around the circle and the neutral expression of virtual agents. The distance from the center of the circle represents the expression's intensity. The spatial relations are used to establish expressions that correspond to each point on the EmotionDisc. This method allows us to obtain an expression continuum between two particular expressions.

Another dimensional model for generating facial expressions of various emotional states has been introduced in Grammer and Oberzaucher [GRA 06]. The characteristic patterns of facial movements are localized in a 2D space of activation and valence. In this work Grammer and Oberzaucher found specific configurations for AUs that characterize the four extremes of the Russell's emotional space (high and low arousal, positive and negative valence) [RUS 80]. This approach allows us to generate new expressions from these two dimensions using AUs [EKM 78]. For this purpose, Grammer and Oberzaucher first of all asked participants to evaluate facial expressions composed of random sets of action units. The evaluation consisted of attributing labels to random expressions. A factorial analysis was carried out on these labels and three factors explaining a large number of variations in interpretations were labeled as a factor of valence, dominance and arousal. The activation and valence factors attributed to the faces were retained to calculate the action unit scores for each of these expressions. As such, the configuration of action units characterizing specific activation and valence values can be obtained using statistical analysis. To verify this model, Ekman's six basic emotions [EKM 75] were reconstructed in the valence and activation spaces. In essence, this approach is an attempt to reconcile the various facial expression theories. It can be used to create facial expressions determined by FACS AUs [EKM 78] for every emotional state described in a 2D space of activation and valence.

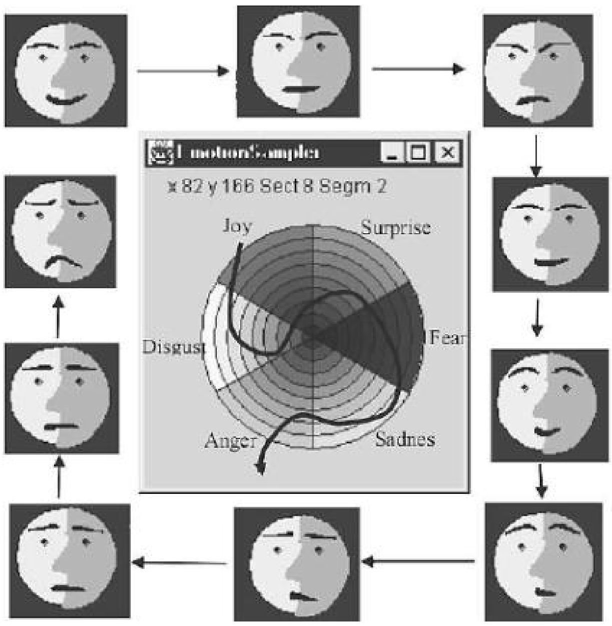


Figure 6.2. *EmotionDisk [RUT 03]*

6.3.3. *Componential approaches to facial expressions*

Other researchers have used a purely componential approach to generate ECA expressions. Wehrle *et al*. [WEH 00], for example, have created a special software for applying Scherer's componential model [SCH 01], which generates expressions by identifying the action units involved. This tool, called a Facial Action Composing Environment (FACE), creates schematic 3D facial expressions [WEH 00]. It allows us to extract an expression's different elements and to vary the dynamic of specific components independently. Furthermore, the elements can be subject to asymmetric movement changes, as with, for example, half smiles.

The number of expressions created can be large and different processes allow us to move beyond the characteristics of basic emotion. Scherer's theory has also been implemented by two research teams, under the direction of Lisetti [PAL 06] and Kollias [MAL 09], respectively, to create artificial expressions. Paleari and Lisetti [PAL 06] used the agent Cherry, a platform with a human avatar, to manually generate expressions by applying Scherer's predictions [SCH 01] with regard to the emotions of anger, fear, happiness and disgust (see Figure 6.3.). The expressions generated are limited in number and the importance of this research lies in the emphasis placed on the temporal relation between the different dynamic elements composing an expression and linked to different appraisal stages.

Figure 6.3. *Agent Cherry displaying a possible evolution of an expression of fear according to Scherer's theory [PAL 06]*

The facial expression is not activated immediately and uniformly, rather the animation parameters are activated at different moments. The final result is an animation composed of a sequence of several configurations of micro-expressions. To do this, Paleari and Lisetti converted the AU related to cognitive evaluation into Cherry's action parameters. They also examined intensities and appropriate temporal constraints for these parameters that might create believable expressions [PAL 06]. Malatesta, Raouzaiou, Karpouzis and Kollias [MAL 09] have similarly used MPEG-4 facial animation by converting AUs into FAPs. They created sequences predicted by componential appraisal theory [SCH 01] and applied these to the ECA Greta [BEV 07]. Each expression is derived using the "addition" of a new AU to the previous AUs.

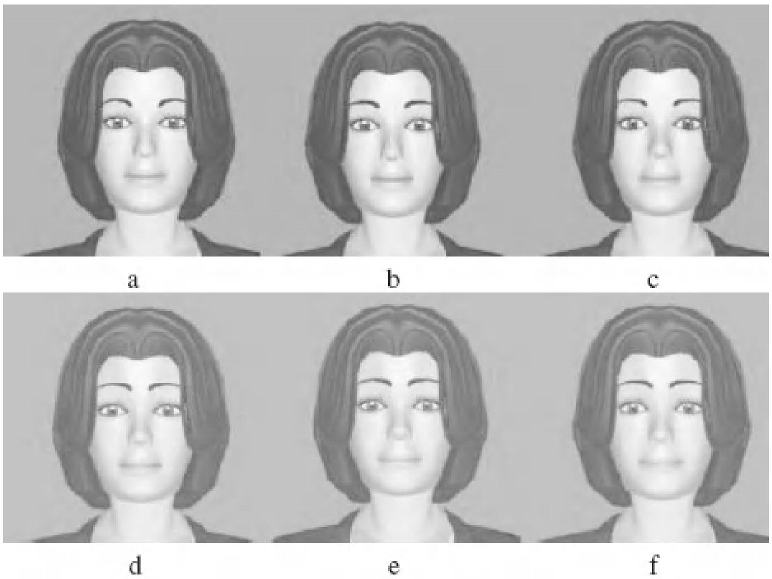


Figure 6.4. *Predictions for intermediary micro-expressions in an expression of fear: a) neutral expression; b) sudden surprise; c) negative valence; d) incoherence; e) goal obstructivion; f) low coping/final expression of fear (according to Scherer's appraisal theory, adapted by Malatesta et al. [MAL 09])*

6.3.4. *Mixtures of emotions and social constraints*

Some researchers have attempted to improve their understanding of facial expressions by including aspects of context and the socio-cultural norms that regulate expressions. As such, Ekman and Friesen have described facial changes that could intervene if the context does not favor the direct expression of an emotion. In some situations, we see that another, more socially acceptable, expression is used [EKM 75]. According to Ekman [EKM 03a], a large number of observed emotional expressions can be explained using the six basic emotions (anger, happiness, fear, disgust, sadness and surprise) and their blends, i.e. expressions involving more than one emotion. Ekman and Friesen [EKM 69] have observed that different types of blends are found in everyday life. Context and socio-cultural norms sometimes encourage the superposition of emotions, concealment (when an expression is masked by another), dissimulation (when an unfelt emotion is expressed) or the inhibition of an emotion. These blends of emotions are achieved, according to Ekman, by a combination of expressions on different parts of the face. For example, when two emotions are superposed, the expression is composed of facial behaviors of one emotion for the upper part of the face and another for the lower part [EKM 75]. The line between the two areas is not always clearly demarcated. For some pairs of emotions the eyes can be considered as being included in the upper part of the face (e.g. anger and sadness), while with some other pairs of emotions they are not [EKM 75]. Ekman has described 18 different expressions of superposition for pairs involving the six emotions [EKM 75, EKM 03a]. The combinations on the upper and lower parts of the face are not all plausible. For example, happiness when superposed with sadness will be expressed by the lower part of the face and sadness will be expressed by the upper part. The opposite pattern never occurs.

Ekman and Friesen have also shown that humans can differentiate between expressions of emotions that are actually felt and those that are faked [EKM 69, FRA 95, GOS 95]. Ekman has proposed a list of deception clues [EKM 89, EKM 95, EKM 03b], i.e. characteristics that can help us differentiate between felt and faked emotions. Given that humans are not capable of consciously controlling all their facial muscles, observing facial behaviors that are produced with difficulty in controlled expressions allows us identify these "faked" and "real" expressions of emotional states. The expression of some emotions could therefore be associated with characteristic changes, such as activation of the orbicularis oculi (contraction of the lower part of the eyelid) when expressing happiness [EKM 03a]. These characteristics of facial actions are not only absent in fake expressions, but they are also difficult to hide when being in some emotional. states. Furthermore, real expressions can be differentiated from faked ones due to their variation in symmetry, synchronization and duration [EKM 75, EKM 89]. Fake expressions are more often asymmetric [EKM 03b], more abrupt [EKM 82, FRA 95] and often expressed for longer periods than expressions of felt emotions [EKM 03b].

Bui [BUI 04] has used Ekman and Friesen's concept that mixtures of emotions are constructed as expressions of different parts of basic emotions [EKM 75]. He has used a set of fuzzy rules to identify mixtures of the six basic emotions [EKM 75]. Sub-set of rules are applied to each pair of emotions. The Bui model allows us to identify, by fuzzy inference, the level of muscular contraction in the final expression. The calculation is carried out in relation to the intensity of the agent's emotional state [BUI 04].

Rehm and André have also studied different types of expressions in complex situations made by ECAs, including those in simulations [REH 05]. Based on Ekman's descriptions [EKM 75, EKM 03a, EKM 03b], they manually defined the parameters for faked expressions. These were characterized by asymmetry and some emotion-specific movements, which are difficult to consciously control. In their study on ECA liars, they showed that characteristics of fake expressions are sufficient for users to be able to unconsiously differentiate between the agents [REH 05].

Niewiadomski and Pelachaud [NIE 07c], also used the face partitioning approach as proposed by Ekman [EKM 75, EKM 03a, EKM 03b]. They developed their own model of expressions for modeling mixtures of emotions including faking, inhibiting, superposing (see Figure 6.5) and masking of an emotion with another (see Figure 6.6). In these expressions, described by Niewiadomski and Pelachaud as "complex expressions" [NIE 07a], each expression is defined by a set of eight areas of the face (eyebrows, upper eyelid, direction of the eyes, lower eyelids, cheeks, nose, lips, and tension of the lips). Each expression is a composition of eight facial areas and each of them can communicate different emotion. Thus several emotions can be expressed on different parts of the face. In the case of an expression of sadness masked by happiness, for example, sadness is expressed in the area of the eyebrows while happiness is expressed in the mouth region.

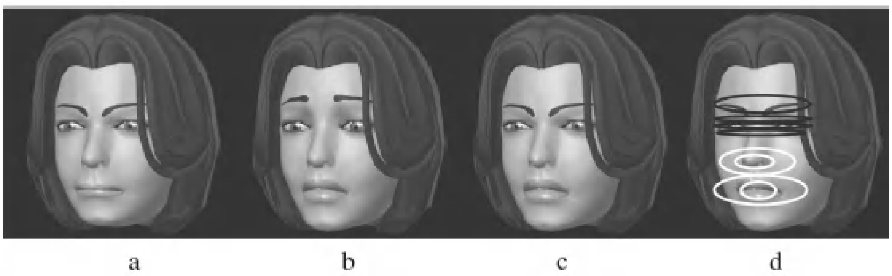


Figure 6.5. *Superposition of the expressions of anger and sadness: a) an expression of anger; b) an expression of sadness; c) superposition of anger and sadness; d) superposition of anger and sadness with markers*

The main task of the algorithm is to attribute emotional expressions to different parts of the face. In order to do this, Niewiadomski and Pelachaud defined a set of fuzzy rules describing the composition of these facial areas. These rules are based on the description proposed by Ekman that refers to the six basic emotions. To generate other complex expressions than ones composed from these six (i.e. ones that are not specified in the algorithm's rules), the algorithm chooses the most appropriate solution.

Another algorithm has been used for this purpose that is based on the notion of fuzzy similarity [NIE 07b]. For this purpose each facial expression is described by a set of fuzzy sets, where each fuzzy set correspondes to one animation parameter (the FAPs in the MPEG-4 format). Fuzzy similarity is used to calculate the degree of visual similarity between two expressions.

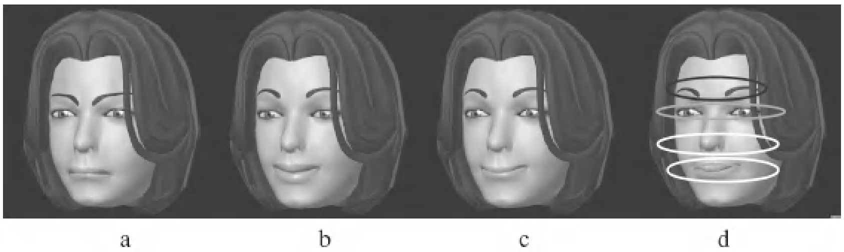


Figure 6.6. *Masking anger with an expression of happiness: a) an expression of anger; b) happiness; c) masking anger with an expression of happiness; d) with markers*

6.3.5. *Sequences of emotional expressions*

In accordance with componential emotion theory in which an emotion is a dynamic episode producing a sequence of response patterns (gestural, facial or vocal) [SCH 07], it could be advantageous for the believability of conversational agents to include more temporal variability and multimodality to their expressions. Despite a large amount of research on the face, some studies have shown that emotions can also be recognized from bodily movements [POL 01, WAL 98]. It is possible to obtain this kind of information through direct observation, whether guided by theory or not.

Other studies have also explored the complexity of emotional expression by focusing on their multimodality. For example, Keltner [KEL 95] has studied the sequence of gestures and movement in embarrassment. This study was based on the frequency of appearance of specific patterns of movement in audiovisual databases.

Shiota *et al*. have studied the three emotions of awe, amusement and pride [SCH 03]. They have demonstrated that expressions of these three emotions are more complex than the simple static images predicted by Ekman's research [EKM 75] in the prototypical form of happiness. These expressions are rather a set of possible signals, which sometimes appear in asynchronous sequences of these different signals. These expressions should not be seen as discrete and not all of the elements composing it have to appear at the same time for it to be recognized as the externalization of a specific emotional state. In an expression of admiration, for example, Shiota *et al*. [SCH 03] observed that the inner part of the eyebrows was raised, the eyes and mouth were more open, the head was tilted forward and there was a large, visible intake of breath. While movements of the eyebrows, eyes and mouth are found in the large majority of expressions of admiration, they only appear in a third of cases. Shiota has similarly analyzed expressions of amusement and pride, providing detailed information on the exact presence of facial movement in terms of AUs but also on their frequency of occurrence. Shiota differentiates these three emotions and moves away from the idea of a shared expression. As such, the Duchenne smile, often considered as the single fundamental element of an expression of positive emotion, cannot be used as a single expression of different positive internal states. These are multimodal sequences that allow differentiation.

Niewiadomski *et al*. [NIE 11] have created an algorithm that can generate emotional expressions composed of different signals (or behaviors) that could scheduled in different ways in time and using several communicative channels (face, gaze, position of the head, gestures). This research was based on the works of Keltner *et al*. [HAI 99, KEL 95, KEL 97], Shiota *et al*. [SCHI 03] and Harrigan and O'Connell [HAR 96]. These works have emphasized the need to move beyond static facial expressions of emotions.

The algorithm by Niewiadomski *et al*. can be used to create multimodal expressions of undetermined duration, while the different signals themselves (or micro-expressions) last for a predetermined amount of time. For each emotional state, a behavior set of is defined including a set of signals expressing particular emotions and a constraint set defining the relations between these signals. From an emotional label, the system generates a sequence of expressions, i.e. an animation of a specific length composed of a sequence of signals of different modalities. The system does this by choosing a coherent subset of the signals from an appropriate behavior set, their durations, the order of appearance, and respecting the constraints defined in a corresponding constraint set. For one single emotional label, this algorithm allows us to generate a large number of animations respecting the previously defined constraints. Figures 6.7 and 6.8 are two examples of animation of the expression of embarrassment.

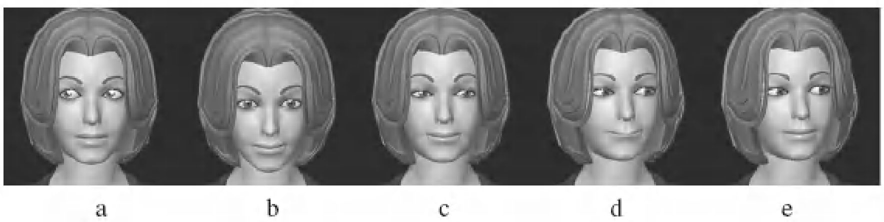


Figure 6.7. *Greta displaying a sequence of signals expressing embarrassment: a) gaze to the right; b) downward head and gaze; c) gaze to the left; d) gaze to the left and a tensed smile; and e) gaze to the left*

Xueni Pan *et al.* [PAN 07] have proposed a different approach for displaying emotions that can be expressed not by static images but by a sequence of signals, such as facial expressions and movements of the head. Sequences of signals have been taken from a video corpus and they have been used to create a "motion graph" from this data.

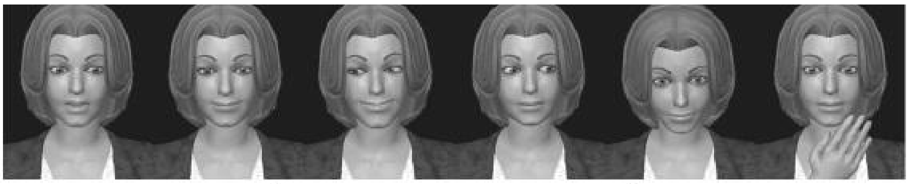


Figure 6.8. *The agent, Greta, displaying an example of a multimodal expression of embarrassment: a) a neutral expression; b) slight smile and gaze to the right; d) gaze to the left; e) forward movement of the head; and f) self-centered gesture*

In this motion graph, the arcs represent the recorded signal sequences and the potential points of transition between them. The different paths in the graph correspond to the different expressions of emotion. The new animations can be generated by a rearrangement of the observed expressions.

6.4. Conclusion

The different models conceptualizing emotions contribute, by their specificities, to a focus among researchers on the different aspects of synthetic facial expressions. With regard to modeling emotional expressions, the contributions of discrete, dimensional and componential theories are not mutually exclusive but are complimentary to each other. Discrete emotion theory provides highly concrete predictions for a specific number of facial expressions. Their universality is a characteristic that has been particularly well studied for creating synthetic expressions given that the objective is to generate largely recognizable expressions. These predictions have been applied to emotion generation in different emotional agents [OST 02, PAN 03, RUT 03]. Furthermore, the unitary nature of these expressions is attractive for modeling due to its simplicity. The theory predicts a continuous development of the expression, a synchrony with one start, an apex reached at the same time by all the AUs and a shared end.

Despite their ease of categorization the expressions defined by this theory remain too simplified and it is therefore necessary to formulate other theories and complimentary studies.

Facial expressions can also be generated based on various dimensions, the most well-known being those of valence and activation. Multi-dimensional spaces can be constructed from the six basic expressions that allow a continuity of expression between different discrete emotions by applying changes linked to each dimension. It is also possible to create facial expressions by not defining the state's position on different dimensions. While the final expressions do not lead to an easy categorization in terms of verbal labels, the possibility of easily generating a large number of different expressions is an important advantage.

The componential approach also allows us to create a large number of emotional expressions by modifying their different components and their temporal development. With regard to the expression of these emotional components, some researchers have divided the face into several parts, each expressing different emotions. Others have focused on the possible sequences of micro-expressions. This method defines expressive behaviors that can appear during a particular emotional state by avoiding repetitive manifestations by the same fixed behavior. This repetition is therefore a weakness of a number of other algorithms determining expressive behaviors.

It is essential to move beyond frozen prototypical expressions and to move towards manipulating the temporal aspects of different elements contributing to expression. The face is subtle means of communication whose different aspects must be animated precisely and in harmony in addition to other modalities, such as bodily movements or even prosody.

6.5. Acknowledgements

We would like to acknowledge that the research presented in this chapter has been partly funded by the EU Framework Program CALLAS grant agreement N 34800.

6.6. Bibliography

[ALB 05] ALBRECHT I., SCHRÖDER M., HABER J., SEIDEL H., "Mixed feelings: expression of non-basic emotions in a muscle-based talking head", *Special issue of Journal of Virtual Reality on Language, Speech & Gesture*, vol. 8, no. 4, pp. 201-212, 2005.

[BEC 05] BECKER C., WACHSMUTH I., PRENDINGER H., ISHIZUKA M., "Evaluating affective feedback of the 3D agent Max in a competitive cards game", TAO J., TAN T., PICARD R. W., (Eds.), *Proceedings of the International Conference on Affective Computing and Intelligent Interaction (ACII)*, Beijing, Springer, pp. 466-473, 2005.

[BEV 07] BEVACQUA E., MANCINI M., NIEWIADOMSKI R., PELACHAUD C., "An expressive ECA showing complex emotions", *Proceedings of the AISB Annual Convention*, Newcastle, pp. 208-216, 2007.

[BRA 05] BRAVE S., NASS C., HUTCHINSON K., "Computers that care: investigating the effects of orientation of emotion exhibited by an embodied computer agent", *International Journal of Human-Computer Studies*, vol. 62, pp. 161-178, 2005.

[BRE 03] BREAZEAL C., "Emotion and sociable humanoid robots", *International Journal of Human-Computer Studies*, vol. 59, pp. 119-155, 2003.

[BUI 04] BUI T. D., Emotions and facial expressions for embodied agents, PhD thesis, University of Twente, Department of Computer Science, 2004.

[COW 99] COWIE R., D OUGLAS -COWIE E., A POLLONI B., TAYLOR J., ROMANO A., F ELLENZW., "What a neural net needs to know about emotionwords", MASTORAKIS N., (Ed.), *Computational Intelligence and Applications. Word Scientific Engineering Society*, pp. 109–114, Society Press, 1999.

[DAR 72] DARWIN C., *The Expression of the Emotions in Man and Animals*, John Murray, London, 1872.

[DUC 99] DUCHENNE G. B. A., *The Mechanism of Human Facial Expression*, CUTHBERTSON R. A., (Ed. & Trans), Cambridge University Press, Cambridge, (Original work published 1876), 1999.

[EKM 69] EKMAN P., FRIESEN W., "The repertoire of nonverbal behavior's: categories, origins, usage and coding", *Semiotica*, vol. 1, pp. 49-98, 1969.

[EKM 72] EKMAN P., "Universals and cultural differences in facial expression of emotion", in: COLE R., (Ed.), *Nebraska Symposium on Motivation*, University of Nebraska Press, Lincoln, pp. 207-283, 1972.

[EKM 75] EKMAN P., FRIESEN W., *Unmasking the Face. A Guide to Recognizing Emotions from Facial Clues*, Prentice-Hall, Englewood Cliffs, 1975.

[EKM 78] EKMAN P., FRIESEN W., *Facial Action Coding System*, Consulting Phsychologists Press, Palo Alto, 1978.

[EKM 82] EKMAN P., FRIESEN W., "Felt, false, miserable smiles", *Journal of Nonverbal Behavior*, vol. 6, no. 4, pp. 238-251, 1982.

[EKM 84] EKMAN P., "Expression and the nature of emotion", SCHERER K., EKMAN P., (Eds.), *Approaches to Emotion*, pp. 319-344, Lawrence Erlbaum, Hillsdale, 1984.

[EKM 89] EKMAN P., "The argument and evidence about universals in facial expressions of emotion", in: MANSTEAD A., WAGNER H., (Eds.), *Handbook of Social Psychophysiology*, p. 485-498, Wiley, Chichester, 1989.

[EKM 92] EKMAN P., "An argument for basic emotions", *Cognition and Emotion*, vol. 6, pp. 169-200, 1992.

[EKM 03a] EKMAN P., "Darwin, deception, and facial expression", *Ann. N.Y. Acad. Sci.*, vol. 1000, pp. 205-221, 2003.

[EKM 03b] EKMAN P., *Emotions Revealed: Recognizing Faces and Feelings to Improve Communication and Emotional Life*, Times Books, 2003.

[ELL 97] ELLIOTT C., "I picked up Catapia and other stories: A multimodal approach to expressivity for 'emotionally intelligent' agents", *Proceedings of the First International Conference on Autonomous Agents*, Marina del Rey, pp. 451-457, 1997.

[ELL 03] ELLSWORTH P., SCHERER K., "Children's ability to control the facial expression of laughter and smiling: knowledge and behavior", in: DAVIDSON R. J., GOLDSMITH H., SCHERER K. R., (Eds.), *Handbook of the Affective Sciences*, pp. 572-595, Oxford University Press, New York, 2003.

[FON 07] FONTAINE J. R., SCHERER K. R., ROESCH E. B., ELLSWORTH P., "The world of emotion is not two-dimensional", *Psychological Science*, vol. 13, pp. 1050-1057, 2007.

[FRA 95] FRANK M., EKMAN P., FRIESEN W., "Behavioral Markers and Recognizability of the Smile of Enjoyment", in: EKMAN P., ROSENBERG E., (Eds.), *What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS)*,pp. 217-238, Oxford University Press, New York, 1995.

[GOS 95] GOSSELIN P., KIROUAC G., DORÉ F., "Components and Recognition of Facial Expression in the Communication of Emotion by Actors", in: EKMAN P., ROSENBERG E., (Eds.), *What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS)*, pp. 243-267, Oxford University Press, New York,1995.

[GRA 06] GRAMMER K., OBERZAUCHER E., "The reconstruction of facial expressions in embodied systems", *ZiF: Mitteilungen*, vol. 2, pp. 14-31, 2006.

[HAI 99] HAIDT J., KELTNER D., "Culture and facial expression: Open-ended methods find more expressions and a gradient of recognition", *Cognition and Emotion*, vol. 13, no. 3, pp. 225-266, 1999.

[HAR 96] HARRIGAN J. A., O'CONNELL D. M., "Facial movements during anxiety states", *Personality and Individual Differences*, vol. 21, pp. 205-211, 1996.

[IZA 71] IZARD C., *The Face of Emotion*, Appleton-Century-Crofts, New York, 1971.

[IZA 93] IZARD C., "Four systems for emotion activation: Cognitive and noncognitive processes", *Psychological Review*, vol. 100, pp. 68-90, 1993.

[KAI 98] KAISER S., SCHERER K. R., *Models of 'Normal' Emotions Applied to Facial and Vocal Expressions in Clinical Disorders, Emotions in Psychopathology*, Oxford University Press, New York, 1998.

[KAI 01] KAISER S., WEHRLE T., "Facial expressions as indicators of appraisal processes", in: SCHRERER K. R., SCHORR A., JOHNSTONE T., (Eds.), *Appraisal Processes in Emotions: Theory, Methods, Research*, pp. 285-300, Oxford University Press, New York, 2001.

[KEL 95] KELTNER D., "Signs of appeasement: Evidence for the distinct displays of embarrassment, amusement, and shame", *Journal of Personality and Social Psychology*, vol. 68, pp. 441-454, 1995.

[KEL 97] KELTNER D., BUSWELL B., "Embarrassment: its distinct form and appeasement functions", *Psychological Bulletin*, vol. 122, pp. 250-270, 1997.

[MAL 09] MALATESTA L., RAOUZAIOU A., KARPOUZIS K., KOLLIAS S., "MPEG-4 facial expression synthesis", in: *Personal and Ubiquitous Computing, Special issue on Emerging Multimodal Interfaces*, Vol. 13, no. 1, pp. 77-83, January 2009.

[MAN 05] MANSTEAD A., FISCHER A., JAKOBS E., "The social and emotional functions of facial displays", in: PHILIPPOT P., FELDMAN R., COATS E., (Eds.), *The Social Context of Nonverbal Behavior (Studies in Emotion and Social Interaction)*, pp. 287-316, Cambridge University Press, New York, 2005.

[MEH 80] MEHRABIAN A., *Basic Dimensions for a General Psychological Theory: Implications for Personality, Social, Environmental, and Developmental Studies*, Oelgeschlager, Gunn & Hain, Cambridge, MA, 1980.

[MEH 86] MEHRABIAN A., *Eating Characteristics and Temperament: General Measures and Interrelationships*, Springer Verlag, New York, 1986.

[MEH 87] MEHRABIAN A., DE WETTER R., "Experimental test of an emotion- based approach to fitting brand names to products", *Journal of Applied Psychology*, vol. 72, pp. 125-130, 1987.

[MEH 95] MEHRABIAN A., "Framework for a comprehensive description and measurement of emotional states", *Genetic, Social, and General Psychology Monographs*, vol. 121, no. 3, pp. 339-361, 1995.

[MEH 96] MEHRABIAN A., "Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament", *Current Psychology: Developmental, Learning, Personality, Social*, vol. 14, pp. 261-292, 1996.

[MEH 97] MEHRABIAN A., "Analysis of affiliation-related traits in terms of the PAD temperament model", *Journal of Psychology*, vol. 131, pp. 101-117, 1997.

[NIE 07a] NIEWIADOMSKI R., A model of complex facial expressions in interpersonal relations for animated agents, PhD thesis, University of Perugia, 2007.

[NIE 07b] NIEWIADOMSKI R., PELACHAUD C., "Fuzzy similarity of facial expressions of embodied agents", in: PELACHAUD C., MARTIN J.-C., ANDRÉ E., CHOLLET G., KARPOUZIS, K., PELÉ D., (Eds.), *Proceedings of the 7th International Conference on Intelligent Virtual Agents (IVA)*, Springer, pp. 86-98, 2007.

[NIE 07c] NIEWIADOMSKI R., PELACHAUD C., "Model of facial expressions management for an embodied conversational agent", in: PAIVA A., PRADA R., PICARD R.W., (Eds.), *Second International Conference on Affective Computing and Intelligent Interaction (ACII)*, Springer, pp. 12-23, 2007.

[NIE 11] NIEWIADOMSKI, R., HYNIEWSKA, S., PELACHAUD, C., "Constraint-based model for synthesis of multimodal sequential expressions of emotions", *IEEE Transactions in Affective Computing*, 2011.

[ORT 90] ORTONY A., TURNER T. J., "What's basic about basic emotions?", *Psychological Review*, vol. 97, no. 3, pp. 315-331, July 1990.

[OST 02] OSTERMANN J., "Face animation in MPEG-4", in: PANDZIC I., FORCHHEIMER R., (Eds.), *MPEG-4 Facial Animation – The Standard Implementation and Applications*, pp. 17-55, Wiley, 2002.

[PAL 06] PALEARI M., LISETTI C. L., "Psychologically grounded avatars expressions", *1st Workshop on Emotion and Computing at KI 2006, 29th Annual Conference on Artificial Intelligence*, Bremen, 14-19 June 2006.

[PAN 07] PAN X., GILLIES M., SEZGIN T. M., LOSCOS C., "Expressing complex mental states through facial expressions", TAO J., TAN T., PICARD R. W., (Eds.), *Proceedings of the International Conference on Affective Computing and Intelligent Interaction (ACII)*, Springer, pp. 745-746, 2007.

[PAN 03] PANDZIC I., FORCHHEIMER R., (Eds.), *MPEG-4 Facial Animation: The Standard, Implementation and Applications*, John Wiley & Sons, New York, 2003.

[PLU 80] PLUTCHIK R., "A general psychoevolutionary theory of emotion" in: PLUTCHIK R. & KELLERMAN H., *Emotion: Theory, Research, and Experience: Vol. 1. Theories of Emotion*, pp. 3-33,, Academic Press, New York, 1980.

[POL 01] POLLICK F., PATERSON H.M., BRUDERLIN A., SANFORD A.J., "Perceiving affect from arm movement", *Cognition*, vol. 82, pp. 51-61, 2001.

[REE 96] REEVES B., NASS C., *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*, Cambridge University Press, Cambridge, 1996.

[REH 05] REHM M., ANDRÉ E., "Catch me if you can – Exploring lying agents in social settings", in: DIGNUM F., DIGNUM V., KOENIG S., KRAUS S., SINGH M. P.,WOOLDRIDGE M., (Eds.), *Proceedings of International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, ACM Press, Utrecht, pp. 937-944, 2005.

[ROS 01] ROSEMAN I. J., SMITH C. A., "Appraisal theory: Overview, assumptions, varieties, controversies", SCHERER K., SCHORR A., T.J., (Eds.), *Appraisal Processes in Emotion: Theory, Methods, Research*, pp. 68-91, Oxford University Press, New York, 2001.

[RUS 80] RUSSELL J., "A circumplex model of affect", *Journal of Personality and Social Psychology*, vol. 39, pp. 1161-1178, 1980.

[RUS 91] RUSSELL J., "In defense of a prototype approach to emotion concepts", *Journal of Personality and Social Psychology*, vol. 60, pp. 425-438, 1991.

[RUS 97] RUSSELL J., FERNANDEZ-DOLS J., "What does a facial expression mean?", in: RUSSELL J., FERNANDEZ-DOLS J., (Eds.), *The Psychology of Facial Expression*, Cambridge University Press, Cambridge, 1997.

[RUT 03] RUTTKAY Z., NOOT H., HAGEN P. T., "Emotion disc and emotion squares: tools to explore the facial expression face", *Computer Graphics Forum*, vol. 22, no. 1, pp. 49-53, 2003.

[SCH 84] SCHERER K., "Emotion as a multicomponent process: A model and some crosscultural data", in: SHAVER P., (Ed.), *Review of Personality and Social Psychology*, pp. 37-63, Sage, Beverly Hills, 1984.

[SCH 87] SCHERER K. R., "Toward a dynamic theory of emotion: The component process model of affective states", *Geneva Studies in Emotion and Communication*, vol. 1, no. 1, pp. 1-98, 1987.

[SCH 01] SCHERER K. R., "Appraisal considered as a process of multilevel sequential checking", in: SCHERER K., SCHORR A., JONHSTONE T., (Eds.), *Appraisal Processes in Emotion: Theory, Methods, Research*, p. 92-119, Oxford University Press, Oxford, 2001.

[SCH 07] SCHERER K. R., ELLGRING H., "Multimodal expression of emotion: Affect programs or componential appraisal patterns?", *Emotion*, vol. 7, pp. 158-171, 2007.

[SCH 06] SCHILBACH L., WOHLSCHLÄGER A., KRÄMER N., NEWEN A., SHAH N., FINK G., VOGELEY K., "Being with virtual others: Neural correlates of social interaction", *Neuropsycholgia*, vol. 44, pp. 718-730, 2006.

[SCH 52] SCHLOSBERG H., "The description of facial expression in terms of two dimensions", *Journal of Experimental Psychology*, vol. 44, no. 4, pp. 229-237, 1952.

[SHI 03] SHIOTA M. N., CAMPOS B., KELTNER D., "The faces of positive emotion: Prototype displays of awe, amusement, and pride", *Annals of the New York Academy of Sciences*, 1000, pp. 296-299, 2003.

[SMI 85] SMITH C., ELLSWORTH P., "Patterns of cognitive appraisal in emotion", *Journal of Personality and Social Psychology*, vol. 48, pp. 813-838, 1985.

[SOR 06] DE SORIANO M. Y., FOXALL G., "The emotional power of place: the fall and rise of dominance in retail research", *Journal of Retailing and Consumer Services*, vol. 13, pp. 403-416, 2006.

[TOM 62] TOMKINS S., *Affect, Imagery, Consciousness: Vol I: The Positive Affects*, Springer, New York, 1962.

[TOM 63] TOMKINS S., *Affect, Imagery, Consciousness: Vol. II. The Negative Affects*, Springer, New York, 1963.

[TOM 84] TOMKINS S., "Affect theory", in: SCHERER K. R., EKMAN P., (Eds.), *Approaches to Emotion*, pp. 163-196, Erlbaum, Hillsdale, 1984.

[TSA 02] TSAPATSOULIS N., RAOUZAIOU A., KOLLIAS S., CROWIE R., DOUGLAS-COWIE E., "Emotion recognition and synthesis based on MPEG-4 FAPs", in: PANDZIC I., FORCHHEIMER R., (Eds.), *MPEG-4 Facial Animation – The Standard, Implementations, Applications*, pp. 141-168, John Wiley & Sons, Chichester, 2002.

[WAL 94] WALKER J., SPROULL L., SUBRAMANI R., "Using a human face in an interface", *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Celebrating Interdependence*, Boston, pp. 85-91, 1994.

[WAL 98] WALLBOTT H., "Bodily expression of emotion", *European Journal of Social Psychology*, vol. 28, pp. 879-896, 1998.

[WEH 00] WEHRLE T., KAISER S., SCHMIDT S., SCHERER K., "Studying the dynamics of emotional expression via synthesized facial muscle movements", *Journal of Personality and Social Psychology*, vol. 78, no. 1, pp. 105-119, 2000.

[WHI 89] WHISSELL C., "The dictionary of affect in language", in: PLUTCHIK R., KELLERMAN H., (Eds.), *Emotion: Theory, Research, and Experience, Volume 4: The Measurement of Emotions, Academic Press*, San Diego, p. 113-131, 1989.

Chapter 7

Emotion Perception and Recognition

7.1. Introduction

Human beings play a vital role in the perception and categorization of emotions using audio, visual and audiovisual stimuli. When implementing an automatic emotion-recognition system, humans are used as a reference point to validate annotation schemes, determine characteristic emotion cues and provide a basis for evaluating automatic classifications. In recent years, parallel to the development of verbal communication technologies, such as automatic speech recognition, interest in human perception has witnessed a revival. Similarly, in connection with the emergence of these fields, there has been a renewed interest in perceptual mechanism understanding. It is worth noting that human perception has served as a reference point that allows us to measure the performances of different automatic systems, as demonstrated by comparative studies into speech transcription [LIP 03, SCH 07] and language identification [MAD 02, MUT 94].

In the field of speech technology, affective computing [PIC 97] (or emotion-oriented computing [SCH 06a]) is a relatively recent development. However, the study of emotions in relation to speech technologies is a less recent subject when considering, for example, the difficulties involved in automatically transcribing emotional speech [ATH 05]. As the affective sciences have become a discipline in themselves over the past decade, emotional phenomena and their physical

---

Chapter written by Ioana VASILESCU.

Emotion-Oriented Systems Edited by Catherine Pelachaud

manifestations have also developed as a topic of research. Research in affective science has moved from merely building corpora illustrating different aspects of emotional phenomena to developing automatic recognition systems based on one or several modalities (mainly verbal and/or gestural)[1].

Emotion-oriented computational studies have focused on different perceptual experiments on the assumption that humans are continually confronted with emotions conveyed by one or more means in their daily interactions (e.g. audio or visual) and that they can manage these emotions effectively for communication. The aim of this research lies in the fact that without a wider communicative context or information from other modalities, humans identify emotions with a significantly higher success rate than that merely attributable to chance [BAN 96, SCH 03]. As a result, if we consider the performance of current automatic emotion-recognition systems for speech, there is an obvious contradiction between emotional states and the ease with which humans categorize emotions when given the same information as an automatic system[2]. Humans have therefore been proposed as a reference point, both in terms of emotion categorization based on stimuli from one or several modalities and with regard to strategies for effective classification[3].

In this chapter we will focus on the various studies into the categorization of verbally expressed emotional phenomena found in recorded speech samples. The research that we will examine has been carried out with the aim of developing systems designed to use this modality, i.e. automatic recognition systems for emotions in speech. However, since some of the studies we will examine are multimodal, this will inevitably involve more modalities than just speech. The perception tests conducted in this framework have been organized around the different development phases of emotion-recognition systems. These consist of three essential stages:

– the acquisition and validation of emotional material (i.e. building and annotating a corpus);

– the identification of objective parameters characterizing the different emotions expressed in the voice and distinguishing them from other emotions; and finally,

– the development of automatic classifiers.

There have been numerous studies seeking to support these steps, mainly focusing on validating emotional content and identifying emotion-characterizing

1 The work of the HUMAINENetwork of Excellence illustrates the various aspects of affective sciences: http://emotion-research.net/.

2 Chapter 5 provides an overview of automatic recognition systems for emotions in speech.

3 When recognizing emotions in speech, humans make use of linguistic parameters such as semantics and syntax, etc., as well as acoustic information.

parameters. We will begin by examining the work in these fields (section 7.2), followed by an overview of the various perception tests conducted in relation to automatic emotion-oriented systems. Section 7.4 will conclude this chapter by examining what we can learn from this research.

7.2. Perception in vocal communication of emotion

In 2003, Scherer conducted a review of perceptual studies using speech samples to categorize emotions [SCH 03]. This was part of a wider effort to bring together areas of study that traditionally belong to separate scientific communities, including areas such as psychology and automatic speech processing[4]. This is also linked to the new and emerging field of affective sciences.

In his review, Scherer provides an overview of current research into how a speaker's emotions are manifested by the voice using the Brunswickian perception technique, also called a "lens model" as a base of research on the vocal communication of emotions. He examined research into the coding (expressions expression) transmission and decoding (impressions) of emotions in speech. The studies examined were positioned in relation to the two poles of the Brunswickian model (i.e. encoding versus decoding) and their main elements. These consist of the speaker's emotional state, the inference made by the hearer about this state and the audio cues that allow us to make the link between the speech signal and a given emotion.

Scherer's review focused on 30 studies since the 1980s into perceptual emotion recognition in speech. These studies show that humans correctly identify emotions at a rate of around 60%, which is much higher than if this was by mere chance. Scherer emphasizes, however, that this result depends strongly on the type of data used (i.e. whether acted, induced or natural[5]). In most studies, acted data have been used, although more recently natural corpora, recorded in real conditions, have been included[6].

Scherer also highlights that there are a number of studies focusing on emotion *discrimination* (i.e. decoding between alternatives) rather than *recognition* (i.e.

4 There is a special issue of *Speech Communication* dedicated to studies of emotion in speech and automatic processing 2003; vol. 40 (1-2).

5 Defining and classifying emotions remains an on-going challenge for those researching emotions, whether they are psychologists, neurologists, linguists or − more recently − affective scientists. This issue is discussed in Chapters 3 and 4 in reference to Scherer, whose work has been a major contribution to the field.

6 More recent studies using natural data have shown that classification scores depend both on the diversity and complexity of the emotions recorded in a corpus [CLA 07].

recognizing a particular category in its own right) in the sense that the majority of experimental paradigms are constructed as follows: given a predefined set of emotions, speech samples are played to listeners who must then label the sample with an emotion from the list. This type of approach, called *discrimination*, highlights the issue of accuracy in relation to chance. This is because, even though among the studies cited, a number of them use statistical credibility measures that are evaluated against chance. It is therefore difficult to evaluate them against a freely chosen experimental configuration.

A rate of around 60% correct emotion classification is mostly found with corpora composed of acted data. This result of 60% correct classification, despite the limitations just mentioned, allows us to make some inferences about the ability of humans to classify emotional speech. This ability seems to be above chance in various experiments involving both natural and modified speech (using speech synthesizing or resynthesizing techniques).

7.3. Experimental paradigms and emotion-oriented automatic systems

This chapter is based on more than 30 years of study into perceptual emotion categorization using speech samples. The majority of this research has been conducted over the past 10 years as the affective sciences have seen a substantial expansion. These studies have mostly been presented at speech and automatic processing conferences (e.g. Interspeech, Speech Prosody, ICPhS (International Congress of Phonetic Sciences) and LREC (International Conference on Language Resources and Evaluation)[7]). Their very number highlights both the current interest in studying emotions in automatic speech processing and the significance attached to human perception that acts as a reference point for categorizing emotional phenomena.

The following sections will focus on the main experimental paradigms used for automatic processing of emotions in speech. These include paradigms for validating emotional content and/or corpora annotation strategies (see section 7.3.1), paradigms for validating objectives and quantifiable parameters of emotional information (see section 7.3.2) and, finally, paradigms designed to compare humans and machines in terms of emotion recognition (see section 7.3.3). This chapter will focus primarily on this first paradigm, since building corpora is an essential stage in the automatic emotional categorization process. The type of corpus used and how it is annotated are crucial for defining and measuring characteristic parameters carrying emotion information and constructing emotion-recognition systems.

---

7 See: http://www.interspeech2011.org, http://www.speechprosody2010.illinois.edu/, http://www.icphs2011.hk, http://www.lrec-conf.org.

7.3.1. *Experiments validating emotional content and/or annotation strategies*

Of the perceptual studies into automatic emotional speech processing, a significant number are dedicated to pre-processing corpora and, more specifically, verifying the presence of different and identifiable non-neutral emotional states found throughout the corpus or even validating the corpus' emotion labeling strategy. These studies into decoding the emotional signals in speech illustrate a continuing theme in emotional speech research, confirming the Scherer's observation [SCH 03]. In effect, before we can proceed to automatically classifying the speech signals obtained from segmentation, it is necessary to ensure that the corpus conveys different emotions that can be categorized according to a limited number of classes. These classes are the result of an annotation process based on various strategies, conducted by a set number of experts.

The annotation process is often accompanied by perceptual validation studies into emotional content. These studies can be carried out either before annotation or, more often, after annotation when the classes identified by the annotators are verified by a statistically representative population sample. This stage is justified by the time and costly nature of annotation when conducted by several annotators. This is because authors prefer to use an annotation carried out by a limited number of annotators who will be subsequently confirmed by validation techniques. Two complementary methods can therefore be used to validate the annotation strategy. The first consists of estimating the level of agreement between the annotators and the kappa coefficients [COH 60] or Cronbach's alpha coefficient. The second is a perceptual validation of a sample sub-set that is representative of a corpus using a population of listeners [CLA 07]. This methodology has also been employed by Devillers and Vasilescu [DEV 03a].

Human perception can therefore be used as reference point, since it allows research to gain a certain level of objectivity with regard to the emotional content of the corpus. This may also lead to new annotations that call into question some of the initial annotators' choices. Following this, any classification from the automatic processing stage is evaluated in relation to this new annotation.

7.3.1.1. *Experimental design and stimuli*

The perception studies that we examine in this section share some common characteristics. For example, experimentation usually takes place following the acquisition, segmentation and annotation phases of the audio corpus by a small number of experts. Segmentation involves attaching emotional labels to audio extracts of varying length that are determined by the characteristic of the corpus and its desired purpose. These initial labels are attached to speech segments and are produced by researchers and/or human experts.

Once these initial stages have been carried out, a number of speech samples are randomly selected by the researchers to represent the range of emotions illustrated in the corpus. A good example of this is dialog corpora recorded in call centers which is often broken down into individual “speaker turns”. The emotional label and the speech samples correspond to one speaker turn by a given locutor [DEV 03a, DEV 07, VID 07], although the sample can be larger or smaller than this. The larger speaker turn unit is the sequence, the smaller speaker turn is the audio chunk. For example, Clavel [CLA 07] has used “audio chunk”, a segment illustrating a single context likely to cause emotional manifestations. A segment was also used, however, as the minimum emotional unit found when several emotions characterize the same speaker turn. It should therefore be noted that perceptual judgment is strongly dependent on the data being used and the aim of the automatic classification being undertaken.

7.3.1.2. *Test populations*

With regard to *test populations*, it should be noted that it is often difficult to differentiate between the annotations themselves and their verification. This is because, according to various studies, emotional content can be verified by two [LEE 01, MOR 06] or even 20 or more listeners [CLA 04, DEV 03c, DEV 05, HER 02]. The difference between studies using fewer listeners (i.e. less than 10) and those using several dozen subjects with the aim of statistical validation allows us to differentiate between “emotional content validation” versus “annotation validation” studies, although the difference between them is not fixed. Validating annotations means that we can verify the interpretations of the corpus’ emotional content in a wider population. Furthermore, by using a perceptual task where the subjects reproduce set phrases on a short test sample, the researchers can statistically verify emotional content as well as gain an objectification of the emotional labels attached by a small number of annotators. Abrilian *et al*. [ABR 06] have varied the number of listeners to between three and forty, showing that the categorizations of emotional content tend to stabilize at around 10 listeners.

7.3.1.3. *Perceptual categorization into emotional classes*

The task of *perceptual categorization* often involves classifying speech samples according to an obligatory choice based on a limited number of emotions or affective states. More often than not, this consists of classifying these emotions into:

– two groups (e.g. positive versus negative emotions);

– four groups (the basic emotions *fear, anger, happiness* and *neutral*); or

– six groups (adding two other emotions of negative valence to the other four emotions e.g. *sadness* and *disgust*, discrete emotions or more complex categorizations).

Categorizations into more than six classes of emotions are often accompanied by meta-classification into, for example, primary versus secondary emotions or even primary versus mixed emotions [CLA 07, CRA 04, DEV 05]. More rarely researchers have used free categorization, often as a second task in the experiment after an initial forced classification [CLA 04, EHR 02].

With regard to annotation validation methods, perceptual categorization by groups of listeners depends strongly on the *annotation strategy* employed. As demonstrated in Chapter 3, annotation strategies are based on emotion representation theories. As a result, perceptual paradigms reflect the different approaches to categorizing emotions. This primarily includes discrete emotion theories [COW 00, EKM 92, ORT 03], which suppose that there is a link between an emotional label and audio stimulus. This theory also supports the notion of emotions being represented by abstract dimensions [COW 03, OSG 75][8].

The majority of perceptual studies set out experimental paradigms based on annotations according to discrete categories. Some research has also used both theories and proposed complex tasks in which subjects are asked to categorize emotions into discrete categories and evaluate the same stimuli in agreement with the representation of emotions on abstract dimensions [CLA 04, CRA 04, DEV 03a, DEV 06]. The three most commonly used abstract dimensions are those described by Osgood *et al*. [OSG 75] including appraisal or "valence" (positive versus negative), arousal (active versus passive) and power (strong versus weak). The valence dimension is generally used by all researchers because it is perceptually intuitive. The arousal dimension is called "intensity" in some research [CLA 07, CRA 04]; similarly the "power" dimension is called "reactivity" by Clavel [CLA 07]. Coding according to "reactivity" has proven to be more applicable to the needs of the SAFE (Situation Analysis in a Fictional and Emotional Corpus (developed by Clavel)) designed for public surveillance. This involves evaluating reactions to threats when a speaker feels fear in an abnormal situation. It should nevertheless be noted that perceptual experiments require an emotion coding scheme where abstract dimensions confirm the ability of listeners to coherently evaluate the emotional content conveyed by the voice. Even if this evaluation is not intuitive, i.e. according to discrete labels, the scores prove to be better than that of chance.

7.3.1.4. *Choosing emotional labels*

With regards to the emotional labels used in perceptual classification, the number and type vary according to several factors − namely the type of corpus, the level of analytical detail of the analysis (complimentary criteria as well as the

---

8 Chapter 1 details emotion representation theories. Some of these theories are also discussed in Chapter 3 in relation to data acquisition and modeling. In this section we will briefly examine the theories that have paved the way for perceptual studies.

difficulty of human and automatic classification) and the aim of the study. The type of corpus (i.e. acted, induced or natural) is the first factor influencing the emotional classes being interpreted. For example, the more natural the corpus (i.e. using data obtained in real conditions); the greater the probability that emotions other than the six "basic" will be present. These complex emotions that can be mixed with other emotions and attitudes are aspects that complicate perceptual classification [SCH 03][9].

The level of the analysis' specificity is in direct relation to the performance of current classification systems. At this current time, while applicative needs require detailed classification, automatic systems often use binary classifications of positive versus negative emotions or even negative versus non-negative and emotion versus no emotion [CLA 08, DEV 05, DEV 07, HIR 05, MOR 06]. Therefore perceptual validation serves to confirm the presence of at least two classes of emotions (e.g. emotion versus no emotion) but further shows that humans interpret more subtle distinctions within these classes. This is particularly applicable to natural corpora that use data recorded in real conditions outside of the laboratory. This is because with corpora of acted data (i.e. recorded in laboratory conditions with humans simulating emotions out of context), perceptual validation only consists of recognizing the emotions previously set out by the researchers and actors using a population of statistically representative listeners.

In contrast, natural corpora obtained *in situ* have a tendency to show a large amount of variability in terms of potential emotional states. Perceptual verification of this variability is measured by the number, range and difference in perception of emotional classes. Often researchers regroup these emotions into meta-classes after the experiment in order to reduce this variability and, as a result, the complexity of automatic classification stage.

For example, Devillers *et al*. [DEV 03a, DEV 03c] have proposed a fixed choice of 10 classes in a perceptual test to verify the content and emotional labels of a corpus consisting of dialogues recorded in call centers (anger, anxiety, doubt, fear, annoyance, satisfaction, surprise, apologetic, neutral and "I don't know"). Following the experiment, the 10 classes were regrouped into four meta-classes that would be used for automatic classification (fear, anger, satisfaction and neutral).

Some researchers [DEV 05, DEV 06, VID 07] have identified and confirmed 20 more nuanced emotional classes in addition to a neutral state in perceptual tests on CEMO (corpus for naturally-occurring dialogs recorded in a real-life medical call center) data from a medical call center. Following this, the 20 classes were regrouped into seven meta-classes (fear, anger, sadness, pain, surprise, positive and

9 The issue of acquiring data and different types of corpora is discussed in Chapter 3.

neutral) and two valences (negative versus positive), which were used as a reference point in the automatic classification process.

The audiovisual corpus of fictional data, known as SAFE [CLA 06, CLA 08] was designed for public surveillance. While perceptual validation of the content and annotation showed a significant variability in emotional states, Clavel then regrouped the emotions into four meta-classes (negative emotions, fear-like emotions, positive emotions and neutral emotions) and used a sub-corpus of extracts labeled *fear* and *neutral* during the automatic classification process. One of the studies providing the most perceptual classes is that of Ehrette *et al*. [HER 02] who identified 20 classes or perceptual attributes corresponding to four initial speaking styles. In this study, the experiment was carried out on a corpus of telephone messages designed as a vocal server.

The final application motivating automatic classification represents a determining factor in the perceptual paradigm, notably for the number of emotional labels proposed by the listeners during perceptual validation. The various research discussed in this section has been carried out for automatic applications in call centers, for public surveillance or for automatic vocal servers. Different emotional labels aim to detect negative values more often for call center-type applications and surveillance (i.e. detecting that a caller is angry or that an individual is scared in a life-threatening situation), and positive values in vocal servers where pre-recorded messages try to present a company in a positive light.

7.3.1.5. *Discussion*

In this section we have examined how human perception is widely used to confirm emotional states conveyed by the voice and to validate annotation strategies for audio corpora previously constructed by a restricted number of annotators. With regards to validating emotional content, perceptual studies have proven to be a reliable means of confirming the saliency of the emotional classes used. When there are a number of emotional classes (i.e. more than four to six basic emotions), researchers often opt to combine these into meta-classes. This is due to the fact that current automatic classification systems can only distinguish between around two and six different emotions at most [DEV 07]. Human interpretations are therefore used to classify emotional content. This also allows us to account for the needs and limits of automatic systems. Thanks to perceptual tests, it is possible to produce a reference classification that ranges from two to several dozen emotions. This classification can also be hierarchical, e.g. from primary to secondary emotions, or even basic emotions versus mixed emotions.

When validating annotation strategies, the stages using human interpretation can lead to a readjustment in the emotional labels or abstract dimensions set out by the researcher. For example, after a perception test it is possible that some samples previously considered to contain non-neutral emotional content are perceived by listeners as neutral. Some emotional labels can also be reconsidered. For example, following a perception test, Devillers *et al*. [DEV 03c] relabeled "happiness" as "satisfaction" and "anger" as "annoyance". These labels were more applicable to the call center corpora and telephone dialog between agents and clients.

Perceptual tests can also lead to new emotional labels, particularly when the test allows free classification or a listener *feedback* phase [CLA 07]. An explanation for this reorganization of emotional labels could lie in the fact that the lack of context and the use of vocal signals alone are not sufficient to categorize speech samples in detail.

One aspect causing divergence between annotations carried out by a limited number of annotators and validation by several dozen listeners is linked to annotation conditions versus perceptual experience. While annotation is often carried out in context (e.g. within a dialog [DEV 03c, DEV 06] or a film sequence [CLA 07]) a perception test places samples outside of context. For example, Clavel *et al*. [CLA 04, CLA 08] have demonstrated differences in emotion perception depending on whether the context is present or absent. In these two studies, the researchers defined "context" as the role of the visual modality in (i.e. whether this is present or absent) the longer sentence from which the sample has been extracted. This is illustrated by Figures 7.1 and 7.2, which show the distribution of degrees of intensity and reactivity by listeners according to the test conditions (e.g. normal versus abnormal situations, i.e. contexts that could endanger life and cause fear-like emotions, with or without video context).

Following the perception test where this factor is varied (presence versus lack of video), it was noted that the intensity scale is used differently according to whether the emotion is normal or abnormal. Furthermore, some emotions are perceived as more intense in a normal situation when the subjects have access to the images accompanying the sounds in order to make a judgment (Figure 7.1). In abnormal situations audio signals are more robust, whereas in normal situations they seem to be less efficient for evaluating an emotion's degree of intensity (i.e. the curve of the two conditions +/– video have a greater proximity for segments in abnormal situations than for emotions in normal situations).

When annotation is done according to abstract dimensions, the perception test can lead to a readjustment of the degree scales. From the same perception test, Clavel has readjusted the scale used for the "reactivity" dimension (–3 to 3, replaced by 0 to 3) [CLA 07]. This measure and new scale proposed were better adapted to the corpus and were deemed to be more intuitive (Figure 7.2.).

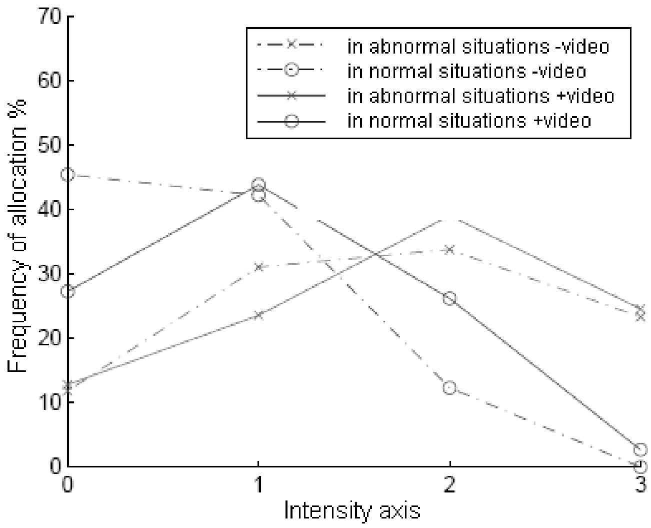


Figure 7.1. *Allocation of degrees of intensity in normal and abnormal situations with and without video context (adapted from [CLA 07])*

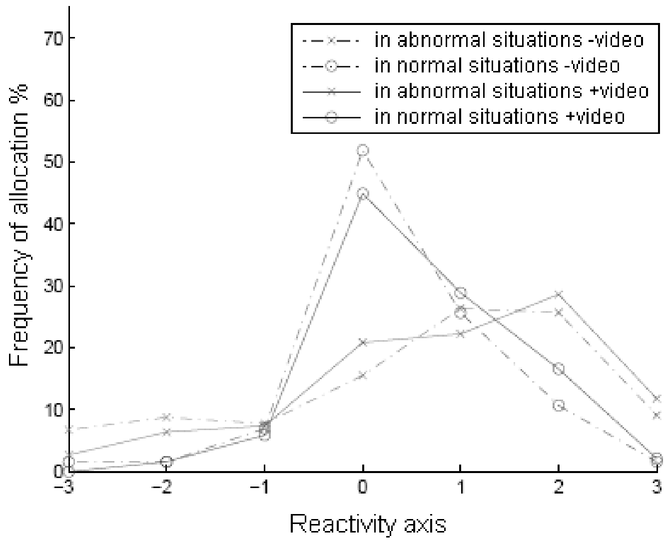


Figure 7.2. *Allocation of the degree of reactivity in normal and abnormal situations with or without video context [CLA 07]*

7.3.2. *Tests for validating measurable parameters of emotional information*

This section focuses on perception tests where the aim is to validate how well measurable speech parameters convey emotional information. Parameters highlighted by perception tests belong to several levels of analysis, specifically acoustic, prosodic and linguistic. This consists of information relating to the spectrum, voiced and unvoiced content of speech, pitch of the voice, energy, voice quality or even lexical level and speech acts. For example, the experimental paradigms used to estimate the role of different audio parameters in emotion recognition are linked to the study of the nature of the underlying voice-emotion inference mechanism (which involves studying how cues produce certain types of emotion inferences in listeners) [SCH 03]. The premise of these studies is that, in daily interactions, humans integrate information from diverse modalities (mainly verbal and gestural). However, experimental categorization of stimuli based on emotional content have shown that even when confronted with a single modality (audio in the case of these studies), humans can correctly classify emotional content with a significantly higher rate than chance.

This observation allows us to hypothesize that, using experimental manipulations it is possible to highlight the most important objective cues characterizing a modality with the final aim obviously being their automatic exploitation. In these studies, the signals have often been manipulated in order to obtain stimuli that preserve or mask certain attributes. For example, Scherer [SCH 03] has employed tests using speech synthesis and resynthesis techniques to preferentially test certain acoustic and prosodic parameters[10]. More recently, other levels of analysis have been highlighted as a potential source of emotional parameters, e.g. voice quality parameters (see Chapter 3).

The majority of perception studies conducted in the past decade have focused on acoustic, prosodic and voice quality parameters found in natural speech or speech modified by synthesis or resynthesis [ARI 07, BAR 07, BUR 00, DEV 03a, EHR 02, GRI 07, IRI 07, LIC 01, MAF 07, MOZ 01, MOZ 02, SCH 06b, SOB 99, YAN 08]. One current experimental procedure consists of evaluating the *natural speech audio parameters* prior to the test according to their statistic association with the judgments of listeners about the emotion perceived to be experienced or simulated.

10 Scherer identifies three experimental paradigms depending on the relationship that the researcher establishes between the objective parameters presented and the listeners: tests establishing a direct relation between acoustic parameters characterizing an emotion/judgment of listeners' perception, emotions masking supposedly robust parameters to ensure their relevance via perception tests, and tests manipulating acoustic parameters by re-synthesizing speech with the aim of arranging the parameters in a hierarchy thanks to perception judgments. For further details, refer to [SCH 03].

This has led to interesting conclusions about the relevant link between a given emotion and a measurable or identified characteristic. The majority of studies show that *acoustic* and *prosodic* parameters, such as formants, fundamental frequency and its variation over time, energy, duration, etc., remain the most robust [CLA 07, DEV 07, HUM 04]. The effects of different emotions on acoustic and prosodic parameters have a physiological basis, since emotional states cause bodily changes that then affect the voice and alter acoustic and prosodic parameters which were previously considered to be neutral. These alterations and the link between alteration and emotion seem to have a perceptual correlation.

More recently, it has been shown that other levels of analysis provide additional important parameters, such as *voice quality*, for example, in parameters such as *jitter* or *shimmer*. Campbell and Mokhtari's precursor study [CAM 03] has already shown the relevance of the normalized amplitude quotient (NAQ) parameter. This parameter measures variations in voice quality, e.g. a "pressed" voice compared to a "breathy" voice that can change in a single locutor depending on the interlocutor, the speaking style used or even the speech act. NAQ has shown that parameters characterizing *tone of voice* are just as important as "classical" prosodic parameters.

The study by Yanushevskaya *et al*. [YAN 08] is a good example of the perceptual correlations of these observations as it combined perception tests and measures of voice quality. The aim of the test was to judge the respective weight of quality (e.g. a cracked, murmuring, strained voice, etc.) versus voice loudness level (estimated according to the variation in intensity). The perception test shows that voice quality is associated more strongly with emotions than with the voice loudness. For example, a whispering voice clearly indicates sadness or annoyance. A study by Grichkovtsova *et al*. [GRI 07] was also based on the role of voice quality in emotion perception.

In terms of *lexical level and dialog patterns*, perceptual studies ([DEV 03c] and [CLA 04]) have shown that listeners recognize the emotions in speech samples based on specific vocabulary or even specific speech acts. It is also possible to indicate anger or fear by using specific words.

Vocabulary is not the only linguistic threshold conveying emotional information. *Backchannels* or *affect bursts* also carry indications about a speaker's emotional state [SCH 06b]. This consists of delexicalized vocalizations used by a speaker to give emotional *feedback* about the informational content of his or her message (e.g. "wow" for admiration or "phew" for relief, etc.). A perception test can confirm the strong link between emotions and *affect bursts* in a cross-cultural context that reiterate the independent nature of some *affect bursts*[11]. Other researchers have focused on the minimum time window needed for the listener to have enough audio information to be able to infer the emotion felt by the speaker that has produced a speech sample [MAF 07]. It has also been shown that information about a speaker and their emotional state accumulates progressively: 250 ms is enough to determine gender, while detecting accent takes around 1.5 seconds and emotional content around 1.7 seconds.

The second experimental procedure leading to perceptual paradigms for evaluating parameters consists of parameter manipulation of speech synthesis–resynthesis. This technique also represents an efficient way of linking some potentially discriminative vocal parameters of emotional states and listeners' perceptions. This consists of modifying some acoustic and prosodic attributes, such as fundamental frequency (*f0*), intensity, energy, etc., and measuring changes in listeners' perceptions allowing us to infer the role of the modified parameter [MOZ 01, MOZ 02, GRI 07, SOB 99]. Among the numerous possible techniques available, the method used by Scherer [SCH 03] is particularly well known for its efficacy. This approach consists of copying the prosody, which then allows a "neutral" speech sample to be used, and varying parameters such as fundamental frequency, intensity and speech rate in order to link particular values of these parameters to a given emotion. As such, it has been observed that a rapid rate can be associated with negative emotions such as anger while a narrow range of (*f0*) − i.e. little variation in this parameter − is perceptually linked to sadness.

To conclude, the *intercultural* aspects of emotion recognition and their objective should also be considered, as has already been outlined by Scherer [SCH 03]. Perception tests on one set of verbal samples played to a range of listeners from different countries (who spoke different languages) and cultures showed that overall the results overlapped. These results suggested a certain universality of some attributes. This overlap has also been noted with speakers of different languages. The listeners were presented with the same audio samples of meaningless phrases, pronounced by an actor to illustrate several emotions that were then translated and recorded in an emotional corpus in the listeners' languages.

Despite this being done, language dependency manifested by variability in the perception of some emotions and their characteristic signals remains to be discussed [BUR 06, SCH 00]. The observations have also been made using both audio samples where the measurable parameters were not changed and where some prosodic and

11 Without a perception study on measures of vocal parameters (in the wider sense), other researchers have also underlined the role of some parameters, such as cries [CLA 08] and verbal disfluencies [DEV 03b] in determining emotional content in speech.

voice quality parameters varied [SCH 00] (level of fundamental frequency, duration and *jitter*) [BUR 06].

7.3.2.1. *Discussion*

In this section we have examined the perceptive paradigms built to evaluate the discriminative power of some objective parameters measured in speech. The parameters discussed here concern acoustics and prosody levels, specifically with regard to traditional parameters such as formants, fundamental frequency, duration, etc. Tests have shown that in addition to this, voice quality and linguistic and vocal parameters also convey emotional information.

Perceptual experimentations into these aspects are often carried out in two experimental frameworks. On the one hand, it involves measuring parameters in natural speech and comparing these measures with the results of perception tests. On the other hand, the parameters are measured by perception tests using stimuli in speech that have been manipulated by speech synthesis–resynthesis techniques. The results show a general correlation between the supposedly important parameters and perceptual categorization. The different parameters evaluated can also be prioritized using perception tests. The final aim of these tests is to measure the discriminative nature of some parameters or descriptors conveying emotional content that can then be used for automatic classification.

7.3.3. *Tests comparing human and automated emotion recognition*

This section will examine studies where the aim has been to compare human and automated classification of emotions conveyed in speech signals. Humans are therefore used as a reference point at the end of the design process for emotion-recognition systems. Given a corpus whose emotional content has been previously validated and an automatic classification method selected, the automatic classification score for the emotional labels allocated during annotation is evaluated using a perception test on the same test set.

This comparative approach suggests that the emerging field of speech emotion recognition proves the need for an evaluation of current performances and that human categorization is seen as a reference point for different systems developed for automatic speech processing. The evaluation process is defined through specific projects for some areas of automatic speech processing, as is the case with automatic speech transcription, for example [PAL 03]. Outside of these evaluations (which have been conducted according to rigorously defined schema involving different actors), comparisons with perceptual processing are often used because they mean we can compare automatic and human performances. Among these comparison strategies, those conducted for automatic speech transcription recognition [LIP 03, SCH 07], or for automatic language identification [MAD 02, MUT 94] are particularly notable.

With regard to automatic emotion recognition, there is an ever-increasing amount of research on this topic because it is at the heart of this field of study. The idea of evaluating current systems performance through evaluation campaigns – as conducted in other related fields such as automatic speech recognition – is beginning to emerge, and has been shown by the HUMAINE network's CEICES initiative [BAT 06]. This is a first step in sharing information on the performance of different emotion-recognition systems based on the same corpus of data recorded in real conditions.

Among the studies into automatic emotion recognition, some have proposed an "internal" evaluation leading to both automatic and perceptual classification that can serve as a reference point [ENO 06, ENO 07, HIR 05, PET 99, POL 98]. These studies are, however, less numerous than those focusing on validating emotional content and annotation (see section 7.3.1) or examining objective emotion characteristic parameters (see section 7.3.2), indicating that the need for evaluation is more recent and less crucial. As a result, the conclusions made here with regard to the success of humans versus automatic systems in recognizing an emotion should be considered with care, given the small number of studies focusing on this issue.

Before examining the results of this comparison, it is first important to focus on the inevitable subjectivity involved in creating automatic classifiers due to the use of human perception to evaluate automatic scores. As in other areas of automatic speech processing, the human reference marked in the annotation conveys a percentage of subjectivity. In areas such as automatic speech transcription, it is possible to speak about error. Human error due to ambiguity in pronunciation can range from 2−5%. In affective sciences, error is more difficult to measure due to the complex nature and ambiguity of emotional phenomena. Consequently annotations, even if multiple, and evaluation using measures of agreement between annotators or perception tests, still have a certain degree of subjectivity.

A good example of this is the work carried out by Clavel [CLA 07], which demonstrates the variation in success rates for automatic emotion-recognition systems depending on reference annotations. Below are the kappa coefficients, k, between the system and the three annotators serving as a reference. The kappa coefficient k measures the level of corrected agreement if based on chance. Kappa k is therefore equal to 0 when the level of agreement is the same as that as obtained by chance. In Table 7.1, the coefficient k between the system and each of the three annotators varies according to the annotator, showing that some annotations are closer to the system's output than others.

System / Ann	κ
Annotator 1 vs System	0.48
Annotator 2 vs System	0.45
Annotator 3 vs System	0.53

Table 7.1. *Rate of corrected agreement, k, between the different annotators Ann1, Ann2 and Ann3 and the automatic classifier (taken from [CLA 07])*

Studies comparing human/machine performance vary according to the complexity of the data being used and the desired aim of the automatic application. Results from tests carried out by Polzin [POL 96], and Petrushin and Waibel [PET 99] show that human performance is comparable to automatic performance where it is a question of recognizing the emotions in a corpus of acted data where the emotional content illustrates four basic emotions (happiness, anger, sadness and fear) and a neutral state. Human and automatic categorizations are around 70%, depending on test conditions. More specifically, humans can correctly categorize the four emotions of the Petrushin and Waibel corpus at a rate of 60%, with errors due to classification as "neutral" with a number of phrases that had previously been labeled as potentially carrying emotional content. The automatic system provides a higher score of around 70% when detecting the four emotions and the neutral state and 77% when identifying "agitation" versus "calm". In the corpus of acted data used by Polzin, the two results − i.e. perceptual and automatic − were comparable and were around 70%.

The studies by Enos *et al*. [ENO 06, ENO 07] and Hirschberg *et al*. [HIR 05] have been used for security projects because they focused on detecting *misleading* speech (i.e. "deceptive speech", according to the authors' own terminology). The corpus of "deceptive speech" is composed of induced data with an acquisition protocol that encourages speakers to lie and indicate when this was the case. A perception test was then carried out to evaluate at what level humans can detect lies, providing lower scores than chance (less than 50%). These rates from humans proved to be less than automatic classification in the present case using machine learning techniques (66.4% when the baseline is 60.2%; higher than chance and perceptual scores).

The researchers in these studies hypothesized that human decoding of lies must be subject to caution whether it is based on "global lies", that is on a wider context as the entire statement is false, or only "local lies" – where the statement is globally true but contains one element of "deception". The tests carried out to date seem to show that machines perform better than people.

7.3.3.1. *Discussion*

This section focused on studies that have evaluated results from automatic emotional speech classification from perception tests to evaluate the same test set. The issue of evaluating automatic emotion classification results being more recent than for other areas of automatic speech processing (such as automatic transcription, where consensus on a "good response" is easier to construct), we have discussed limited number of works outlining the issue. This does not, however, allow us to make any concrete conclusions. It is nevertheless possible to note that, depending on the corpus being used, human success at recognizing emotions is higher or lower than automatic recognition. This suggests that categorizing emotions is a complex task for humans and that continued efforts are needed to create systematic evaluations of work carried out in automatic emotion recognition.

7.4. Conclusion

We have seen in this chapter that human perception is used at all stages of development when creating classification systems for emotions in speech. These phases are based around validating emotional content and annotating corpora, and identifying and validating measureable parameters or descriptors that allow us to link an emotion to an audio extract and estimate the system's success.

This parallel between humans and automatic systems is not new. Other speech-related research, such as automatic transcription or automatic language identification, also uses human perception processing to learn and self-evaluate. An explanation for this lies in the aim of *artificial intelligence* to create technological tools that have the appearance of human intelligence. In this respect, affective sciences form a recent branch of science that has developed to the extent that this field has accepted and scientifically proved the role of emotions in rational decision making, and human intelligence generally [PIC 97]. Creating computers with the ability to identify human emotions and react to them appropriately and constructively is therefore a fully justified aim.

Using the different experimental paradigms presented in this chapter, we have seen that affective sciences essentially rely on subjective content with regard to experimental validation of the different construction stages for an automatic classifier. Human perception remains a highly useful tool for comparing annotators' interpretations, whether at the annotation stage or for testing measurable parameters characterizing emotions. In the absence of real evaluation metrics such as those for automatic speech transcription, human perception provides a primary reference point. Finally, the diversity of perception tests has shown that this technique has improved in accuracy, helping to develop other techniques in closely-related fields, such as experimental psychology. For example, the connection between research into automatic speech transcription and psycholinguistics has enabled a better understanding of the effective perceptual processes that are likely to have automatic encoding [SCH 07].

7.5. Bibliography

[ABR 06] ABRILIAN S., DEVILLERS L., MARTIN J.-C., "Annotation of emotions in real life video interviews: variability between coders", *International Conference on Language Resources and Evaluation LREC*, 24-26 May, Genoa, Italy, 2006.

[ARI 07] ARIMOTO Y., OHNO S., IIDA H., "Acoustic features of anger utterances during natural dialog", *Interspeech*, August 27-31, Antwerp, Belgium, 2007.

[ATH 05] ATHANASELIS T., BAKAMIDIS S., DOLOGLOU I., COWIE R., DOUGLAS-COWIE E., COX C., "ASR for emotional speech: Clarifying the issues and enhancing performance", *Neural Networks*, vol. 18, pp. 437- 444, 2005.

[BAN 96] BANSE R., SCHERER K., "Acoustic profiles in vocal emotion expression", *Journal of Personality and Social Psychology*, vol. 70/3, pp. 614-636, 1996.

[BAR 07] BARRA R., MONTERO J., MACIAS-GUARASA J., GUTIERREZ ARRIOLA J., FERREIROS J., PRADO J., "On the limitations of voice conversion techniques in emotion identification tasks", *Interspeech*, Anvers, 2007.

[BAT 06] BATLINER A., STEIDL S., SCHULLER B., SEPPI D., LASKOWSKI K., VOGT T., DEVILLERS L., VIDRASCU L., AMIR N., KESSOUS L., AHARONSON V., "Combining efforts for improving automatic classification of emotional user states", *IS-LTC*, October 16-17, Ljubljana, Slovenia, 2006.

[BUR 00] BURKHARDT F., SENDLMEIER W., "Verification of acoutical correlates of emotional speech using formant-synthesis", *ISCA Tutorial and Research Workshop (ITRW) on Speech and Emotion*, September 5-7, Newcastle, Northern Ireland, 2000.

[BUR 06] BURKHARDT F., AUDIBERT N., MALATESTA L., TÜRK O., ARSLAN L., AUBERGÉ V., "Emotional prosody – Does culture make a difference?", *Speech Prosody*, May 2-5, Dresden, Germany, 2006.

[CAM 03] CAMPBELL N., MOKHTARI P., "Voice quality: the 4th prosodic dimension", *International Congress of Phonetic Sciences ICPHS*, 3-9 August, Barcelona, Spain, 2003.

[CLA 04] CLAVEL C., I. VASILESCU I., DEVILLERS L., EHRETTE T., "Fiction database for emotion detection in abnormal situations", *International Conference on Spoken Language Processing ICSLP*, October 4-8, Jeju Island, South Korea, 2004.

[CLA 06] CLAVEL C., VASILESCU I., RICHARD G., DEVILLERS L., "De la construction du corpus émotionnel au système de détection: le point de vue applicatif de la surveillance dans les lieux publics", special issue "Interaction émotionnelle", *Revue des Sciences et Technologies de l'Information*, vol. 20, no. 4-5, 2006.

[CLA 07] CLAVEL C., Analyse et Reconnaissance des Manifestatiions Acoustiques des Emotions de Type peur en Situations Anormales, Télécom Paris/ENST, Paris, 2007.

[CLA 08] CLAVEL C., VASILESCU I., DEVILLERS L., RICHARD G., EHRETTE T., "From fear-type emotion recognition toward threat diagnosis in surveillance systems", *Speech Communication*, vol. 36/50, pp. 445-356, 2008.

[COH 60] COHEN J., "A coefficient of agreement for nominal scales", *Educational and Psychological Measurement*, vol. 20, pp. 27-46, 1960.

[COW 00] COWIE R., "Emotional states expressed in speech. In describing the emotional states expressed in speech", *ISCA ITRW on Speech and Emotion: Developing a Conceptual Framework for Research*, pp. 224-231, 2000.

[COW 03] COWIE R., CORNELIUS R., "Describing the emotional states that are expressed in speech", *Speech Communication*, vol. 40, no. 1-2, pp. 5-32, 2003.

[CRA 04] CRAGGS R., WOOD M. M., *A Categorical Annotation Scheme for Emotion in the Linguistic Content of Dialogue*, Springer-Verlag, Berlin-Heidelberg, 2004.

[DEV 03a] DEVILLERS L., VASILESCU I., "Reliability of lexical and prosodic cues in two real-life spoken dialog corpora", *International Conference on Language Resources and Evaluation LREC*, May 26-28, Lisbon, Portugal, 2003.

[DEV 03b] DEVILLERS L., VASILESCU I., LAMEL L., "Emotion detection in a task-oriented dialog corpus", *IEEE International Conference on Multimedia Expo (ICME)*, July 6-9, Baltimore, USA, 2003.

[DEV 03c] DEVILLERS L., VASILESCU I., MATHON C., "Acoustic cues for perceptual emotion detection in task-oriented human-human corpus", *International Congress of Phonetic Sciences ICPhS*, Barcelona, 2003.

[DEV 05] DEVILLERS L., VIDRASCU L., LAMEL. L., "Challenges in real-life emotionannotation and machine learning based detection", *Journal of Neural Networks*, vol. 18, no. 4, pp. 407-422, 2005.

[DEV 06] DEVILLERS L., VIDRASCU L., "Représentation et détection des émotions dans des données issues de dialogues enregistrés dans des centres d'appels: des émotions mixtes dans des données réelles", special issue, "Interaction émotionnelle" *Revue des Sciences et Technolgies de l'Information*, vol. 20, no. 4-5, pp. 447-476, 2006.

[DEV 07] DEVILLERS L., Les Émotions dans les Interactions: Perception, Détection et Generation, LIMSI-CNRS/University of Paris XI, Orsay, 2007.

[EHR 02] EHRETTE T., CHATEAU N., D'ALESSANDRO C., MAFFIOLO V., "Prosodic parameters of perceived emotions in vocal server voices", *Speech Prosody*, 11-13 April, Aix-en-Provence, France, 2002.

[EKM 92] EKMAN P., "An argument for basic emotions", *Cognition and Emotion*, vol. 6, pp. 169-200, 1992.

[ENO 06] ENOS F., BENUS S., CAUTIN R., GRACIARENA M., HIRSCHBERG J., SHRIBERG E., "Personality factors in human deception detection: Comparing human to machine performance", *Interspeech – International Conference on Spoken Language Processing ICSLP*, September 17-21, Pittsburgh, USA, 2006.

[ENO 07] ENOS F., SHRIBERG E., GRACIARENA M., HIRSCHBERG J., STOLCKE A., "Detecting deception using critical segments", *Interspeech*, Anvers, 2007.

[GRI 07] GRICHKOVTSOVA I., LACHERET A., MOREL M., "The role of intonation and voice quality in the affective speech perception", *Interspeech*, Anvers, 2007.

[HIR 05] HIRSCHBERG J., BENUS S., BRENIER S., ENOS F., GILMAN S., GIRAND C., GRACIARENA M., KATHOL L., PELLOM B., SHRIBERG E., STOLCKE A., "Distinguishing deceptive from non-deceptive speech", *Eurospeech*, September 4-8, Lisbon, Portugal, 2005.

[HUM 04] HUMAINE, "Preliminary plans for exemplars: signals and signs", *Workpackage 4 Deliverable*, HUMAINE 2004.

[IRI 07] IRIONDO I., PLANET S., SOCORO J., ALIAS F., MONZO C., MARTINEZ E., "Expressive speech corpus validation by mapping subjective perception to automatic classification based on prosody and voice quality", *International Conference of Phonetic Sciences ICPhS*, 6-10 August, Saarbrucken, Germany, 2007.

[LEE 01] LEE C., NARAYANAN S., PIERACCINI R., "Recognition of negative emotions from the speech signal", *IEEE Automatic Speech Recognition and Understanding Workshop ASRU*, 9-13 December, Trento, Italy, 2001.

[LIC 01] LI-CHIUNG Y., CAMPBELL N., "Linking form to meaning: The expression and recognition of emotions thourgh prosody", *Fourth ISCA ITRW on Speech Synthesis*, August 29–September 1, Perthshire, Scotland, 2001.

[LIP 03] LIPPMANN N., "Speech recognition by machines and humans: benchmarking human performance for for continuous speech recognition", *Speech Communication*, vol. 22, no. 99, pp. 1-15, 2003.

[MAD 02] MADDIESON I., VASILESCU I., "Factors in human language identification", *International Conference on Spoken Language Processing ICSLP,* September 16-20, Denver, Colarado, 2002.

[MAF 07] MAFFIOLO V., CHATEAU N., CHENADEC G. L., "Temporal organization in listeners' perception of the speakers' emotions and characteristics: Away to improve the automatic recognition of emotion-related states in human voice", *ACII*, 12-14 September, Lisbon, Portugal, 2007.

[MOR 06] MORRISON D., SILVA L. D., "Voting ensembles for spoken affect classification", *Journal of Network and Computer Applications*, vol. 30, pp. 1356-1365, 2006.

[MOZ 01] MOZZICONACCI S., "Modeling emotion and attitude in speech by means of perceptually based parameter values", *User Modeling and User-Adapted Interaction*, vol. 11, pp. 297-326, 2001.

[MOZ 02] MOZZICONACCI S., "Prosody and emotions", *Speech Prosody*, Aix en-Provence, 2002.

[MUT 94] MUTHUSAMY Y., JAIN N., COLE R., "Perceptual benchmarks for automatic language identification", *IEEE International Conference on Acoustics, Speech, and Signal Processing ICASSP*, April 19-22, Adelaide, Australia, 1994.

[ORT 03] ORTONY A., TURNER T., "What's basic about basic emotion?", *Psychological Review*, vol. 97, pp. 315-331, 2003.

[OSG 75] OSGOOD C., MAY W., MIRON M., *Cross-cultural Unniversal of Affective Meaning*, University of Illinois Press, Urbana, 1975.

[PAL 03] PALLETT D. S., "A look at the NIST's benchmark tests: past, present, and feature", *Automatic Speech Recognition and Understanding ASRU IEEE Workshop*, Nov 30-Dec 4, St. Thomas, Virgin Islands, 2003.

[PET 99] PETRUSHIN V., "Emotion in speech: Recognition and application to call centers", *Artificial Neural Networks in Engineering*, St. Louis, Missouri, 1999.

[PIC 97] PICARD R., *Affective Computing*, MIT Press, Cambridge, 1997.

[POL 98] POLZIN T., WAIBEL A., "Detecting emotions in speech", *Cooperative Multimodal Communication, 2nd International Conference, CMC'98*, BUNT H., BEUN R.J. (eds), Tilburg, The Netherlands, 1998.

[SCH 00] SCHERER K., "A cross-cultural investigation of emotion inferences from voice and speech: Implications for speech technology", *International Conference on Spoken Language Processing ICSLP*, Beijing, China, 2000.

[SCH 03] SCHERER K., "Vocal communication of emotion: A review of the research paradigms", *Speech Communication*, vol. 40, pp. 227-256, 2003.

[SCH 06a] SCHRÖDER M., COWIE R., "Developing a consistent view on emotion-oriented computing", in: RENALS S. and BENGIO S. (Eds), *MLMI 2005*, LNCS 3869, Springer-Verlag, Berlin-Heidelberg, 2006.

[SCH 06b] SCHRÖDER M., HEYLEN D., POGGI I., "Perception of non-verbal listener feedback", *Speech Prosody*, Dresden, 2006.

[SCH 07] SCHARENBORG O., "Reaching over the gap: A review of efforts to link human and automatic speech recognition research", *Speech Communication – Special Issue on Bridging the Gap between Human and Automatic Speech Processing*, vol. 49, pp. 336-347, 2007.

[SOB 99] SOBIN C., ALPERT M., "Emotion in speech: The acoustic attributes of fear, anger, sadness, and joy", *Journal of the Psycholinguistic Research*, vol. 28, no. 4, pp. 347-365, 1999.

[VID 07] VIDRASCU L., *Analyse et Détection des Émotions Verbales dans les Interactions Arales*, LIMSI-CNRS/University of Paris XI, Orsay, 2007.

[YAN 08] YANUSHEVSKAYA I., GOBL C., CHASAIDE A. N., "Voice quality and loudness in affect perception", *Speech Prosody*, Campinas, Brazil, 2008.

PART 3

Functions

Chapter 8

The Role of Emotions in Human−Machine Interaction

8.1. Introduction

The rise of new technologies has placed computer systems at the heart of our daily lives. These new advances increasingly seek to take on roles normally carried out by humans, such as those of the tutor, advisor or animator. Indeed, research has shown that users tend to interact with these virtual agents in the same way as they would with humans [REE 96]. As such, one of the aims in this field of research is to provide machines with social intelligence to the extent that they can interact with humans naturally and easily.

The considerable amount of research into emotions and their link to the brain and social behavior has led to the creation of the term *emotional intelligence*, as proposed by Salovey and Mayer [SAL 00]. Emotional intelligence refers to the ability to perceive and express emotions, to understand and reason with emotions as well as regulating them within ourselves and others. This intelligence allows an individual to better integrate into society and to achieve his or her aims in his or her professional and personal life more quickly [GOL 97]. Researchers are currently interested in integrating some form of emotional intelligence into computer systems. As Minsky has highlighted [MIN 86], "the question is not whether intelligent machines can have any emotions, but whether machines can be intelligent without any emotions". It is as a result of this that artificial intelligence has developed as a

---

Chapter written by Valérie MAFFIOLO and Magalie OCHS.

Emotion-Oriented Systems Edited by Catherine Pelachaud

field of research, also known as *affective computing* [PIC 97].The aim is therefore to use emotions to improve computer performance, of which there are three main field of research:

− producing systems capable of recognizing users' emotions;

− modeling emotional processes in computer systems;

– designing systems that can adapt to users' emotions.

The first of these fields − involving the design of systems that can recognize user emotions − mainly examines how to use physiological, somatic (e.g. blushing), vocal, lexical, semantic and dialogical indicators to identify users' emotions. This can be achieved using external sensors that record visual and sound signals. Emotion-recognition systems also sometimes rely on modeling users' emotional processes, which then allows us to identify at which point an emotion occurs during an interaction. This constitutes the second of the research fields that studies how to model emotional sub-processes in virtual agents[1] (such as triggering an emotion, its expression and influence on cognition) to allow us to identify when and how the agent will express an emotion. The third area of research attempts to create systems that can account for the users' emotions by adapting their mode of interaction, dialog, etc.

Researchers in affective computing pursue different objectives when integrating emotions into computer systems:

− *Improving adaptability, autonomy and, more generally, virtual agent performance in problem solving*: research in human and social sciences has shown that emotions are useful, if not indispensable, mechanisms for the correct functioning of an individual's cognitive processes [SCH 00]. For example, they allow us to remember important information or to select an action adapted to the situation at hand. In computational models, emotions are therefore used as a heuristic for selecting an action [CAN 03], a memory, researching information [VEL 98] or choosing a reasoning strategy [OLI 03]. These models are specifically designed for robots or, more generally, agents working in uncertain environments that may have several different aims simultaneously and limited resources.

− *Improving virtual agents' believability*: an agent is *believable*[2] when it gives the appearance of being alive [BAT 94], i.e. when the user feels that he or she is interacting with a virtual agent, alive, which has its own emotions, beliefs and

1 *Virtual agent* here is taken to mean a computational being with a certain degree of autonomy. It is not necessarily represented by a virtual character.

2 This refers to the believability of the agent as perceived by the user on the basis of the virtual agent's appearance and animation and not on the grounds of the agent's responses.

wishes [RIZ 00]. A virtual agent with emotional behavior[3] can therefore increase its believability. To create emotionally believable virtual agents, models are designed to identify at which moment during interaction an emotion should be expressed and the (verbal or non-verbal) behavior that follows this [DER 03, ELL 92, GRA 04]. Furthermore, research has shown that emotionally believable agents can increase user satisfaction [BAR 02, BRA 05].

–*Improving user engagement and performance:* the user can feel a number of emotions during his/her interaction with a computer [PIC 97]. He/she might feel happiness when having successfully completed a task or frustration when the system fails. Studies have shown [BOW 92, ISE 00] that an individual's emotions strongly influence their cognitive abilities (memory, concentration, reasoning, etc.). Some researchers [BOT 06, OCH 08] have focused on creating systems that can manage users' emotions to improve their engagement and performance when carrying out tasks. This requires systems to be designed that can identify the user's emotions and his/her type of behavior to optimize the interaction.

The aims pursued by researchers in affective computing are guided by the context of the desired application. The role played by emotions (the virtual emotions of the agent's or the real ones of the user) in an interaction depends mainly on their context in which they appear. In this chapter, we will explore the different roles that emotions can play, both on a human and machine level in different contexts of application. More specifically, we will examine the fields of interactive systems for information and assistance (section 8.2.), video games (section 8.3) and intelligent tutorial systems (section 8.4). For each of these domains, we will analyze the benefits of using emotions in this context as well as research carried out and existing tools. To conclude the chapter, will we analyze the various research perspectives in these different domains.

8.2. Interactive information and assistance systems

Interactive systems are those that involve user interaction to organize the consultation of information. Numerous services have been developed based on interactive systems, in particular in telecommunications. This includes interactive vocal services (such as automatic information services) and virtual information services over the Internet that integrate virtual assistants.

---

3 Emotional behavior includes non-verbal expressions (facial, gestural expressions or specific actions) and verbal expressions (what and how something is said).

8.2.1. *Uses of emotions in interactive systems*

Emotions play a highly important role in interactive systems. With the question of ergonomics and service usability comes the issue of emotional processing. As *appraisal theories* imply [SCH 00], in human–machine interaction, each main event will be subjectively evaluated by the user, depending on whether the event represents progress or an obstacle towards his or her aims. The user's emotional reaction to obstacles will involve an adaption of his/her cognitive and motor behavior to overcome or avoid the problem. At the first obstacle, the emotional reaction will probably be moderate and the adaptation will be minimal (e.g. in the case of a vocal interaction, repeating the same thing but moving closer to the microphone). If the obstacles increase in number, the emotional reaction will increase in intensity (which, depending on the individual, could potentially lead to anger or desperation) and could lead to increasingly significant adaptations, or even a break in strategy (e.g. hanging up violently). Negative emotions experienced by the user can reduce their satisfaction, affect their perception of the interface, and prevent them from concentrating on and remembering information [BOW 92, ISE 00].

A system that can identify user emotions and select appropriate responses (including responses to influence the user's emotions) or even generate emotional expressions themselves therefore has real significance and is what we call an *affective loop* [HOO 02]. Proposing an interaction based on such a system will allow us to improve interaction and, more generally, to increase satisfaction, performance and user engagement. It will also improve user perception of the service and task [BRA 05, KLE 99, OCH 08, PRE 05]. As such, since the user's expressions of emotions convey information about his/her mental state [KEL 01], the ability of a system to recognize the user's emotions can enable the system to better understand the user [PIC 97]. This will allow the system to adapt to the user's behavior, such as preventing a disrupted interaction situation, for example.

8.2.2. *Current research and tools*

Current interactive systems often have difficulty conducting interactions with humans. Due to the quality of the media or the variability of human behavior, "intelligent" systems can sometimes fail to understand user demands. Since the *Computers Are Social Actors* (CASA) paradigm [REE 96], it has been known that human–machine interactions present certain similarities with human interactions. One way of improving the quality of human–machine interaction is to develop computer systems that can perceive and understand human behavior and respond to it appropriately [PAN 07]. Human behavior can range from a simple linguistic message to a non-verbal act of communication of a cognitive state and/or emotion.

In academic literature on the subject, there have been a number of studies focusing on the primary emotions or the *Big Six* (happiness, anger, sadness, disgust, fear and surprise) [EKM 76]. More recently, so-called complex emotions that more precisely describe the range of daily emotions have been studied. This includes emotion-related states [COW 03], mixed emotions [VID 05] and mental states [BAR 04].

In order to improve human–machine interactions, it is evident that computer systems need not only to account for what individuals feel in terms of the word "emotion" but also whether they are confused, skeptical or sure of themselves. These states can be expressed in different ways, including facial expressions, head movements, speech (including non-linguistic vocal characteristics), body gestures [KNA 06] and specific physiological reactions [PIC 01]. (For a more detailed description of emotion recognition techniques, see Chapters 4, 5 and 7.)

Carrying out studies into multimodal expressions of human behavior is the best way of clarifying information and, as such, getting a better indication of the emotion and cognitive states [LEC 06]. Technical advances have brought an increasing amount of functionality, allowing interaction (audio, video, haptic, tactile, etc.) with the machine to develop beyond a simple mouse and keyboard. Equally, a machine can use these modalities to improve communication with the user.

In order to recognize user emotions and determine how a machine can communicate emotions perceptively, it is firstly necessary to know how emotions are expressed and perceived. A large amount of research has focused on collecting and annotating corpora of human–machine interaction or human expressions of emotion. We will examine some of this research in this section (8.2.2), but for a more comprehensive overview, the reader should consult Chapter 3.

8.2.2.1. *Collecting and annotating expressions of emotions*

Currently there are three ways of collecting emotional expressions [DOU 07]. The first method consists of collecting *acted expressions* [BAN 06], which involves asking actors to act emotions. However, in an automatic recognition problem, a system built on a corpus of acted expressions has certain limitations [BAT 03, ZEN 07] due to the fact that acted and spontaneous emotions can be expressed differently. Collecting *natural data* is therefore the second method. In this approach, expressions of emotions appearing in daily life are recorded. For example, the EmoTV1 database [ABR 05] contains recordings of televised interviews. The construction of such corpora can prove to be difficult, particularly in the case of human–machine interaction. The third method involves *recording data induced in laboratory conditions*, which are non-acted expressions of emotions. The main advantage of this approach lies in the quality of the signals recorded and the control of stimuli [AUB 03, LEC 06]. This method often relies on the Wizard of Oz paradigm, which allows us to simulate human–machine interaction while researchers control the machine's actions. This method therefore allows a greater amount of flexibility than the second method in terms of technical conditions and managing the interaction. An example of this is the research carried out by Zara *et al.* [ZAR 09] who used audio-video interactions between pairs of subjects (an interaction constrained by a board game and its rules), to build a corpus of spontaneous expressions.

Regarding annotating emotional corpora, this can be split between *encoding emotion-related content* and *encoding emotional signals*. Encoding emotion-related content consists of inferring the emotion expressed based on the observations by annotators of multimodal behaviors [LEC 07]. Annotation techniques often reveal either a categorically discrete approach or a continuous dimensional approach. Encoding emotional signals consists of annotating the behavior observed in each of these modalities in detail. This provides highly relevant data to compensate for the fact that automatic extraction of parameter values for each useful modality is not yet optimal. For facial signals, Ekman and Friesen [EKM 78] have developed a standardized annotation method called a *Facial Action Coding System* (FACS). Vocal expressions are the most commonly studied feature [MAF 06] but there is no standard annotation. Some researchers have proposed a standard annotation of multimodal expressions [ALL 06], but this has not yet been adopted across the board. An increasing number of researchers are using a tool called ANVIL (http://www.anvil-software.de/), which enables the annotation of multimodal data [KIP 01]. Note that the context of the interaction with the system is highly important for understanding what and why a piece of information has been communicated. The task of the user and his/her dialog with other people is relevant information for a computer system designed to be sensitive to its user's emotions. Some studies have examined multimodal behavior, examining context specifically with regards to annotating and analyzing the corpus [ALL 06, ZAR 07].

8.2.2.2. *Automatic emotion recognition*

Work on constructing interactive information and assistance systems that can automatically recognize user emotions have mainly focused on the voice. Traditionally, such systems have been built using corpora of acted expressions [DEL 96, POL 99]. More recently, studies have tended to use real data, motivated by the need to improve human−machine interactions in real-life everyday situations, such as with robots, for example [OUD 02], or voice services [ANG 02, BAT 03, CHA 04, LEE 05]. The majority of these systems rely on constructing a learning and classification model that consists of matching the values of the parameters calculated from vocal signals and the emotional labels attached to them (encoding the content relating to emotion). However, the studies disagree on two issues, the first being which parameters to extract − whether linguistic (e.g. lexical or syntactic) or paralinguistic (often relating to acoustics [CHA 05, COW 01]); the second being how to combine them.

The vast majority of learning model-based systems only rely on acoustic parameters [COW 01, DEL 96, OUD 02, VER 06], although this does present some disadvantages. For example, a purely acoustic system cannot distinguish between linguistic (relating to words and sentence structure) and paralinguistic (relating to the speaker's state, regional accent, etc.) intonation. It is also unable to detect an emotion unless the overall intonation leaves the linguistic intonation framework. In order to be able to account for both linguistic and paralinguistic information, some systems rely on a combination of these two types of parameters [ANG 02, BAT 03, BAT 06, LEE 05, POL 99, SCH 04b]. This combination relies on either fusing the two parameters or fusing the decisions from a model constructed from acoustic parameters and another constructed on linguistic parameters. This approach does not allow us to specifically account for interaction between acoustic and linguistic parameters. The parameter fusion approach therefore seems to be the best option. A common problem for all learning and classification methods also lies in the choice and length of the window on which the parameters are calculated [LEE 04, MAF 07].

Automatic emotion recognition remains problematic, given that the emotions expressed by users when interacting with an information and assistance system are generally of weak intensity (except in some contexts, such as medical situations). Furthermore, for more efficient recognition it seems to be important to integrate contextual information into the information system and, in particular, the cause of the emotion, e.g. the fact that the triggered emotion is linked to an event that is external or linked to the interaction.

8.2.2.3. *Interface expressivity*

Embodied conversational agents or ECAs are virtual characters that can communicate with users, ranging from delivering information to teaching [CAS 00]. A significant amount of research has focused on improving ECA autonomy [RUT 02] while other studies have sought to integrate emotions for improved user−ECA interaction [CAS 00, NIJ 01, RUT 02]. This approach has tended to improve several aspects of the interaction, addressing the interface design, user perception [NIJ 01], engagement [DAR 02], performance in task achievement and, more generally, the usability of a service based on the interactive system.

8.2.2.3.1. Expressing emotions in ECAs

Emotions can be expressed via different modalities. For emotions expressed via the face, those cited by Ekman and Friesen [EKM 75] are the most commonly used expressions for ECAs today, while recently the MPEG-4 coding scheme has also been proposed [PAN 03]. The animation of the ECAs face, *Greta* [PEL 02], for example, is based on this approach [DER 03].

Several computational models have been developed to allow ECAs to express emotions via the face. The complexity of the latter depends on the emotional phenomena considered (e.g. the combination of emotions, the expression of faked emotions, etc.). DuyBui [DUY 04], for example, has proposed a series of fuzzy logic rules that allow us to link an emotion vector (representing the intensity of the six basic emotions) to a facial muscle contraction vector (representing the level of contraction of 19 muscles of the face). This model is particularly convenient to the expression of combined emotions. In *The Affective Reasoner* [ELL 92], the virtual character has 72 facial expressions to express the 24 emotions modeled with three different possible intensities with an algorithm that allows us to identify how to pass from one expression to another. When one person speaks, a facial expression cannot simply be presented, as Bevacqua *et al.* [BEV 04] have highlighted. In their research, they developed a model that could account for lip movements when the virtual agent spoke with a facial expression showing an emotion (for further details on modeling facial expressions, see Chapter 6).

Emotions can also be expressed by gestures. To model gestures with an emotionally expressive virtual agent requires the analysis of individuals' gestural behavior during interpersonal interactions. Corpora are annotated to compare an individual's emotions with their gestures (movements of the head, torso and hand gestures) in, for example, audiovisual interviews [MAR 06] or during human game interactions [ZAR 07]. As such, the EMOTE system has been developed to model the influence of emotion on a virtual agent's gestures [BAD 02]. This allows us to modify virtual agents' gestural expressivity. A set of parameters are used to control the virtual agents' body movements (e.g. the trajectory and speed of movements). These parameters are then modified depending on the ECAs emotion and personality. The EMOTE system acts as a filter on the movements of the pre-recorded ECA bodily movements. The computational model proposed by Hartmann *et al.* [HAR 05] directly alters the gestural behavior of the virtual personality. The character's expressivity is described by six attributes (the quantity of movement, amplitude, duration, fluidity, strength and repetitions). Similarly, Egges *et al*. [EGG 05] have focused on modeling the virtual agent's bodily movement depending on its emotions when not carrying out any particular action. This research can also be applied to the problem of a virtual agent expressing emotion while waiting for the user to perform an action, e.g. awaiting a request. On the basis of a 2D representation of emotions (valence and activation), Egges *et al*. combined the characteristics of waiting movements to emotional dimensions. For example, pauses between changes of position are as short as the arousal dimension is large.

ECAs can also be provided with speech, most commonly synthetic. The current challenge in vocal synthesis is to propose a synthesized vocal that is emotional, or at least expressive. Several emotional vocal synthesis systems have been developed. The most commonly used method is *speech synthesis by rules*. This consists of creating the speech directly using acoustic rules without using pre-recorded speech, resulting in an unnatural robotic voice. This method, however, offers a large degree of flexibility and control of acoustic parameters. It is therefore particularly applicable for acoustic emotion modeling. The first emotional speech synthesis system, called *Affect Editor*, is based on this approach [CAH 90].

More recently, Schroeder [SCH 04] has defined some emotional prosody rules. This represents emotion in a 3D space (valence, activation and power). Using research based on academic literature and analysis of emotional databases, Schroeder defines the linear coefficient highlighting the effect of an emotional dimension on acoustic parameters.

Ball [BAL 02] has proposed using a Bayesian network to represent the effect of emotions on the voice. As such, the level of activation and valence of the emotion are linked to tonality and speed and volume of the voice.

In *The Affective Reasoner* [ELL 92], the tone, volume and speed of the virtual agent's speech vary depending on the emotion. More precisely, for each of the 24 emotions, three literal values are allocated to these variables. The development of an interactive system integrating synthetic emotional speech brings significant improvements on three levels: the users are more stimulated and engaged in the interaction [DAR 02]: the user can better understand and memorize information because the attention given to speech is led by prosodic patterns that are precisely controlled in emotional synthetic speech [JON 93]: ECAs built with synthetic emotional speech are considered to be more believable [RUT 04].

Emotions can also be expressed via linguistic sentences [KER 00]. Researchers such as [STR 04] have specifically focused on elaborating emotional lexicon. This consists of identifying, structuring and organizing words that have emotional connotations. An example of a dictionary of affective terms is WordNet-Affect [STR 04]. In this, terms are grouped according to categories (happiness, sadness, anger, surprise, sadness and disgust) and according to type (verb, adjective, adverb and nouns). This dictionary has been used, for example, for automatic recognition of emotions expressed by the user in utterances [MA 05]. Emotions can also be expressed using *diminutive affective suffixes* (e.g. "framemakers", insults, swearing, affectionate terms such as "dear", exclamations such as "ah!" or interjections such as "hmm"). In terms of Speech Act theory, a specific category of communicative acts, called *expressive acts*, expresses the emotional mental state of the speaker. For example, by saying "I am happy", the speaker is expressing his or her mental state of happiness. As Kerbrat-Orecchioni has highlighted [KER 00], "all words and constructions produced in a favorable context can convey affective connotations". In computational models of virtual agents, pre-recorded emotional utterances (e.g. "hooray", "congratulations" or "I am happy") are generally used to express the agent's emotions [BAL 02, ELL 92, LES 00].

There have been several ECAs expressing emotions via a variety of modalities. For example, at the entrance of the German Research Center for Artificial Intelligence (DFKI), there is an information point with a virtual receptionist [AND 01]. He or she can respond to visitors' questions or initiate conversations on various subjects (news, research projects and DFKI staff). The emotions of the virtual agent are expressed through different modalities, such as gesture or voice intonation. Another example is the ECA Greta [PEL 02], who communicates using not only speech but non-verbal behavior. Capable of expressing emotions, Greta can easily be manipulated using the XML files in which the behavior of the agent can be described.

8.2.2.3.2. The Impact of emotional virtual agents on interaction

Some researchers [BAR 02, BRA 05, KLE 99, PAR 04, PRE 05], have focused on the impact of emotion-expressing virtual agents on human interaction. Research by Bartneck [BAR 02], for example, has shown that a user experiences greater *enjoyment* when interacting with an emotionally expressive virtual character than with a character expressing no emotion. The work by Brave *et al*. [BRA 05] has similarly shown that the user perceives a virtual character as being more agreeable when expressing emotion as opposed to when it expresses none. Research has focused specifically on the impact of a virtual agent expressing emotions that it thinks the user is feeling (also called empathy emotions) on interaction. For example, in research carried out by Prendinger [PRE 05], frustration is deliberately induced in the user by simulating system failures. The results of experiments show that the user has less stress and frustration when faced with a character expressing the emotions that it believes the user to be feeling (empathy emotions) than with an agent expressing no emotion.

Using the same method of simulating system failures to generate frustration, Klein [KLE 99] has shown that expressions of empathic emotions by the agent (via text message) reduce the *level of frustration* felt by the user. Recent research [PIC 07] has shown that a system that interrupts a user is preferred when the system's behavior is marked by empathy rather than when no empathy is expressed. The system appears to be less perturbing and frustrating. In the works carried out by Ochs *et al*. [OCH 08], it has been shown that a virtual character used as an information system is perceived significantly more positively when expressing empathy. Similarly, according to results found by Brave *et al*. [BRA 05], the user feels more "supported" (i.e. more assisted and less alone) when faced with an agent expressing empathy. In other research [KLE 99], results have shown that in the same conditions of frustration, users' playing time with virtual agents expressing empathy is higher than those expressing no emotion. Similarly, in research carried out by Partala and Surakka [PAR 04], the results show that user performance when carrying out a task is better when the virtual agent expresses empathy. In comparison with agents expressing no emotion or non-emphatic emotions, emotionally emphatic virtual agents therefore seem to favor a more significant improvement in interaction.

This research remains limited however. Furthermore, the context of application of these evaluations is mainly that of games where the virtual character plays the role of an assistant [BAR 02, BRA 05, PAR 04]. The results also seem to depend on the culture of the participants themselves [BEC 05, BRA 05] and sometimes appear to be contradictory [BAR 02, PAR 04]. As Becker *et al*. have demonstrated [BEC 05], expressions of emotions by a virtual character can distrub the interaction when they are inappropriate to the situation.

With regard to academic literature in this area, more evaluations are necessary, in particular with regard to each socio-cultural context and for each specific application. This will allow us to evaluate the real impact of emotionally expressive agents on interactions and, more generally, create an interactive system with the ability to recognize, manage and express emotions.

8.3. Video games

Nietzsche famously wrote that: "In every real man a child is hidden that wants to play". The video games industry has seen a rapid global expansion in recent years. The widening of its target audience (adults, women, families, etc.), improved console performance, the growth in importance of *serious gaming*[4] and the development of new accessories has created a context of application in which emotions are important.

4 Serious gaming is a combination of a game with a serious or practical dimension conveying a particular message (educational, informative, advertising, etc.), for training, learning or exchanging data/goods.

8.3.1. *The importance of emotions in video games*

8.3.1.1. *Emotions in gaming systems*

One of the main challenges for the gaming industry is to develop non-player characters (NPCs) whose behavior (actions, expressions, utterances, etc.) are believable. As the animators at Disney have highlighted, the expression of an emotion creates the illusion of life for an artificial being: “From the earliest days, it has been the portrayal of emotions that has given the Disney characters the illusion of life” [THO 81]. Furthermore, an expression of an emotion also conveys information. For example, it reflects a particualr individual’s mental state [KEL 01] in relation to a given situation. Fear, for instance, characterizes the presence of danger. In interpersonal relations, individuals generally use others’ expressions of emotions to help them evaluate the emotional charge of a situation and as such determine their actions [MAN 01]. Similarly, in video games, expressions of emotion by NPCs can be used to communicate information to the user about the situation, for example to warn them about imminent danger.

Research has shown that emotion is generally accompanied by *action tendencies* (e.g. fleeing following fear) that allow an individual to face an event [FRI 86]. In video games, emotions can be used as a heuristic to determine appropriate and believable actions for NPCs in a given situation. The emotions of an NPC can be calculated dynamically during the game according to the choices made by the player character. Using these emotions to determine adaptive behavior therefore allows us to improve the autonomy of NPCs whose behavior is currently almost entirely scripted.

As Stern has highlighted [STE 03], emotions in NPCs are essential for creating an *emotional relationship* between the gamer and virtual characters. A human–machine emotional relationship is defined as a long-term interaction in which the player and virtual character pay attention to each other’s emotions, communicate them, express empathy and, as such, establish a connection. The success of virtual pets such as Dogz and Catz [PET 95], Tamagotchis [BAN 96] and the Aibo robot [SON 99] are good examples of this. Similarly, in the role-playing game Fable II (Lionhead Studios), an emotional relationship is established between the player and a virtual dog. The dog’s behavior is based on artificial behavior and responds to three basic rules, which are to avoid annoying the player, to love him/her unconditionally, and to take care of him/her. The dog’s behavior also depends on the context of the game. For example, if the player enters a shop, the dog will wait at the door. The dog also helps the player to accomplish tasks (such as fighting enemies, for instance). The player does not have direct control over the dog but can interact with it (e.g. praise or punish it for its actions), leading it to learn throughout the game how to gain the attention of the player when there is a threat or to help him/her discover hidden treasure.

Finally, emotions in video games can be used to create the illusion that characters are alive, as a conveyer of information in gaming situations, to improve the autonomy of NPCs and to potentially encourage an emotional relationship between the user and virtual character.

8.3.1.2. *Player emotions*

Video games are also a source of emotions for users. One of the main aims for game designers is to maximize the pleasure, happiness, enjoyment and well-being of the player. In the 1970s, Mihaly Csikszentmihalyi introduced the concept of *flow* to characterize the concept of complete concentration on an activity with a high level of pleasure and satisfaction, which can be seen as specific emotional states [CHE 07]. During a *flow* experience, the player, completely immersed in the game, loses all notion of time. The *flow* experience requires a balance between the challenge of the game and the ability of the player (see Figure 8.1). A challenge that is too difficult for the player can cause anxiety or boredom. The aim is therefore to keep the player in the *flow* zone for as long as possible.

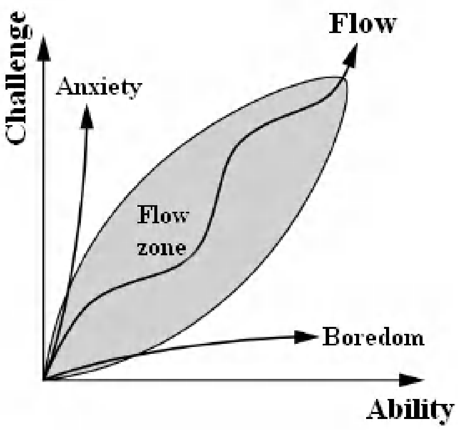


Figure 8.1. *The flow zone*

Some games, such as the ones we examine in the following sections, use the user's emotions for teaching or therapeutic purposes. For example, the Ico[5] game is used by psychologists for therapeutic requirements to observe children's reactions

5 Sony, 2001.

when faced with tests. In particular, the child can be disoriented when the game asks the player to choose between letting go of a princess' hand or taking care of her. Similarly, the *Virtual Iraq*[6] game was recently developed to help soldiers returning from Iraq to relive traumatic events so that they could deal better with these experiences.

In some games, the user's emotions are used as a variable of the game to improve their engagement. More precisely, the game is designed so that the user's emotions affects how the game progresses. A number of interactive systems use individuals' facial and bodily expressions as an entry point. For example, boxing games and, more generally, games developed for the Wii and Wii Fit use bodily movements (http://us.wii.com/). Observations of the face and body can be used differently according to the user, so that they control the game and can consciously use their movements and expressions to control the interface. Inversely, the application can adapt itself to the user's emotional state (this state can be interpreted from the user's behavior). We can therefore distinguish between *voluntary* and *involuntary* control. With *voluntary* control, the user consciously produces facial expressions, movements of the head and gestures to control the game. This includes commands that allow the user to navigate in a game environment or animate virtual characters (e.g. the virtual character showing a similar facial expression to that of the player).

As expressions and movements are made consciously, they do not necessarily reflect the user's real emotional state. For example, PKR.com is an online poker site where the player can display their emotions (i.e. neutral, angry, happy, relaxed or sad), therefore controlling the temperament and tone of the player character representing the user. Over the course of the game, the player can indicate their emotion via facial expressions to trick the other players. This is achieved by activating emotion indicating commands by typing / in the chat box, followed by the indicator name of the emotion so that when, for example, the player types "*/SMILE*", their virtual character smiles.

During the first E3 gaming conference [FRA 09], Microsoft unveiled project Natal, which aimed to combine gaming systems with the body (players' movement using an RGB camera and a 3D depth sensor) and speech recognition (using a microphone), removing the need for joysticks.

With *involuntary* control, the system detects the player's spontaneous facial or bodily expression and infers an emotion from it which it uses to adapt the game. This adaptation can affect the appearance of the game environment, the mode of interaction, experience and engagement, and the feedback and strategy followed by

---

6 University of Southern California, 2007.

the game or actors in the game. For example, *Mindball*, (http://www.mindball.se/) is a game mixing *voluntary* and *involuntary* control. In this ping-pong style game, the two players control a ball using brain waves. The most relaxed player is the one who wins the game. The players wear headbands and sit opposite each other at a *Mindball* table. The ball is placed along a table so that it rolls towards one of the players. The electric activity manifested by the brain, also called EEG (electroencephalography), is detected by the sensors (electrodes) attached to the headband, which is connected to a bio-detector system. In this game, the *voluntary* control comes from the player being in a state of intense relaxation. The *involuntary* control is characterized by a lack of control over brain waves, which can vary during the game. An example of a system using involuntary control unveiled during the E3 gaming conference was the *Wii Vitality Sensor*, currently under development by Nintendo, which is a sensor placed on the ends of the thumbs that transmits information (e.g. pulse, tenseness, etc.) to the machine [FRA 09]. This is an accessory that could be used in future Nintendo sports *coaching* software or, more generally, to adapt the game according to the user's level of interest.

8.3.2. *Current research and tools*

8.3.2.1. *Towards believable emotional virtual characters*

In order to be able to develop *emotionally believable* virtual characters, we need to be able to identify which emotion the character should express, when and how Several computational models have been proposed for this purpose, which we examine in the following section.

The *Em* architecture [REI 96], for example, aims to provide a series of tools allowing the designers of virtual characters to provide them with emotions (including a reasoning capacity to identify which emotion to express, when and how). Based on the OCC model [ORT 88], an emotion generator containing a set of emotion trigger rules is proposed. There are also several functions to calculate the intensity of the emotion, temporal decay and the way they are combined. Emotions cannot influence the character's actions directly but can effect specific characteristics of a character's behavior (such as aggressive or energetic behavior) *Em* architecture offers large a degree of flexibility due to its simple initial structure, allowing the construction of an emotional virtual character. Several virtual characters have been given emotional capabilities using *Em* architecture tools. A good example of this is *Lyotard* virtual chat (for more details, see [REI 96]).

The *Affective Reasoner* is an emotion simulation platform for virtual characters [ELL 92]. Each character is given specific goals, principles and preferences that are defined by the designer. These elements are used by the agent to appraise the situations it encounters in the virtual world. As such, the same situation can trigger different emotions depending on the character. Twenty-four emotions are represented in the *Affective Reasoner*. For each of these emotions, trigger conditions are designed independently of the application. Each situation is represented to enable easy identification of the emotions triggered. The intensity of the emotions is calculated according to the values of the intensity variables. These are determined in relation to specific characteristics of the situation and the character (its role, mood, personality and relationships with other characters). In response to the triggered emotions, depending on their personality the characters select one or several actions from a database. These actions can vary in type (somatic, verbal, non-verbal, etc.).

The *EMotion and Adaptation* (EMA) model proposed by Gratch and Marsella [GRA 04] involves *appraisal* and *coping* sub-processes. The *appraisal* sub-process allows a virtual character to determine its own emotions following an event. The originality of this *EMA* lies in the modeling of the *coping* sub-process. This allows virtual characters to identify the behavior to adopt when faced with emotions of strong intensity. There are also different *coping strategies*. For example, faced with a negative emotion, the virtual character can adopt a *positive reinterpretation strategy* where it focuses on the positive consequences of the event triggering the negative emotion. It can also adopt an *acceptance strategy*, where it recognizes the negative event as unavoidable. In this case, the character chooses to abandon the intention which, by its failure, has triggered the negative emotion. Several coping strategies can be used simultaneously. These strategies can modify the situation. This is then re-evaluated and new emotions are triggered *a priori*. This *EMA* model has been used to create virtual characters that can populate a military virtual environment [GRA 04].

The SCREAM (*SCRipting Emotion-based Agent Minds*) tool enables us to determine a virtual character's emotional reaction based on the character's personal goals, standards and attitudes. The particularity of this model is the distinction between the felt emotions and those that are expressed. A *social filter* module determines which emotion to express according to the character's role (represented by power and social distance) and based on its personality (represented by the dimensions of extraversion and agreeability). As such, if the interlocutor has more power than the character, negative emotions will be replaced by a neutral expression. Inversely, it is the personality that will determine which emotion to express. For example, a disagreeable and extravert character tends to express more negative emotions than positive ones.

In the Fantasy *A* [PAI 01] video game two characters fight in a duel by casting spells. The characters' emotions are used to determine their actions. The player takes on the role of an apprentice magician. To control the characters in the game, the player uses the *SenToy* tangible user interface. This interface is represented by a doll. The player indicates his or her emotions (anger, fear, surprise, victory, sadness and happiness) to the other characters in the game by simulating gestures that are characteristic of these emotions using the doll. For example, by placing the arms of the *SenToy* in front of the doll's face, fear is generated in the character representing the player. The characters act according to their emotions and those of their enemy. Based on work by Lazarus [LAZ 91], Ekman [EKM 93] and Darwin [DAR 72], a set of rules applies an action tendency to each emotion. For example, anger can lead to attack and fear can lead to avoidance or fleeing. The choice of action is also guided by the emotion's adversary. For example, if the adversary is happy, then the character will assume that he/she is about to attack (because he/she feels at ease in the duel). In some cases, the character's personality will determine their actions. Analysis of the game has shown that players have found the interaction enjoyable. With regard to tools for emotional expression in virtual characters, refer to section 2.2 and Chapters 3 and 6.

8.3.2.2. *Towards games that account for player emotions*

To be able to account for player emotions requires us to be able to identify the emotions expressed through the game and determine how to adapt to them. Using external sensors (cameras, microphones, physiological sensors), current tools can recognize player emotions using facial expressions, voice or even physical reactions [PIC 97]. A example of the use of voice analysis in games is *Voxler*, a middleware development team focused on vocal interaction with the aim of developing a vocal interface solution that has, among other things, the ability to detect emotions in the voice in real time (in the context of the collaborative projects *Affective Avatars* and *Romeo*). Currently, the system detects activation (weak/strong) as a marker of expressivity.

Glasgow Caledonian University has opened a laboratory specifically for studying the emotive interactions of players in video games. Named, the *eMotion Laboratory*, it contains all the requirements for playing games but is also equipped with tinted windows, infrared cameras that trace eye movements, and pupil dilation and pressure and perspiration sensors on the gamepads, etc., i.e. severaltools needed to measure the excitement and emotions felt by the player [SYK 03]. Based on the latest developments in neurotechnology, *Emotiv Systems* (www.emotiv.com/) has developed a wireless headset that can analyze electrical activity in the human brain and transmit this information in real time to a computer, which allows the player to act on the interface and specifically to project their own facial expressions onto their avatar. To date, no evaluation results of this system have been published.

The MIT Media Lab is also leading a project called *AffQuake* [PIC], which aims to account for physiological signals indicating a player's emotions in the video game

Quake II[7]. Several modifications have been made so that the player character sees its behavior modified according to the physiological signals. As such, when a player is scared, his or her character also displays this emotion and jumps backwards. Furthermore, depending on the player's excitement, measured by his or her electrodermal response (the level of skin conductivity), the size of his or her character changes. In terms of game play, a taller player character signifies that the player can see further, although this also makes them an easier target.

An example of a game managing the player *flow* is *Buzzword* (a Tetris-style game). In this game, each block has a word attached. The aim of the game is to place the falling blocks on different piles in order to create relevant associations between words. To create a state of *flow*, a module that adapts the difficulty of the game (DDA – *Dynamic Difficulty Adjustment* module) is incorporated into the system. This system is used to place the player in a position of confronting often more challenging word associations, but which are sufficiently adapted so that the player wants to continue and does not stop playing. Since the system does not know the player's ability, the difficulty of the game is increased or reduced based on the player's successes and failures.

Other research projects seeking to integrate an affective dimension into games, including games studies and research laboratories, are currently being undertaken. An example of this is the DEEP project (Dialogs taking into account Experience, Emotions and Personality, adapted to computer games), whose objectives include using emotional models to increase the believability of non-player characters using less predictable behaviors guided by their emotions [OCH 09].

Multidisciplinary research into emotions has tended to highlight the importance of accounting for emotions in virtual player-character interactions to enrich video games. The dimensions of the games examined so far focus mainly on:

– the realism of the virtual environment, which affects the player's feeling of immersion;

– accounting for signals (physiological, visual, etc.) from the player, which increases the game's adaptability and playability; and

– the desire to keep the player in a state of *flow*.

While several models and tools have been proposed, however, the problem of evaluation and its real impact on the game remains an issue.

---

7 Developed by IDSoftware in 1997.

8.4. Intelligent tutoring systems (ITS)

8.4.1. *The importance of emotions in ITS*

8.4.1.1. *The learner's emotions*

Some interactive systems [ALE 06, ELL 97, MEL 08] have been developed for teaching purposes. Called Intelligent Tutorial Systems (ITS), they are used as a training tool. The user will feel a number of emotions during their interaction with this kind of system. For example, negative emotions can appear when the user's expectations are not satisfied while maintaining a concept can lead to positive emotions. The learner's emotions play a determining role for effective learning. Research has shown that in effect an individual's emotions will strongly influence their cognitive abilities [BOW 92, ISE 00]. Depending on the type and intensity of an emotion, reasoning and memory capacity can vary. As such, some tasks are more suited to negative emotions while others favor positive ones. Similarly, while some emotions reduce cognitive ability, others increase it [ISE 00]. ITSs that are capable of influencing learner emotions by adapting the interaction to the recognized emotions can therefore optimize the learner's abilities and improve learning. The development of such capacities requires the integration of a real-time detection module for the user's emotions and a module determining the best teaching strategy into the system. Note that in this research context, attention has focused more on studying emotions than cognitive states, such as confusion, curiosity or boredom. Research has shown that these cognitive states frequently appear during learning [LEH 08].

8.4.1.2. *Emotions in ITS*

In ITSs, animated conversational agents are often used to create virtual tutors or learning companions. The expression of emotion by these virtual characters can therefore be used to mark their agreement or disagreement, satisfaction or disappointment or even show empathy towards the learner. This indicates a certain level of interest in the user and his or her performance. Research has indeed shown that the expression of emotion by educational virtual characters improves learning [ELL 97, LES 00].

8.4.2. *Current research and tools*

Given the close relationship between emotion and learning, several works of research have focused on the development of ITSs including an emotional dimension [ALE 06, ELL 97, MEL 08]. For example, the virtual tutor *Eve* [ALE 06], which teaches math in individual classes to elementary school children, can detect the emotions fear, happiness, anger, disgust, sadness and surprise among the pupils in real time using their facial expressions. During teaching, *Eve* also expresses emotions using facial expressions. *Eve*'s reactions depend on the previous interactions with the student (responses to the student's questions and emotions and those of *Eve*). A system of case-based reasoning, developed by analyzing real interactions between tutors and pupils, is used to recommend *Eve*'s actions.

Autotutor [MEL 08] is an intelligent tutorial system, developed by the MIT Media Lab, which aims to help users learn physics and informatic with advice, complementary information and summaries as well as *feedback* on the learner's responses. This system manages emotions of confusion, engagement, boredom and frustration of the learner during the interaction. These emotions are detected using verbal and non-verbal indicators, extracts of dialog, bodily movement and facial expressions. These indicators have been selected using learning supervised algorithms used on data from 28 students interacting with *Autotutor*. Once the emotions have been detected, *Autotutor* can respond to them, specifically those of confusion, frustration and boredom. The strategy used to face boredom and frustration is the expression of empathy by the virtual tutor. Furthermore, the tutor defines sub-aims in cases where there is frustration to encourage the learner. For confusion the tutor encourages the learner, then expresses empathy by redirecting the learner if he/she is still confused.

The TLTS (*Tactical Language Training System*) [JOH 05] is designed specifically to train military staff to communicate with civilians in foreign countries. This includes teaching foreign languages and cultural information about that country. The learner is accompanied by a virtual tutor, which adapts its behavior to the learner's emotions in order to motivate them in the learning process.

To learn how to manage critical situations, which are often accompanied by strong emotions, the military training system, *Mission Rehearsal Exercise*, has been developed [HIL 03]. Virtual humans in this system adopt emotional behavior depending on how the situation develops. This behavior is based on the EMA emotional model [GRA 04].

The *FearNot!* system [PAI 04] has been created to increase sensitivity in children aged between eight and 12 to violent behavior at school. The child portrays in a virtual environment an invisible friend of a virtual character that has been a victim of violence. The child advises on the behavior to adopt. The aim is therefore to generate feelings of empathy of the child towards this virtual character [PAI 04].

Developing an ITS with an emotional dimension requires the ability to design emotionally expressive virtual characters and systems that are capable of recognizing user emotions. These capablities have been discussed (see sections 8.2.2 and 8.2.3).

To date, there has been little research analyzing the impact of affective tutors both with regard to learning performance and user satisfaction when interacting with an emotional ITS. This therefore remains an ongoing issue. Furthermore, questions remain about which emotional state is best for optimal learning. Once these questions have been answered, this type of knowledge should be included in ITS design.

8.5. Discussion and research perspectives

Today, the technology presented in this chapter still has major challenges. Despite this, recently there has been a veritable coming together of academic research and developing industry that, via cooperative projects has allowed a real technology transfer. This has required research teams to focus on issues such as operational application of systems that can recognize human emotions and/or include models of emotional processes such as real-time problems or even rejection thresholds in detection systems.

In spite of this, some issues remain. For example, in interactive vocal services, where a simple emotion recognition process becomes a part of sales approaches, aspects of architecture with other components (e.g. speech recognition, dialog management, vocal synthesis modules), performance and interface need to be taken into account.

If emotion recognition is applied to a system based on vocal dialog, this should not restrict overall system performance. The client should not have to wait too long for a response from the system. Furthermore, it is important to identify how emotion recognition can help fulfill the application's overall objective. This may mean that the system does not focus so much on detecting anger, for example, in vocal signals; but on the ability of the emotion-recognition module to identify the moment in the interaction where the caller's emotional state changes, which may then impact on the smooth running of the interaction with the system.

Furthermore, it is essential to account for the specificities of communication support and the user's characteristics. With regard to support, some technical aspects, such as network quality or screen size, can have a direct impact on performance; relevancy even on the importance on the emotion modules involved. For example, an emotional facial expression module for a virtual character may not be effective for human–virtual character interaction with cell phones due to the generally small size of the screen. With users, the age, type, knowledge and familiarity with technical tools are elements that are not always taken into account in affective computing but can strongly affect user perception and performance. A child today, being raised in a gaming environment, is already at greater ease than an adult when interacting with computer systems.

Beyond questions of mode, it seems undeniable that, in the future, machines (systems, interfaces, robots, etc.) will increasingly integrate concepts and functions from affective computing. Providing machines with emotional abilities is a dream that has existed for the past 40 years, fed by literature (Philip K. Dick's 1966 novel *Do Androids Dream of Electric Sheep*?), cinema (ranging from Stanley Kubrick's *2001: A Space Odyssey*, 1968, to Steven Spielberg's *A.I. Artificial Intelligence*, 2001) but also current research. While knowledge about emotion genesis and expression in living organisms is still in question, it is possible to rationalize a certain number of these organisms' functions using computational models and implement these in machines in order to establish an emotional relationship between machine and user [STE 03]. It also remains to be seen to what extent, but also in which contexts, humans will accept interaction with an emotionally intelligent machine.

8.6. Bibliography

[ABR 05] ABRILIAN S., DEVILLERS L., MARTIN J., "EmoTV1: Annotation of real-life emotions for the specification of multimodal affective interfaces", *HCI International*, Las Vegas, USA, 2005.

[ALE 06] ALEXANDER S., SARRAFZADEH A., "Easy with Eve: A functional affective tutoring system", *Workshop on Motivational and Affective Issues in ITS-8th International Conference on Intelligent Tutoring Systems*, Taiwan, 2006.

[ALL 06] ALLWOOD J., JOKINEN K., CERRATO L., NAVARRETTA C., PAGGIO P., "A coding scheme for the annotation of feedback turn management and sequencing phenomena", *5th International Conference on Language Resources and Evaluation*, Genoa, Italy, 2006.

[AND 01] ANDRÉ E., KLESEN P., GEBHARD P., ALLEN S., RIST T., "Integrating models of personality and emotions into lifelike characters", PAIVA A., (Ed.), *Affective Interactions: Towards a New Generation of Computer Interfaces*, pp. 150-165, Springer-Verlag, New York, 2001.

[ANG 02] ANG J., DHILLON R., KRUPSKI A., "Prosody-based automatic detection of annoyance and frustration in human-computer dialog", *International Conference on Spoken Language Processing*, Colorado, USA, 2002.

[AUB 03] AUBERGE V., AUDIBERT N., RILLIARD A., "Why and how to control the authentic emotional speech corpora", *EuroSpeech*, Geneva, Switzerland, 2003.

[BAD 02] BADLER N., ALLBECK J., ZHAO L., BYUN M., "Representing and parameterizing agent behaviors", *Proceedings of the Computer Animation Conference*, pp. 133-143, Geneva, Switzerland, 2002.

[BAL 02] BALL E., “A Bayesian heart: Computer recognition and simulation of emotion”, TRAPPL R., PETTA P., PAYR S., (Eds.), *Emotions in Humans and Artifacts*, pp. 303-332, MIT Press, Cambridge, 2002.

[BAN 96] BANDAI, *Tamagotchi Keychain Toy*, www.bandai.com, 1996.

[BAN 06] BANZIGER T., PIRKER H., SCHERER K., “Gemep - Geneva multimodal emotion portrayals: a corpus for the study of multimodal emotional expressions”, *LREC’06 Workshop on Corpora for Research on Emotion and Affect*, Genoa, Italy, 2006.

[BAR 02] BARTNECK C., eMuu - An embodied emotional character for the ambient intelligent home, PhD thesis, Eindhoven University of Technology, 2002.

[BAR 04] BARON-COHEN S., GOLAN O., WHEELWRIGHT S., HILL J., *Mind Reading: The Interactive Guide to Emotions*, Jessica Kingsley Publishers, Philadelphia, 2004.

[BAT 94] BATES J., “The role of emotion in believable agents”, *Communications of the ACM(CACM)*, vol. 37, no. 7, pp. 122-125, 1994.

[BAT 03] BATLINER A., FISCHER K., HUBER R., SPILKER J., NOTH E., “How to find trouble in communication”, *Speech Communication*, vol. 40, pp. 117-143, 2003.

[BAT 06] BATLINER A., STEIDL S., SCHULLER B., SEPPI D., LASKOWSKI K., VOGT T., DEVILLERS L., VIDRASCU L., AMIR N., KESSOUS L., AHARONSON V., “Combining efforts for improving automatic classification of emotional user states”, *Language Technologies*, pp. 240-245, Infornacijska Druzba (Information Society), 2006.

[BEC 05] BECKER C., WACHSMUTH I., PRENDINGER H., ISHIZUKA M., “Evaluating Affective Feedback of the 3D Agent Max in a Competitive Cards Game”, TAO J., TAN T., PICARD R.W. (Eds.), *Proceedings of the International Conference on Affective Computing and Intelligent Interaction (ACII)*, pp. 466-473, Springer, Beijing, China, 2005.

[BEV 04] BEVACQUA E., MANCINI M., PELACHAUD C., “Speaking with emotions”, *Proceedings of the Artificial Intelligence and Simulated Behavior (AISB) conference Symposium on Motion, Emotion and Cognition*, Leeds, UK, 2004.

[BOT 06] BOTHEREL V., MAFFIOLO V., “Regulation of emotional attitudes for a better interaction: field study in call centers”, *Human Factors in Telecommunication (HFT)*, Nice, France, 2006.

[BOW 92] BOWER G., “How might emotions affect learning?”, CHRISTIANSON S.-A. (Ed.), *Handbook of Emotion and Memory*, pp. 3-31, Lawrence Erlbaum Associates, Hillsdale, 1992.

[BRA 05] BRAVE S., NASS C., HUTCHINSON K., “Computers that care: Investigating the effects of orientation of emotion exhibited by an embodied computer agent”, *International Journal of Human-Computer Studies*, vol. 62, pp. 161-178, 2005.

[CAH 90] CAHN J. E., “The generation of affect in synthesized speech”, *Journal of the American Voice I/O Society*, vol. 8, pp. 1-19, 1990.

[CAN 03] CANAMERO L., "Designing emotions for activity selection in autonomous agents", in: TRAPPL R., PETTA P., PAYR S. (Eds.), *Emotions in Humans and Artifacts*, pp. 115-148, MIT Press, Cambridge, 2003.

[CAS 00] CASSELL J., SULLIVAN J., PREVOST S., CHURCHILL E., *Embodied Conversational Agents*, MIT Press, Cambridge, 2000.

[CHA 04] CHATEAU N., MAFFIOLO V., BLOUIN C., "Analysis of emotional speech in voicemail message: The influence of speaker's gender", *International Conference of Spoken Language Processing*, Jeju Island, Korea, 2004.

[CHA 05] CHATEAU N., MAFFIOLO V., "Les émotions et leur expression dans les interactions humaines et homme-machine", in: CHOQUET B., STERN D., *Cybermonde*, Hermes, Lavoisier, 2005.

[CHE 07] CHEN J., "Flow in games (and everything else)", *Communications of the ACM*, vol. 50, pp. 31-34, 2007.

[COW 01] COWIE R., DOUGLAS-COWIE E., TSAPATSOULIS N., VOTSIS G., KOLLIAS S., FELLENZ W., TAYLOR J., "Emotion recognition in human-computer interaction", *IEEE Signal Processing Magazine*, vol. 18, no 1, pp. 32-80, 2001.

[COW 03] COWIE R., CORNELIUS R., "Describing the emotional states that are expressed in speech", *Speech Communication*, vol. 40, pp. 5-32, 2003.

[DAR 02] DARVES C., OVIATT S., COULSTON R., "The impact of auditory embodiment on animated character design", *International Joint Conference on Autonomous Agents and Multi-Agent Systems, Embodied Agents Workshop*, Bologna, Italy, 2002.

[DAR 72] DARWIN C., *The Expression of the Emotions in Man and Animals*, John Murray, London, 1872.

[DEL 96] DELLAERT F., POLZIN T., WAIBEL A., "Recognizing emotion in speech", *International Conference on Spoken Language Processing*, Philadelphia, USA,1996.

[DER 03] DE ROSIS F., PELACHAUD C., POGGI I., CAROFIGLIO V., DE CAROLIS B., "From Greta's mind to her face: Modelling the dynamics of affective states in a conversational embodied agent", *International Journal of Human-Computer Studies*, vol. 59, no. 1-2, pp. 81-118, 2003.

[DOU 07] DOUGLAS-COWIE E., COWIE R., SNEDDON I., COX C., LOWRY O., MCRORIE M., MARTIN J., DEVILLERS L., ABRILIAN S., BATLINER A., AMIR N., KARPOUZIS K., "The HUMAINE Database: Addressing the collection and annotation of naturalistic and induced emotional data", *ACII*, pp. 488-500, Springer-Verlag, Berlin-Heidelberg, Lisbon, Portugal, 2007.

[DUY 04] DUY BUI T., Creating emotions and facial expressions for embodied agents, PhD thesis, University of Twente, 2004.

[EGG 05] EGGES A., MAGNENAT-THALMANN N., "Emotional communicative body animation for multiple characters", *First International Workshop on Crowd Simulation*, Lausanne, Switzerland, 2005.

[EKM 75] EKMAN P., FRIESEN W., *Unmasking the Face. A Guide to Recognizing Emotions from Facial Clues*, Prentice Hall Trade, 1975.

[EKM 76] EKMAN P., FRIESEN W., *Pictures of Facial Affect*, Consulting Psychologists Press, Palo Alto, 1976.

[EKM 78] EKMAN P., FRIESEN W., *Facial Action Coding System*, Consulting Psychologists Press, Palo Alto, 1978.

[EKM 93] EKMAN P., "Facial expression and emotion", *American Psychologist*, vol. 48, pp. 384-392, 1993.

[ELL 92] ELLIOT C., The affective reasoner: A process model of emotions in a multi-agent system, PhD thesis, Northwestern University, 1992.

[ELL 97] ELLIOTT C., RICKEL J., LESTER J., "Integrating affective computing into animated tutoring agents", *Proceedings of the IJCAI (International Joint Conference on Artificial Intelligence) Workshop on Animated Interface Agents: Making Them Intelligent*, Nagoya,1997.

[FRA 09] FRAISSARD G., "Jeux vidéo: les dernières innovations de Microsoft, Nintendo et Sony", *Le Monde*, June 2009.

[FRI 86] FRIJDA N., *The Emotions*, Cambridge University Press, Cambridge, 1986.

[GOL 97] GOLEMAN D., *L'Intelligence Émotionnelle*, Robert Laffont, Paris, 1997.

[GRA 04] GRATCH J., MARSELLA S., "A domain-independent framework for modeling emotion", *Journal of Cognitive Systems Research*, vol. 5, no. 4, pp. 269-306, 2004.

[HAR 05] HARTMANN B., MANCINI M., PELACHAUD C., "Implementing expressive gesture synthesis for embodied conversational agents", *Gesture in Human-Computer Interaction and Simulation, 6th International Gesture* Workshop, pp. 188-199, Vannes, France, 2005.

[HIL 03] HILL R., GRATCH J., MARSELLA S., RICKEL J., SWARTOUT W., TRAUM D., "Virtual Humans in the Mission Rehearsal Exercise System (2003)", *KI Embodied Conversational Agents*, 2003.

[HOO 02] HOOK K., "Evaluation of Affective Interfaces", *AAMAS Workshop on Embodied Conversational Agents: Let's Specify and Compare Them*, Bologna, Italy, 2002.

[ISE 00] ISEN A., "Positive affect and decision making", in: LEWIS M., HAVILAND-JONES J., (Eds.), *Handbook of Emotions*, pp. 417-435, Guilford Press, 2000.

[JOH 05] JOHNSON W., VILHJALMSSON H., SAMTANI P., "The tactical language training system", *Conference on Artificial Intelligence and Interactive Digital Entertainment*, Kos, Greece, 2005.

[JON 93] JONES M. R., YEE W., "Attending to auditory events: the role of temporal organization", in: *Thinking in Sound: The Cognitive Psychology of Human Audition*, pp. 69-112, Oxford University Press, New York, 1993.

[KEL 01] KELTNER D., HAIDT J., "Social functions of emotions", in: MAYNE T. J., BONANNO G. A., Eds., *Emotions: Current issues and Future Directions*, pp. 192-213, The Guilford Press, New York, 2001.

[KER 00] KERBRAT-ORECCHIONI C., "Quelle place pour les émotions dans la linguistique du XXème siècle. Remarques et aperçus", *Les émotions dans les interactions*, pp. 33-74, Presses Universitaires de Lyon, Lyon, 2000.

[KIP 01] KIPP M., "Anvil - A generic annotation tool for multimodal dialogue", *7th European Conference on Speech Communication and Technology*, Aalborg, Denmark, 2001.

[KLE 99] KLEIN J., MOON Y., PICARD R., "This computer responds to user frustration", *Proceedings of the Conference on Human Factors in Computing Systems*, ACM Press, NewYork, pp. 242-243, 1999.

[KNA 06] KNAPP M. L., HALL J. A., *Nonverbal Communication in Human Interaction*, Thomson Wadsworth, Belmont, 2006.

[LAZ 91] LAZARUS R. S., *Emotion and Adaptation*, Oxford University Press, New York,1991.

[LEC 06] LE CHENADEC G., MAFFIOLO V., CHATEAU N., COLLETTA J., "Creation of a corpusof multimodal spontaneous expressions of emotions in human-machine interaction", *5th International Conference on Language Resources and Evaluation*, Genoa, Italy, 2006.

[LEC 07] LE CHENADEC G., MAFFIOLO V., CHATEAU N., "Analysis of the multimodal behaviour of users in HCI: the expert viewpoint of close relations", *4th Joint Workshop onMultimodal Interaction and Related Machine Learning Algorithms (MLMI)*, Washington, USA, 2007.

[LEE 04] LEE C., YILDIRIM S., BULUT M., "Emotion recognition based on phoneme classes", *International Conference on Spoken Language Processing*, Jeju Island, Korea, 2004.

[LEE 05] LEE C., NARAYANAN S., "Toward detecting emotions in spoken dialogs", *IEEE Trans. on Speech and Audio Processing*, vol. 13, no. 2, pp. 293-303, 2005

[LES 00] LESTER J., TOWNS S., CALLAWAY C., VOERMAN J., FITZGERALD P., "Deictic and emotive communication in animated pedagogical agents", *Embodied Conversational Agents*, pp. 123-154, MIT Press, Cambridge,2000.

[MA 05] MA C., PRENDINGER H., MITSURU I., "Emotion estimation and reasoning Basedon affective textual interaction", *First International Conference on Affective Computing and Intelligent Interaction (ACII)*, Beijing, China, 2005.

[MAF 06] MAFFIOLO V., DAMNATI G., BOTHEREL V., GUIMIER DE NEEF E., MAILLEBUAU E., "Multilevel features for annotating application-driven spontaneous speech corpora", *Language Resources and Evaluation (LREC)*, Genoa, Italy, 2006.

[MAF 07] MAFFIOLO V., CHATEAU N., LE CHENADEC G., "Temporal hierarchical organizationin listeners' perception of the speakers' emotions and characteristics: a way toimprove the automatic recognition of emotion-related states in human voice", *Proceedings of the International Conference Affective Computing and Intelligent Interaction*, Lisbon, Portugal, 2007.

[MAN 01] MANSTEAD A., FISCHER A., "Social appraisal: The social world as object of and influence on appraisal processes", in: SCHERER K. R., SCHORR A., JOHNSTONE T. (Eds.), *Appraisal Processes in Emotion: Theory, Research*, Application, pp. 221-232, Oxford University Press, Oxford, 2001.

[MAR 06] MARTIN J.-C., CARIDAKIS G., DEVILLERS L., KARPOUZIS K., ABRILIAN S., "Manual annotation and automatic image processing of multimodal emotional behaviours: Validating the annotation of TV interviews", *Fifth International Conference on Language Resources and Evaluation (LREC 2006)*, Genoa, Italy, 2006.

[MEL 08] MELLO S., JACKSON T., CRAIG S., MORGAN B., CHIPMAN P., WHITE H., PERSON N., KORT B., EL KALIOUBY R., PICARD R., GRAESSER A., "AutoTutor detects and responds to learners affective and cognitive states", *Workshop on Emotional and Cognitive Issues at the International Conference of Intelligent Tutoring*, Montreal, Canada, 2008.

[MIN 86] MINSKY M., *The Society of Mind*, Simon and Schuster, New York, 1986.

[NIJ 01] NIJHOLT A., "Towards multi-modal emotion display in embodied agents", *5th Biannual Conference on Artificial Neural Networks and Expert Systems*, University of Otago, Dunedin, New Zealand, 2001.

[OCH 08] OCHS M., PELACHAUD C., SADEK D., "An empathic virtual dialog agent to improve human-machine interaction", *Proceedings of the International Conference Autonomous Agent and Multi-Agent Systems (AAMAS)*, Estoril, Portugal, 2008.

[OCH 09] Ochs M., Sabouret N., Corruble C., "Simulation de la dynamique des émotions et des relations sociales de personnages virtuels", Revue d'Intelligence Artificielle (RIA), Numéro spécial Modélisation et Décision pour les Jeux, vol. 23, no. 2-3, pp. 327-357,2009.

[OLI 03] OLIVEIRA E., SARMENTO L., "Emotional advantages for adaptability and autonomy", in: ROSENSCHEIN J., WOOLDRIDGE M., Eds., *Proceedings of International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, pp. 305-312, ACM Press, Melbourne, Australia, 2003.

[ORT 88] ORTONY A., CLORE G., COLLINS A., *The Cognitive Structure of Emotions*, Cambridge University Press, Cambridge, 1988.

[OUD 02] OUDEYER P.-Y., "Novel useful features and algorithms for the recognition of emotions in human speech", *First Conference on Speech Prosody*, Aix en Provence, France, 2002.

[PAI 01] PAIVA A., MACHADO I., PRADA R., "Heroes, villains, magicians: Dramatics personaein a virtual story creation environment", *Proceedings of the Conference on Intelligent User Interfaces (IUI)*, pp. 129-136, Sante Fe, USA, 2001.

[PAI 04] PAIVA A., DIAS J., SOBRAL D., WOODS S., HALL L., "Building empathic lifelike characters: the proximity factor", PAIVA A., AYLETT R., MARSELLA S., Eds., *International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, Workshop on Empathic Agents, New York, USA, 2004.

[PAN 03] PANDZIC I., FORCHHEIMER R., Eds., *MPEG-4 Facial Animation: The Standard, Implementation and Applications*, Wiley, New York, 2003.

[PAN 07] PANTIC M., PENTLAND A., NIJHOLT A., HUANG T., "Machine understanding of human behavior", *20th International Conference on Artificial Intelligence, International Workshop on AI for Human Computing AI4HC'07*, Hyderabad, India, 2007.

[PAR 04] PARTALA T., SURAKKA V., "The effects of affective interventions in human computer interaction", *Interacting with Computers*, vol. 16, pp. 295-309, 2004.

[PEL 02] PELACHAUD C., CAROFIGLIO V., CAROLIS B. DEROSIS F., POGGI I., "Embodied contextual agent in information delivering application", *First International Joint Conference on Autonomous Agents & Multi-Agent Systems*, Bologna, 2002.

[PET 95] PETZ, *Dogs and Cats: Your virtual Petz*, www.petz.com, 1995.

[PIC] PICARD R., *affect.media.mit.edu*, Affective Computing Lab.

[PIC 97] PICARD R., *Affective Computing*, MIT Press, Cambridge, 1997.

[PIC 01] PICARD R. W., VYZAS E., HEALEY J., "Toward machine emotional intelligence:Analysis of affective physiological state", *IEEE Transactions on Patterns Analysis and Machine Intelligence*, vol. 23, pp. 1175-1191, 2001.

[PIC 07] PICARD R., LIU K., "Relative subjective count and assessment of interruptive technologies applied to mobile monitoring of stress", *International Journal of Human-Computer Studies*, vol. 65, pp. 396-375, 2007.

[POL 99] POLZIN T., Detecting verbal and non-verbal cues in the communication of emotions, PhD thesis, Carnegie Mellon University, Pittsburg, 1999.

[PRE 02] PRENDINGER H., DESCAMPS S., ISHIZUKA M., "Scripting affective communication with life-like characters in web-based interaction systems", *Applied Artificial Intelligence Journal*, vol. 16, no. 7-8, pp. 519-553, 2002.

[PRE 05] PRENDINGER H., MITSURU I., "The empathic companion: A character-based interface that addresses users' affective states", *Applied Artificial Intelligence Journal*, vol. 19, pp. 297-285, 2005.

[REE 96] REEVES B., NASS C., *The Media Equation: How People Treat Computers, Television,and New Media Like Real People and Places*, Cambridge University Press, New York, 1996.

[REI 96] REILLY S., Believable social and emotional agents, PhD thesis, Carnegie Mellon University, 1996.

[RIZ 00] RIZZO P., "Why should agents be emotional for entertaining users? A critical analysis", *Affective Interactions: Towards a New Generation of Computer Interfaces*, pp. 166-181, Springer-Verlag, New York, 2000.

[RUT 02] RUTTKAY Z., DOORMAN C., NOOT H., "Evaluating ECAs - What and how?", *AAMAS02 Workshop on 'Embodied conversational agents - let's specify and evaluate them!'*, Bologna, Italy, 2002.

[RUT 04] RUTTKAY Z., PELACHAUD C., *From Brows to Trust: Evaluating Embodied Conversational Agents*, Kluwer Academic Publishers, Norwell, 2004.

[SAL 00] SALOVEY P., BEDELL B., DETWEILER J., MAYER J., "Current directions in emotional intelligence research", in: LEWIS M., HAVILAND-JONES J. (Eds.), *Handbook of Emotions*, pp. 504-520, The Guilford Press, New York, 2000.

[SCH 00] SCHERER K., "Emotion", in: HEWSTONE M., STROEBE W. (Eds.), *Introduction to Social Psychology: A European Perspective*, pp. 151-191, Blackwell Publishers, Oxford, 2000.

[SCH 04a] SCHROEDER M., Speech and emotion research: An overview of research frameworks and a dimensional approach to emotional speech synthesis, PhD thesis, Saarland University, 2004.

[SCH 04b] SCHULLER B., RIGOLL G., LANG M., "Speech emotion recognition combining acoustic features and linguistic information in a hybrid support vector machine – belief network architecture", *International Conference on Acoustics, Speech, and Signal Processing*, pp. 577-580, Montreal, Canada, 2004.

[SON 99] SONY, *Aibo Robot Dog*, www.aibo.com, 1999.

[STE 03] STERN A., "Creating emotional relationships with virtual characters", in: TRAPPL R., PETTA P., PAYR S., Eds., *Emotions in Humans and Artifacts*, pp. 333-361, MIT Press, Cambridge, 2003.

[STR 04] STRAPPARAVA C., VALITUTTI A., "WordNet-Affect: an affective extension of WordNet", *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC 2004)*, Lisbon, Portugal, 2004.

[SYK 03] SYKES J., BROWN S., "Affective gaming: measuring emotion through the gamepad", *CHI '03 Extended Abstracts on Human Factors in Computing Systems*, ACM, New York, pp. 732-733, 2003.

[THO 81] THOMAS F., JOHNSTON O., *The Illusion of Life: Disney Animation*, Disney Editions, New York, 1981.

[VEL 98] VELASQUEZ J. D., "Modeling emotion-based decision-making", CANAMERO L., (Ed.), *Emotional and Intelligent: The Tangled Knot of Cognition*, pp. 164-169, AAAI Press, Orlando, 1998.

[VER 06] VERVERIDIS D., KOTROPOULOS C., "Emotional speech recognition: Resources, features, and methods", *Speech Communication*, vol. 48, pp. 1162-1181, 2006.

[VID 05] VIDRASCU L., DEVILLERS L., "Annotation and detection of blended emotions in real human-human dialogs recorded in a call center", *IEEE International Conference on Multimedia & Expo '05*, Amsterdam, The Netherlands, 2005.

[ZAR 07] ZARA A., MAFFIOLO V., MARTIN J.-C., DEVILLERS L., "Collection and annotation of a corpus of human-human multimodal interactions: Emotion and others anthropomorphic characteristics", *Affective Computing and Intelligent Interaction*, Lisbon, Portugal, 2007.

[ZAR 09] ZARA A., MAFFIOLO V., ADAM J., DUBOIS M., “From mediated human-human interactions to human-virtual character interactions: Selection and vategorization of expressive gestures”, *International Gesture Workshop*, Bielefield, Germany, 2009.

[ZEN 07] ZENG Z., HU Y., ROISMAN G., WEN Z., FU Y., HUANG T., “Audio-visual Spontaneous Emotion Recognition”, in: *Artificial Intelligence for Human Computing*, pp. 72-90, Springer, New York, 2007.

Chapter 9

Music and Emotions

9.1. The growing importance of music in society

With the low cost and increasing simplicity of production technology, music is now omnipresent in our daily lives. From public transport to concert venues, theatres or even elevators, it is now difficult to spend even two hours without hearing music [SLO 01]. According to a recent report, time spent listening to music now exceeds that spent on other activities, such as watching television or going to the movies [REN 03]. As Zentner has suggested, "music must be uniquely rewarding to its listeners" for it to occupy such an important space in our daily lives [ZEN 08]. However, what is so special about music? What makes it so satisfying to humans?

Since antiquity, the emotions evoked by music have been proposed as a possible answer to these questions. Aristotle, inspired by Plato's *The Republic*, gave a relatively detailed description of the emotional effects of each type of music. For example, while the Mixolydian mode tends to make people sad, the Phrygian mode makes them enthusiastic or happy (*Politics*, book VIII). At the start of the 17th Century, Monteverdi revived this idea with a *rappresentativo* style, which encouraged writing freedom (dissonance, rhythm, etc.) and strengthened the expression of feeling. Since Plato, and in parallel with the development of musical practice, music and emotions have been a favorite subject of philosophy, as

---

Chapter written by Donald GLOWINSKI and Antonio CAMURRI.

Emotion-Oriented Systems Edited by Catherine Pelachaud

demonstrated by the impressive number of theories on the subject [BUD 85, DAV 94b, ROB 05].

For nearly a century, theories on the ability of music to relieve, regulate or alter mood and emotion have been confirmed by empirical research [JUS 01, JUS 02, SCH 01a]. The majority of studies have focused on how listeners perceive emotions expressed in music and on the link with these perceived emotions and musical and sound parameters [CLY 77, COO 59]. More recently, studies have explored situations in which music demonstrates its emotional power since, for example, music is an effective means of inducing emotions in laboratories [VAS 02, WES 96]. Similarly, in our daily lives, music is used to modify consumer behavior [ALP 05, BRU 90] or as a tool for treating emotional disorders in therapeutic applications [GOL 04].

The emotional power of music has also been noted among very young subjects. For example, Zentner and Kagan have noted emotional reactions to music in infants as young as four months [ZEN 96, ZEN 98]. Indeed, the intensity of pleasure gained from music cannot be rivaled by other more "mundane" sources of pleasure. Blood, Zatorre, Meon and Levetin have discovered that the regions of the brain activated by the emotions evoked when listening to music are similar to those felt by those activated by intense pleasure such as sex, food and drugs [BLO 01, MEN 05]. These studies thus confirm that music is generally appreciated because of the emotional reward that it gives listeners [ZEN 08].

The central role of emotions in music can explain the interest in developing new computer interfaces, specifically "emotional" interfaces [PIC 97]. Indeed, the possibility of using music to influence emotional reactions (or vice versa) and modeling users' emotional reactions to modify music in real time offer new perspectives on expressive interactions with the computer. In this chapter, we will examine the recently developed computer systems (interfaces) that allow musical synthesis through emotions. In this field we can distinguish between two types of research. The first focuses on communicating specific emotions using computerized modifications of sound samples and pieces of music. The second concentrates on the possibility of interaction between humans and musical content, where this is synthesized by the computer. We will also examine the theories and models of emotions that have contributed to the development of these new musical interfaces. Note that we will keep two essential aspects of emotion that are the subject of consensus within the scientific community, that:

– emotion is a reaction to events that are considered important to an individual's needs, aims and interests; and

– emotion includes physiological, behavioral and cognitive components [IZA 77].

9.2. Recognizing emotions and structural characteristics in music

The emotions recognized in a piece of music have two sources: the composer and the interpreter. The perception and triggering of an emotion by music depends on numerous multiple factors in interaction, however, such as sound parameters, audience, performance and context [SCH 01a]. The first of these factors (sound parameters) has received the most attention and has given way to several successful implementations. We will now examine some of the methods employed to link emotion recognition and sound parameters with their results and use in the creation of computational models.

9.2.1. *Understanding listeners' emotional reactions*

The most common method of gauging listeners' emotional reactions involves asking participants to listen to music and then report their impressions in a questionnaire. The responses can take the form of: (i) free comments; (ii) choice from a set list of emotions (multiple choice questionnaries); or (iii) assessing the ability of a term to describe the music in question. This type of research has revealed that there are a number of elements in sound aspects that modify emotional responses when listening to music. These elements are links to *score features*, e.g. pitch (high/low), intervals (short/long), harmony (consonant/dissonant) and rhythm (regular/irregular). The dynamic between these sound elements is also important and depends largely on instrumental interpretation (or performance features), e.g. rhythmic accents, articulation (staccato/legato), variations in timbre (spectral richness, playing mode, etc.). For a more detailed overview, see Gabrielsson and Lindstrom [GAB 01b].

A number of major studies have focused on the *perception* of emotion by listeners. Sloboda, for example, has highlighted that it is possible to judge music on the basis of conventional characteristics without experiencing the emotions in question [SLO 01]. Furthermore, Kallinen *et al*. have shown that emotions perceived as negative by the listener can also be felt as positive by others [KAL 05].

Recent laboratory studies have compared emotion perception and induction using musical extracts that are unknown to the participants. Interestingly, these studies show that the emotions evoked are mostly similar to the emotions perceived [KAL 06, SCH 07].

This research, based on analyzing subjective comments, has revealed a large number of sound parameters that had previously been thought of as present in vocal modulations used to evoke emotion. An analysis of 142 scientific articles on vocal expression (101 articles) and musical interpretation (42) have allowed Juslin and

Laukka to identify several similarities [JUS 03]. Three acoustic parameters in particular evolve similarly in vocal and musical expression. For example, the rate of speech and rhythm, vocal intensity and sound intensity increase for anger and happiness while they decrease for sadness and tenderness. Similarly, energy at high frequencies of the spectrum (a component of timbre) increase in vocal and instrumental expressions of happiness and anger. According to Juslin, the similarity between these two forms of expressions justifies the evolutionist approach to music: expressive qualities of music are derived from vocal strategies developed through evolution to communicate emotion or regulate the emotional states of those around us (e.g. mother and child) [JUS 03b, TRA 08].

9.2.2. *A categorical or dimensional approach?*

Whether a methodological choice or theoretical postulate, the terms used in questionnaires given to gain listeners' opinions almost invariably relate to the basic (also called discrete) emotions, such as anger, fear, surprise, happiness or sadness (this is a categorical approach). In other cases, the terms used to describe these emotions refer to the level of intensity (arousal) and positivity (valence) – the dimensional approach– two dimensions normally found in Russel's *affective circumplex* model and its variants [RUS 80]. Studies in this area therefore often rely on a categorical or dimensional approach. However, as Zentner *et al*. have demonstrated [ZEN 08], these approaches can be inadequate for characterizing musically-induced emotions. Music therefore seems to be able to induce a wider range of emotional states than these semantic descriptors (labels) suppose.

The semantic descriptors resulting from these two canonical approaches do not do justice to the range of emotions evoked by music. Several attempts have been made to create descriptors that are adapted to the specific phenomenological situation of listening to music [ASM 85, RIG 64]. As Scherer has noted [SCH 01], however, these approaches are eclectic by their very nature, with a choice of emotional descriptors that are highly reliant on researchers' own personal perceptions.

A systematic and an empirical taxonomy of terms also need to exist, which describe emotions in relation to music. The recent statistical studies carried out by Zentner *et al*. which involved asking 1,393 participants to classify semantic descriptors in music extracts from various genres (classical, rock, jazz, etc.) was a first step towards this [ZEN 08].

These studies highlight that terms such as *melancholy* or *enchantment* describe listeners' experiences better than the usual terms of *sadness* or *happiness*. Studies based on a categorical or dimensional approach (largely the majority of studies) therefore do not consider the specificity of emotions evoked by music. They have, however, revealed some musical characteristics that can evoke emotional reactions in the listener.

9.3. Rules for modeling musical expression of emotions

The identification of themes and structures that are characteristic of some emotions (often the basic emotions) has enabled the development of models, rules regulating musical emotional expression and the implementation of expressive automatic musical synthesis systems [GAB 03, GAB 01a, GAB 03, JUS 03a, SCH 99]. In this vein, Sundberg, Friberg, Bresin and their team have developed a rules-based interpretation system called the *Director Musices-KTH rule system*, which generates expressive musical interpretations [BRE 00, FRI 95, FRI 99, FRI 00, FRI 06a].

These rules modify different aspects of musical interpretations, such as phrasing for example. A musical phrase normally describes a pattern in an arc with a slow and soft beginning followed by acceleration, growing in intensity in the middle and changing a slow conclusion *in diminuendo*. The *phrase arch rule* modifies the characteristics of this theme and therefore determines the phrase's temporal and dynamic progression [GAB 87, REP 92]. Another aspect of interpretation on which these rules are based is articulation. Articulation determines how a note is played, i.e. *legato* or *staccato*. The *punctuation rule* selects short fragments of melody, identifies the last note, lengthens it, follows it with a short pause and emphasizes or decreases the section's melodic continuity.

Finally, the *duration contrast* rule acts on the duration that separates the start of a note from the beginning of the next note. The musical effect of these rules is to lengthen the long notes and shorten the short notes, therefore increasing the contrast in length between them [FRI 91, SUN 82]. The punctual application of these rules and the modification of more general parameters, such as rhythm, sound intensity and articulation, on the whole of the section have allowed Bresin and Friberg to generate various types of emotional expressions. Four of these are illustrated in Figure 9.1, which includes happiness, sadness, anger and love [BRE 00].

These music expressivity rules were designed to "humanize" the computer's performance on MIDI (*Musical Instrument Digital Interface*) scores for almost 20 years [FRI 95, FRI 06a]. The MIDI format is widely used in music technology to code information sequences related to the musical set but does not include expressive indication and "mechanical" musical reproductions. The model developed by Friberg *et al.* is now used in several software applications that can automatically "animate" a music score.

	Happiness	Sadness	Anger	Love
Overall changes				
Tempo	somewhat fast	slow	fast	slow
Sound level	medium	low	high	low
Articulation	*staccato*	*legato*	somewhat *staccato*	*legato*
Rules				
Phrase arch	small	widened	negative	small
Final *ritardando*	minimal			minimal
Punctuation	large	reduced	average	reduced
Duration contrast	large	negative	large	

Figure 9.1. *Parameters of musical expressivity rules suggested by Friberg et al. to model emotional content of musical synthesis [FRI 06b]*

The evaluation procedure for these rules takes place across four stages. The first stage consists of selecting and encoding a set of samples in the MIDI format, most commonly taken from Western classical music or popular music used in video games [LIV 07]. Second, the piece is played to participants in a 'neutral' form, i.e. without any expressive tonality. At the third stage, the section is reproduced with several modifications by applying the model's rules in successive listenings. Finally, the participants are asked to categorize the reproductions according to the expressive nuances that they have recognized.

Beyond its practical uses, i.e. making computerized scores expressive, the application of these rules illustrates the success of analysis-by-synthesis methods, which have proven effective for verifying how some musical parameters determine which emotion is perceived and felt [CAN 04, GAB 85, POL 04].

9.4. Towards a continuous measure of emotional reactions to music

The recognition of emotions in a musical context is frequently measured using questionnaires. The use of this technique, used as standard in psychology, can underestimate the profoundly temporal nature of music [MEY 56, MEY 67] and the possibility that emotion can vary while listening to the same piece of music. Alternative procedures have also been developed to provide a continuous measure of subjects' emotional reactions [NIE 83, SCH 01b].

Nelson, for example, has developed a research technique that consists of recording tensions felt by a subject while listening to a piece of music. During this process, the participant is asked to press on levers with a force comparable to the tension experienced. The more tension the participant feels, the more pressure he/she applies to the levers. Equally, as he/she relaxes, the tension decreases.

Several experiments have supported using a measure of tension as a reference technique for evaluating emotions in musical contexts. Madsen and Fredrickson [MAD 93], for example, have analyzed the correlations between a measure of continuous tension and aesthetic judgments by participants while listening to classical music (in this case the first movement of Haydn's 104th symphony) [MAD 93]. Their study confirmed that tension can be considered as an overall attribute of musical experience. Musical factors in tension have also been identified, such as texture, dynamics, intensity, harmonic relations and implicit expectations, based on the experience that each acquires in a musical genre [BIG 06, KRU 96, MEY 56]. Participants' age, musical abilities and level of familiarity with the musical stimulus being used also affect evaluations of tension [VIN 06].

Krumhansl, on the other hand, has noted that judgements about the force of an emotion correlate significantly with evaluations of tension [KRU 97a]. His study involved asking participants to listen to several pieces of music that were characteristic of three basic emotions – including sadness (Albinoni's *Adagio*), fear (Moussorgsky's *La nuit sur le Mont Chauve*) and happiness (Spring in Vivaldi's *Four Seasons*). Participants were asked to adjust the position of a mouse cursor to describe how the force of the emotion varied temporally while listening to the section of music. Through this, Krumhansl has demonstrated that tension can serve as a more accurate indicator of emotion than participants' emotional reactions by correlating this measure of tension to physiological changes including cardiac, vascular, respiratory and electro dermal functions [CHA 08, KRU 97a].

9.5. Multimodality in musical experience

Whether through questionnaires or continuous tension measures, studies have confirmed that emotions can be communicated via music [GAB 96, JUS 97a, JUS 97b]. In this way, some sound and musical characteristics have been highlighted in experiments as conveyors of this emotional communication. However, there is a multitude of different styles and means of expressing emotion that seem to be an obstacle to modeling based on these musical structures alone. Musical expression involves several modes of expression, as do expressions of emotions generally [DAV 03, LIV 06, SCH 85, WAL 86].

A series of studies have shown that emotions in music are communicated not only via sound but by facial expressions and other bodily movements. Adopting an ethnographic and historical view, Thompson, Graham and Russo [THO 05] have highlighted that visual information influences our perception and experience of music. The influence of bodily movements on the perception of music has been empirically studied [DAH 07, HEN 07, TIM 06, VIN 04, WAN 05]. Similarly, Dahl and Friberg have demonstrated how emotions can be communicated by musicians' movements (in this case, a Marimba player, a saxophonist and a bassoonist) [DAH 07]. In particular, they have identified the role of different parts of the body during the expression of an emotion. For example, movements of the head play a predominant role. Similar results have also been found by Davidson [DAV 94a] in an experiment where he showed participants videos of pianists, and observed that participants paid the greatest attention to the head to identify different expressive tonalities of musical interpretations.

In a series of studies carried out into the perception of how clarinetists play, Vines *et al*. have used a measure of continuous tension to approach the problem of multimodal integration (mutual influence of audio and visual information) in emotion recognition [VIN 06]. A pioneering study led by Krumhansl and Schenck [KRU 97b] using this continuous measure had already identified a similarity between emotional reactions of three groups of participants that had (i) a ballet video; (ii) a sound track; and (iii) the video and soundtrack played to them. The reaction of the third group in which participants watched and listened could even be modeled as a combination of the judgments made by the participants in the two other conditions.

The significance of the visual modality has been re-verified by a study by Vines *et al*. in which 30 participants split into three groups: (i) watched a video; (ii) listened to sound track; and (iii) watched and listened to an interpretation of the second part of Stravinsky's clarinet solo performed by two clarinetists. The participants used a sliding potentiometer to evaluate continuous tension (a measure correlated with emotion). The application of a functional analysis method using FDA (functional data analysis) data [PAS 07, VIN 05] correlated the dynamic of participants' reactions with the pieces' progression in detail. The results showed that visual and audio modalities convey emotional information independently, but also relay musical structure almost identically. They showed that musicians' movements had a decisive role in increasing or reducing the emotional "impact" of music and in anticipating changes in emotional intensity.

These studies confirm that recognizing emotions during a musical performance is a multisensory experience in which sound and images interact. Both the music itself and watching the musician playing it have a role in listeners' emotional inference and induction processes. The understanding of visual and sound interactions supposes, however, that we can characterize each of the elements at work in emotional communication in detail. If we adopt the Brunswickian theoretical perspective used by Juslin in musical contexts [JUS 00], we can explain emotional communication using the relation between the *sender* and the *receiver*.

Analysis must therefore not only consider perception (by focusing on the proximal indicators perceived by the *receiver*) but also expression (by showing the distal indicators produced by the *sender*) [JUS 97b]. As we have seen, elements linked to musical structure have been examined in detail although the specific nature of gestures contributing to the expressivity of musical content has remained unexplored.

The relative absence of a shared protocol for encoding the entirety of bodily movements, as there is for facial expression [EKM 84], has in part limited the scope of quantitative studies. Recent technological developments have enabled the development of motion capture and real-time analysis instruments that have allowed us to characterize gestural dynamics and their relation to expressivity in an unprecedented way (www.vicon.com and www.eyesweb.org). The correlation between bodily movements and the acoustic characteristics of musical interpretation has been studied by Dahl [DAH 07]. Interpretations may have a happy or sad tonality that is characterized by movements of the body corresponding more widely to higher intensity sounds. With interpretations of a sad tonality, more fluidic movements underlie a *legato* musical articulation. In a recent video analysis of the bodily movements of a cellist and pianist, Castellano *et al*. have not only highlighted the importance of head movements in communicating emotion, but have also specified the expressive dynamics of musical performance. A growing number of studies are now focusing on identifying multimodal characteristics and, in particular, gestural characteristics of musical expression [GLO 08, LEM 07, TIM 06, WAN 05]. The results of these studies have already been used to develop new musical expression interfaces [BER 08, CAS 07].

9.5.1. ***A multimodal research platform for musical expression***

In order to best use new technological resources for analyzing movement and music in order to understand and model musical expressivity, Camurri *et al*. have created a multimodal research platform. This platform allows us to explain the different stages needed to analyze modes of musical expression and, specifically, those linked to gesture.

Expressive gestures: The notion of expressive gestures is central in this type of analysis [CAM 05b]. An attempt to define the term is provided by Camurri *et al*. based on the conception by Kurtenbach and Hulteen of gesture as a “movement of the body which contains information” [KUR 90]. From this perspective, expressive gesture is characterized by the fact that the information it contains and conveys is linked to emotion. This content superimposes itself using possible denotive meanings (for example a finger pointed towards an object to identify it) but is independent of it. Based on this point of view, gesture can be considered as expressive in the sense that it conveys what Cowey *et al*. have called “implicit messages” [COW 01] and what Hashimot has called “kansei” [HAS 97].

High-level information on expressivity: emotion recognition, strength of audience and emotional reaction.
Level 4 – Concepts and structures Modeling techniques (i.e. classification as basic emotions, predicting the intensity of the audience’s emotional reaction): algorithms for multiple regression, neural networks, support vector machines, decision trees, and Bayesian networks.
[Input] Segmented gestures and accompanying parameters (i.e. directivity, impulsiveness, rigidity, etc.). Trajectories representing gestures in semantic spaces.
Level 3 – Intermediary characteristics and cards Gesture segmentation techniques: segmentation of movement (distinction between pauses and phases of movements), segmentation of musical extracts into musical phases. Representing gestures as trajectories in semantic spaces (Laban’s *effort* space, “energy-articulation” space).
[Input] Descriptors of movement or sound content, i.e. amount of energy –intensity, amount of expansion of the body – spectral frequency, melodic contours, fluidity, coarseness, etc.
Level 2 –Low-level characteristics Computer processing the signal (audio and video signal) and statistical measures.
[Input] Preprocessed images for detecting point movement and trajectories (i.e. trajectories of different parts of the body, musicians in a given space), preprocessed MIDI data and audio signals for detecting low-level spectral and temporal characteristics.
Level 1 – Physical signal Analyzing the video and audio signal: image segmentation techniques (background subtraction), movement detection, monitoring movement (i.e. detection and monitoring techniques for color or optical flux) and audio filtering techniques.
[Input] Data from several sensors, i.e. data from TV cameras, positions of localization systems, accelerometer data, audio signals, MIDI messages, etc.

Figure 9.2. *Diagram of a quadripartite conceptual framework for analyzing expressive gestures*

In the framework of the European *Multisensory Expressive Gesture Applications* (MEGA) project, Camurri *et al*. have developed a quadripartite conceptual framework for analyzing expressive gestures [CAM 05a]. The analysis process can be summarized in four main points (see Figure 9.2):

(i) identifying the type of data characterizing the musician’s activity or that of the user;

(ii) then identifying an appropriate collection of descriptors to describe the expressive gesture;

(iii) implementing these descriptors in the form of algorithms to obtain a continuous description of the movement; and

(iv) analyzing the data obtained by applying these descriptors to obtain high-level information characterizing the gestures' expressivity.

9.5.1.1. *Level 1*

Level 1 (the physical signal) corresponds to techniques used to collect data recorded by sensors (video cameras, micro sensors, sensory sensors, *on-body sensors*). Several sources are used to get a detailed description of the movement in question.

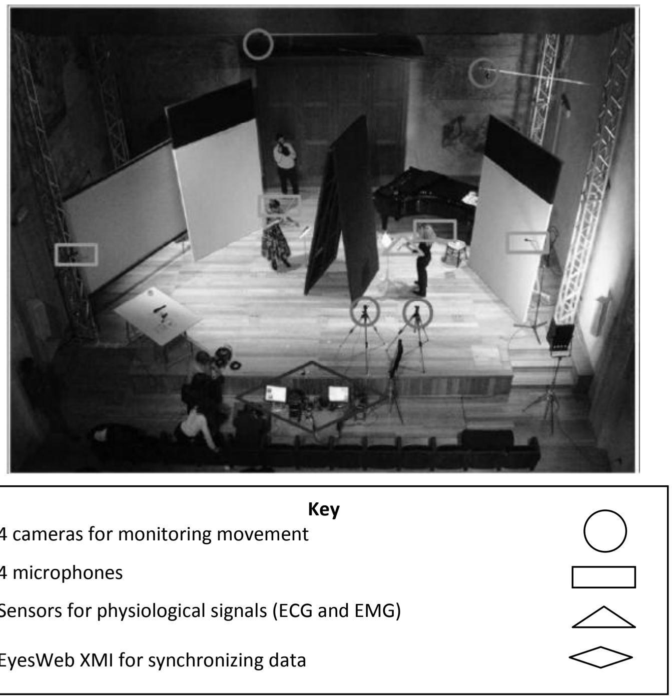


Figure 9.3. *An example of experimental tools for recording multi-sensory data in a musical performance in real time (using two violinists, Chiara Noera (right) and Lucid Marucci (left), in the Casa Paganini Research Centre [GLO 08]*

This first level is illustrated in a recent experiment called the "Premio Paganini" [GLO 08], which was designed for studying multimodal characteristics of musicians' emotional processes in concert. In a concert hall (the *Casa Paganini* in Genoa) and a public event (the *Premio Paganino* international violin competition), a series of video (cameras), audio (microphones) and physiological (EGG (electroencephalogram signal), GSR (Galvanic skin response), EMG (electromyogram signal)) sensors were installed to analyze the violinists performances (see Figure 9.3). New informatic solutions based on the EyesWeb software platform were developed to synchronize and visualize the data coming from the sensors (see Figure 9.4).

9.5.1.2. *Level 2*

Level 2 (the descriptors) contains the series of descriptors of the type of movement and are based on the measures taken from level 1. In the case of musical interpretation, for example, the descriptors include cinematic measures: the amount of movement, the amount of contraction/expansion of the arms in relation to the body, etc. These descriptors come from research in psychology and artistic practice. For a more detailed examination of this, see [BLA 07, BOO 98, WAL 98]. Reference to Rudolf Laban's effort theory has allowed us, for example, to model the qualitative dimensions of gestures involved in expressivity such as rigidity or fluidity [LAB 47, LAB 63].

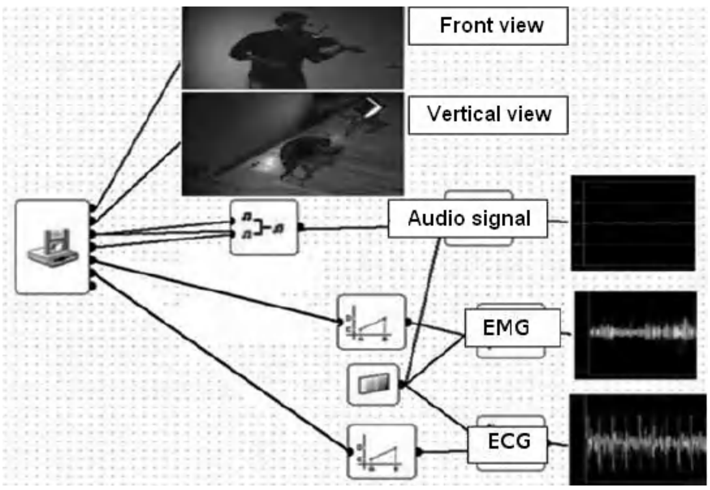


Figure 9.4. *The use of EyesWeb XMI to synchronize and visualize multisensory data of musical performances in real time (using the violinist Alessio Gabriele in the Casa Paganini Research Centre) [GLO 08]*

9.5.1.3. *Level 3*

Level 3 (segmentation and representation) concerns two stages: first segmentation – identifying the gestures in a sequence of movements; and second, representation – representing these gestures in an adequate space. The difficulty is in identifying the relevant segments in the flow of the initial data (music, movement) and applying descriptors of expressivity to them. In an analysis of an interpreter's movements, for example, a fragment of his/her performance will be segmented into a series of gestures, each defined according to the various speeds and direction. Each gesture is then described by a vector containing the values taken by each of the descriptors of expressivity.

All of these measures allow us to represent a gesture in a multidimensional space containing qualitative and expressive descriptors of movement (e.g. fluid versus rigid). The sequence of gestures taking place in time and space are therefore identified as trajectories in spaces of continuous dimensions. The trajectories are then analyzed to find similarities among them in order to group them into *clusters*. To get a good overview of the data gathered, factorial analysis techniques are applied (analysis according to main components, analysis according to independent components, etc). These techniques allow us to reduce a large volume of data into a smaller, reduced set, with what might be considered irrelevant or insignificant data removed [CAN 04, POL 04].

9.5.1.4. *Level 4*

Level 4 (categories and concepts) describes the process of extracting high-level expressive information, i.e. extracting the meaning of what is expressed. In principle, this level can be considered as a network linking characteristics of gestures to concepts or categories. Musical performance, for example, can be analyzed in terms of conveyed emotional intentions. Machine learning techniques are frequently used to classify sequences of movements into emotional categories [CAM 03]. These techniques include various procedures such as statistics (e.g. multiple regression, generalized linear techniques), fluidic logic or probabilistic reasoning systems (Bayesian networks), neural networks (classic back-propagation networks, Kohonen networks), *support vector machines* and decision trees, etc.

9.5.1.5. *Overview*

These four levels of analysis for multimodal content of musical expressivity, specifically for expressive gestures, suggest a clear path that has demonstrated its efficacy in research. Such an approach can be successfully followed when designing public applications. The precision brought by this type of study has highlighted, scientifically and methodically, essential aspects of musical expressivity that benefit from the creation of new musical expression interfaces.

9.6. Multimodal emotional synthesis in a musical context

Accounting for multisensory context in expressivity is crucial when creating new musical interfaces. For several years now the music informatics community has sought to identify new expressive interaction paradigms using computers (for an overview, refer to the NIME or *New Interfaces for Musical Expression* Conference, www.nime.org). This frequently involves referring to musical instruments. Virtual instruments, hyper instruments and other enhanced instruments represent a successful area of research [BEL 06]. As early as 1965 in *Variation V*, John Cage and Merce Cunningham used Theremin to musically transcribe dancers' movements. The use of sensors was developed in the 1970s and 80s to create invisible instruments that, in reality, were active spaces in which the user's every movement is captured, analyzed and used to generate and control sound samples. These developments led to the creation of *multimodal environments* that favor continuous and multimodal interaction with the user. This renewed form of interaction can be considered as an extension of possible interactions developed with virtual instruments [CAM 05a, CAM 99].

Castellano *et al*. have developed a multimodal environment based on expressive and emotional communication, which allows complete freedom of movement [CAS 07]. The system, designed using the EyesWeb XMI software platform, can analyze the user's movements by focusing on the combined variations of two bodily characteristics recognized as expressive, which are the amount of movement and the subject's spatial occupation [CAM 03]. The values of the movement's expressive characteristics are *mapped* in real time using acoustic musical expressivity parameters using a specialized library from the Pure Data software (www.puredata.org). In order to strengthen emotional interaction, the system projects the subject's silhouette, which it colors according to their emotion. The integration of different multimodal aspects of musical expressivity therefore aims to reinforce the subject's "emotional engagement".

Based on this principle of multimodal engagement, Bernhardt and Robinson have added machine learning techniques to this multimodal environment [BER 08]. During the systems learning phase, a piece of music is first selected for the type of dominant emotion that it conveys. Then the user is asked to move freely according to how the music makes him/her feel. Following this learning process, the system can recognize emotions on the basis of the user's natural (i.e. unconstrained) movements. This approach has the advantage that the movement's rules do not need to be well defined in the interface in advance and the users do not need to learn gestures prescribed by the researcher or the software (see Figure 9.5).

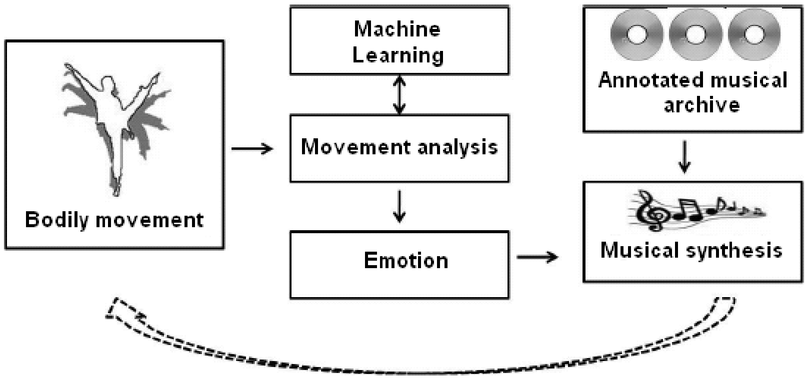


Figure 9.5. *Diagram of a musical–emotional interface. Musical synthesis is controlled by the body's expressive movements [BER 08]*

Less directly, multisensory factors of musical expressivity have favored interaction between humans and virtual creatures, such as avatars. Following the example of DiPaola *et al.* [DIP 04], Mancini *et al*., have created an application that visualizes expressive qualities in a musical performance using a virtual head (e.g. the avatar, Greta) [MAN 07]. Four acoustic parameters have been selected to identify the emotional intention of the musical interpretation, e.g. sound intensity, tempo, articulation (*legato-staccato*) [FRI 05]. Each of these parameters is linked to one of the parameters that conditions expression by the head: rotation and acceleration, duration and frequency of head movement, the fluidity of movement. This visualization of musical performance in real time using an expressive face can lead to new forms of musical applications. For example, a portable audio player with a video screen provides the user with responsive visual feedback when he/she searches the database. A musician who needs to practice can be accompanied by a virtual teacher that adapts its response and expression according to the user's musical intentions.

The majority of these musical expressivity interfaces focus on the individual user. However, a musical experience can be split into both the audience and the musician, since we can play to practice, in groups, orchestras, etc. This collective aspect of music remains relatively ignored in the design of emotional music interfaces.

9.7. The social active listening paradigm: the collective aspect of emotion

9.7.1. *Example: Mappe per Affetti Erranti*

Camurri *et al.* have created an installation called *Mappe per Affetti Erranti* (literally, "maps for wandering effects") that has shown new possibilities for "emotional" interaction between users and pre-recorded musical content. The musical material chosen was two Baroque choral pieces – Percell's *The Fairy Queen*, and *King Lear*. These pieces were interpreted by four professional singers, depending on the pieces' possible emotional explorations, which will then be provided to the users. Each singer was recorded separately in the studio and this was then reused separately during installation.

The place of installation (a stage in this case) was "split" into four spaces corresponding to the four voices in the polyphonic music. The presence of a user in one of these spaces triggered the reproduction of one of the four voices. By crossing the whole stage, the user encountered different spaces and was able to listen to each of the voices. If the user stayed in one space, he/she heard the voice associated with this mood separately. If he/she remained immobile, the music slowly ended.

The user was able to alter the voice he or she was listening to in different ways. For example, he/she was able to change the sound parameters by opening his/her arms, increasing the density of the voices (i.e. listening to two or more voices in unison). If the user moved towards the back of the stage, the level of echo increased while if he/she stayed at the front of the stage, the voice became harder. This emotional space was itself divided into four sections corresponding to the four pre-recorded expressive interpretations of the piece of music.

The users were therefore able to experience the music in two ways. First, by crossing the space, he/she could explore the music's polyphonic structure and by crossing the emotional space were able to discover several different possible interpretations. One user alone, however, could only listen to and alter one voice at a time. The user could not listen to the polyphony as a whole and the voices composing it. Only a group of users was able to have a complete experience of the *Mappe per Affetti Erranti*. The piece of music therefore only had its polyphony extended if there were at least the same number of users as there were voices playing in the installation. Furthermore, since each user controled the voice's interpretation based on his/ her movements, the whole piece was executed with the same expressive intention only where all the users moved with the same expressivity. Consequently, if users moved with conflicting expressive intentions, the resulting sound was incoherent and chaotic. In contrast, if users moved with the same type of movement and worked together, all the voices came together coherently, creating a harmonious and expressive interpretation of the piece.

This type of interface revives the traditional approach to musical expressivity. Research into computational models of emotion (*affective computing*) and models of expressive gestures normally focus on individual users. However, in this kind of interface the system does not recognize expressivity based only on the type of movement made by the users; the type of cooperation and interaction between different actors is also considered. The challenge for developing such collaborative interfaces is in estimating a group's expressive intention, not only as an average of expressive intentions conveyed by each of these elements but as the result of more complex in-group dynamics of interaction (e.g. training, synchronization, etc.) [CLA 04, DEM 08]. Such dynamics of social interaction and their relation to emotional expression remain largely underexplored in the domain of musical expression interfaces.

9.8. Conclusion and perspectives

It seems clear that music and emotions are linked. The nature of this relationship has therefore been the subject of a growing number of scientific studies. The range of disciplines examining this relationship illustrates the complexity of the processes involved, for example neuroscience, experimental psychology, physiology, clinical studies, etc. Analysis of the relation between the musical score characteristics and audience interpretations and reactions to them has allowed us to identify acoustic properties governing how emotions are expressed and conveyed. Other sensory modalities (visual information and bodily movements) have also been examined with regard to their role in the emotional communication process.

Research into the emotions evoked by music has led to prototypes and new multimodal interfaces for artistic expression and creativity. Since these interfaces are interactive and "sensitive" to emotions, they have been the subject of major attention because they can enable music to echo users' emotions and stimulate players' enjoyment. The possibility to freely interact with musical content through unconstrained and ever-evolving movement and intention has introduced a new way of experiencing and sharing music. The music informatics community has therefore reached a new level of maturity that allows it to be seen as an independent area of research whose scientific productions and technological achievements will feed our understanding of the emotional processes evoked by music.

9.9. Bibliography

[ALP 05] ALPERT M., ALPERT J., MALTZ E., "Purchase occasion influence on the role of music in advertising", *Journal of Business Research*, vol. 58, no. 3, pp. 369-376, Elsevier, 2005.

[ASM 85] ASMUS E., "The development of a multidimensional instrument for the measurement of affective responses to music", *Psychology of Music*, vol. 13, no. 1, 1985.

[BER 08] BERNHARDT D., ROBINSON P., "Interactive control of music using emotional body expressions", *Proceedings of the ACM Conference on Computer-Human Interaction CHI'08*, ACM Press, New York, 2008.

[BEV 06] BEVILACQUA F., RASAMIMANANA N., FLÉTY E., LEMOUTON S., BASCHET F., "The augmented violin project: research, composition and performance report",*Proceedings of the 2006 Conference on New Interfaces for musical expression*, pp. 402-406, 2006.

[BIG 06] BIGAND E., *Musiciens et Mon-musiciens Perçoivent-ils la Musique Différemment*?, De Boeck University, 2006.

[BLA 07] BLAKE R., SHIFFRAR M., "Perception of human motion", *Annual Review of Psychology*, vol. 58, 2007.

[BLO 01] BLOOD A., ZATORRE R., "Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion", *Proceedings of the National Academy of Sciences*, vol. 98, no. 20, 2001.

[BOO 98] BOONE R., CUNNINGHAM J., "Children's decoding of emotion in expressive body movement: The development of cue attunement", in: *Developmental Psychology*, vol. 34, pp. 1007-1016, APA American Psychological Association, 1998.

[BRE 00] BRESIN R., FRIBERG A., "Emotional coloring of computer-controlled music performances", *Computer Music Journal*, vol. 24, no.4, pp. 44-63, 2000.

[BRU 56] BRUNSWIK E., *Perception and the Representative Design of Psychological Experiments*, University of California Press, Berkeley, 1956.

[BRU 90] BRUNER G., "Music, mood and marketing", *Journal of Marketing*, vol. 54, no. 4, pp. 94-104, 1990.

[BUD 85] BUDD M., *Music and the Emotions: The Philosophical Theories*, Routledge Kegan & Paul, London, 1985.

[CAM 99] CAMURRI A., RICCHETTI M., TROCCA R., "EyesWeb-toward gesture and affect recognition in dance/music interactive systems", *Multimedia Computing and Systems, 1999, IEEE International Conference on*, vol. 1, 1999.

[CAM 03] CAMURRI A., LAGERLÖF I., VOLPE G., "Recognizing emotion from dance movement: comparison of spectator recognition and automated techniques", *International Journal of Human-Computer Studies*, vol. 59, pp. 213-225, July 2003.

[CAM 05a] CAMURRI A., DE POLI G., FRIBERG A., LEMAN M., VOLPE G., "The MEGA project: analysis and synthesis of multisensory expressive gesture in performing art applications", *Journal of New Music Research*, vol. 34, no. 1, pp. 5-21, Routledge, 2005.

[CAM 05b] CAMURRI A., VOLPE G., DE POLI G., LEMAN M., "Communicating expressiveness and affect in multimodal interactive systems", *IEEE Multimedia*, pp. 43-53, IEEE Computer Society, 2005.

[CAN 04] CANAZZA S., DE POLI G., DRIOLI C., RODA A., VIDOLIN A., "Modeling and control of expressiveness in music performance", *Proceedings of the IEEE*, vol. 92, no. 4, pp. 686-701, 2004.

[CAS 07] CASTELLANO G., BRESIN R., CAMURRI A., VOLPE G.,"Expressive control of music and visual media by full-body movement", *Proceedings of the 7th International Conference on New Interfaces for Musical Expression*, pp. 390-391, ACM Press, New York, 2007.

[CHA 08] CHAPADOS C., LEVITIN D., "Cross-modal interactions in the experience of musical performances: Physiological correlates", *Cognition*, Elsevier, 2008.

[CLA 04] CLAYTON M., SAGER R., WILL U., "In time with the music: the concept of entrainment and its significance for ethnomusicology", *ESEM Counterpoint*, vol. 1, pp. 1-82, 2004.

[CLY 77] CLYNES M., *Sentics: The Touch of Emotions*, Anchor Press, New York, 1977.

[COO 59] COOKE D., *The Language of Music*, Oxford University Press, London, 1959.

[COW 01] COWIE R., DOUGLAS-COWIE E., TSAPATSOULIS N., VOTSIS G., KOLLIASS., FELLENZ W., TAYLOR J., "Emotion recognition in human-computer interaction", *Signal Processing Magazine*, vol. 18, no. 1, pp. 32-80, 2001.

[DAH 07] DAHL S., FRIBERG A., "Visual perception of expressiveness in musicians' body movements", *Music Perception*, vol. 24, no.5, pp. 433-454, 2007.

[DAV 94a] DAVIDSON J., "What type of information is conveyed in the body movements of solo musician performers", *Journal of Human Movement Studies*, vol. 6, pp. 279-301, 1994.

[DAV 03] DAVIDSON R., SCHERER K., GOLDSMITH H., *Handbook of Affective Sciences*, Oxford University Press, New York, 2003.

[DAV 94b] DAVIS M., "Folk music psychology", *The Psychologist*, vol. 7, no.12, 1994.

[DEM 08] DEMEY M., LEMAN M., BOSSUYT F., VANFLETEREN J., "The musical synchrotron: using wireless motion sensors to study how social interaction affects synchronization with musical tempo", *Proceedings of the 8th International Conference on New Interfaces for Musical Expression*, 2008.

[DIP 04] DIPAOLA S., ARYA A., "Affective communication remapping in MusicFace system", *Proceedings of the 10th European Conference on Electronic Imaging and the Visual Arts (EVA '04)*, 2004.

[EKM 83] EKMAN P., FRIESEN W., EMFACS-7: Emotional Facial Action Coding System, Unpublished manual, University of California, San Francisco, 1983.

[FRI 91] FRIBERG A., FRYDÉN L., BODIN L., SUNDBERG J., "Performance rules for computer-controlled contemporary keyboard music", *Computer Music Journal*, vol. 15, no. 2, pp. 49-55, 1991.

[FRI 95] FRIBERG A., A quantitative rule system for musical performance, Unpublished Doctoral dissertation, Royal Institute of Technology, Sweden, 1995.

[FRI 99] FRIBERG A., SUNDBERG J., "Does music performance allude to locomotion? A model of final ritardandi derived from measurements of stopping runners", *The Journal of the Acoustical Society of America*, vol. 105, 1999.

[FRI 00] FRIBERG A., COLOMBO V., FRYDEN L., SUNDBERG J., "Generating musical performances with Director Musices", *Computer Music Journal*, vol. 24, no. 3, pp. 23-29, 2000.

[FRI 05] FRIBERG A., SCHOONDERWALDT E., JUSLIN P., "CUEX: An algorithm for extracting expressive tone variables from audio recordings", *Acoustica United with Acta Acoustica*, vol. 93, pp. 411-420, 2005.

[FRI 06a] FRIBERG A., "pDM: An expressive sequencer with real-time control of the KTH music-performance rules", *Computer Music Journal*, vol. 30, no. 1, pp. 37-48, 2006.

[FRI 06b] FRIBERG A., BRESIN R., SUNDBERG J., "Overview of the KTH rule system for musical performance", *Advances in Cognitive Psychology*, vol. 2, no. 2-3, pp. 145-161, 2006.

[GAB 85] GABRIELSSON A., "Interplay between analysis and synthesis in studies of music performance and music experience", *Music Perception*, vol. 3, no. 1, pp. 59-86, 1985.

[GAB 87] GABRIELSSON A., "Once again: the theme from Mozart piano sonata in A Major (K 331)", *Action and Perception in Rhythm and Music*, vol. 55, pp. 81-103, 1987.

[GAB 96] GABRIELSSON A., JUSLIN P., "Emotional expression in music performance: between the performer's intention and the listener's experience", *Psychology of Music*, vol. 24, no. 1, 1996.

[GAB 01a] GABRIELSSON A., "Emotions in Strong Experiences with Music", *Music and Emotion: Theory and Research*, Oxford University Press, Oxford, 2001.

[GAB 01b] GABRIELSSON A., LINDSTROM E., "The influence of musical structure on emotional expression", *Music and Emotion: Theory and Research*, pp. 223-248, Oxford University Press, Oxford, 2001.

[GAB 03] GABRIELSSON A., JUSLIN P., "Emotional expression in music", *Handbook of Affective Sciences*, GOLDSMITH H.H., DAVIDSON R.J., SCHERER K.R. (eds), pp. 503-534, Oxford University Press, New York, 2003.

[GLO 08] GLOWINSKI D., CAMURRI A., VOLPE G., NOERA C., COWIE R., MCMAHON E., KNAPP B., JAIMOVICH J., "Using induction and multimodal assessment to understand the role of emotion in musical performance", *Proceedings of the 2008 Conference on Emotion in Human-Computer Interaction*, Liverpool John Moores University, 2008.

[GOL 04] GOLD C., VORACEK M., WIGRAM T., "Effects of music therapy for children and adolescents with psychopathology: a meta-analysis", *Journal of Child Psychology and Psychiatry*, vol. 45, no. 6, pp. 1054-1063, 2004.

[HAS 97] HASHIMOTO S., "KANSEI as the third target of information processing and related topics in Japan", *Proceedings of the International Workshop on KANSEI: The Technology of Emotion*, AIMI (Italian Computer Music Association) and University of Genoa, 1997.

[HEN 07] HENBING L., "A gesture-based typology of sliding-tones in Guqinmusic", *Journal of New Music Research*, vol. 36, no. 2, pp. 61-82, 2007.

[IZA 77] IZARD C., *Human Emotions*, Plenum Press, New York, 1977.

[JUS 97a] JUSLIN P., "Perceived emotional expression in synthesized performances of a short melody: Capturing the listener's judgment policy", *Musicae Scientiae*, vol. 1, no. 2, pp. 225- 256, 1997.

[JUS 97b] JUSLIN P., "Emotional communication in music performance: A functionalist perspective and some data", *Music Perception*, vol. 14, p. 383-418, 1997.

[JUS 00] JUSLIN P., "Cue utilization in communication of emotion in music performance: relating performance to perception", *Journal of Experimental Psychology: Human Perception, Performance*, vol. 26, no. 6, pp. 1797-813, 2000.

[JUS 01] JUSLIN P., SLOBODA J., *Music and Emotion: Theory and Research*, Oxford University Press, Oxford, 2001.

[JUS 02] JUSLIN P., ZENTNER M., "Current trends in the study of music and emotion: Overture", *Musicae Scientiae*, vol. 6, no. 1; SPI, pp. 3-22, 2002.

[JUS 03a] JUSLIN P., LAUKKA P., "Communication of emotions in vocal expression and music performance: Different channels, same code?" *Psychological Bulletin*, vol. 129, no.5, pp. 770-814, 2003.

[JUS 03b] JUSLIN P., LAUKKA P., "Emotional expression in speech and music evidence of cross-modal similarities", *Annals of the New York Academy of Sciences*, vol. 1000, no.1, pp. 279-282, 2003.

[KAL 05] KALLINEN K., "Emotional ratings of music excerpts in the western art music repertoire and their self-organization in the Kohonen neural network", *Psychology of Music*, vol. 33, no.4, 2005.

[KAL 06] KALLINEN K., RAVAJA N., "Emotion perceived and emotion felt: same and different", *Musicae Scientiae*, vol. 10, no. 2, 2006.

[KRU 96] KRUMHANSL C., "A perceptual analysis of Mozart's piano sonata K. 282: Segmentation, tension, and musical ideas", *Music Perception*, vol. 13, pp. 401-432, 1996.

[KRU 97a] KRUMHANSL C., "An exploratory study of musical emotions and psychophysiology", *Canadian Journal of Experimental Psychology*, vol. 51, no.4, pp. 336–353, 1997.

[KRU 97b] KRUMHANSL C., SCHENCK D., "Can dance reflect the structural and expressive qualities of music?: A perceptual experiment on Balanchine's choreography of Mozart's Divertimento n. 15", *Musicae Scientiae*, vol. 1, pp. 63-86, 1997.

[KUR 90] KURTENBACH G., HULTEEN E., "Gestures in human-computer communication", *The Art of Human-Computer Interface Design*, pp. 309-317, 1990.

[LAB 47] LABAN R., LAWRENCE F., *Effort*, Macdonald and Evans, London, 1947.

[LAB 63] LABAN R., *Modern Educational Dance*, Macdonald and Evans, London, 1963.

[LEM 07] LEMAN M., *Embodied Music Cognition and Mediation Technology*, M Press, Cambridge, 2007.

[LIV 06] LIVINGSTONE S., THOMPSON W., "Multimodal affective interaction", *Music Perception*, vol. 24, no.1, p. 89-94, 2006.

[LIV 07] LIVINGSTONE S., "Controlling musical emotionality: an affective computational architecture for influencing musical emotions", *Digital Creativity*, vol. 18, no. 1, pp. 43-53, 2007.

[MAD 93] MADSEN C., FREDRICKSON W., "The experience of musical tension: A replication of Nielsen's research using the continuous response digital interface", *Journal of Music Therapy*, vol. 30, pp. 46-46, 1993.

[MAN 07] MANCINI M., BRESIN R., PELACHAUD C., "A virtual head driven by music expressivity", *IEEE Transactions on Audio Speech and Language Processing*, vol. 15, no. 6, 2007.

[MEN 05] MENON V., LEVITIN D., "The rewards of music listening: Response and physiological connectivity of the mesolimbic system", *Neuroimage*, vol. 28, no. 1, pp.175-184, 2005.

[MEY 56] MEYER L., *Emotion and Meaning in Music*, University of Chicago Press, Chicago-London, 1956.

[MEY 67] MEYER L., *Music, the Arts and Ideas*, University of Chicago Press, Chicago London, 1967.

[NIE 83] NIELSEN F., *Oplevelseaf Musikalsk Spænding*, Akademisk Forlag, 1983.

[PIC 97] PICARD R., *Affective Computing*, MIT Press, Cambridge, 1997.

[POL 04] POLI G., "Methodologies for expressiveness modelling of and for music performance", *Journal of New Music Research*, vol. 33, no.3, pp. 189-202, 2004.

[RAM 07] RAMSAY J. O., SILVERMAN B.W., "Functional data analysis", *Biometrics*, vol. 63, no.1, pp. 300-301,2007.

[REN 03] RENTFROW P., GOSLING S., "The Do Re Mi s of everyday life: the structure and personality correlates of music preferences", *Journal of Personality and Social Psychology*, vol. 84, no. 6, pp. 1236-1256, 2003.

[REP 92] REPP B., "Diversity and commonality in music performance: An analysis of timing microstructure in Schumann Traumerei", *The Journal of the Acoustical Society of America*, vol. 92, 1992.

[RIG 64] RIGG M., "The mood effects of music: A comparison of data from four investigators", *The Journal of Psychology*, vol. 58, pp. 427-438, 1964.

[ROB 05] ROBINSON J., "Deeper than reason: Emotion and its role in literature", *Music, and Art*, Oxford University Press, Oxford, 2005.

[RUS 80] RUSSELL J., "A circumplex model of affect", *Journal of Personality and Social Psychology*, vol. 39, no.6, pp. 1161-1178, 1980.

[SCH 85] SCHERER K., WALLBOTT H., "Analysis of nonverbal behavior", *Handbook of discourse analysis*, vol. 2, pp. 199-230, 1985.

[SCH 01a] SCHERER K., ZENTNER M., "Emotional effects of music: production rules", *Music and Emotion: Theory and Research*, JUSLIN P., SLOBODA J. (eds), pp. 361-392, Oxford University Press, 2001.

[SCH 99] SCHUBERT E., Measurement and time series analysis of emotion in music, PhD thesis, University of New South Wales, 1999.

[SCH 01b] SCHUBERT E., "Continuous measurement of self-report emotional response to music", *Music and Emotion: Theory and Research*, p. 393-414, Oxford University Press, 2001.

[SCH 07] SCHUBERT E., "The influence of emotion, locus of emotion and familiarity upon preference in music", *Psychology of Music*, vol. 35, no.3, 2007.

[SLO 01] SLOBODA J., O NEILL S., IVALDI A., "Functions of music in everyday life: An exploratory study using the experience sampling method", *Musicae Scientiae*, vol. 5, no. 1, pp. 9-32, 2001.

[SUN 82] SUNDBERG J., FRYDÉN L., ASKENFELT A., "What tells you the player is musical? A study of music performance by means of analysis-by-synthesis", *KTH Speech Transmission Laboratory Quarterly Progress and Status Report,* vol. 23, pp. 135-148, 1982.

[THO 05] THOMPSON W., GRAHAM P., RUSSO F., "Seeing music performance: Visual influences on perception and experience", *Semiotica*, vol. 2005, no. 156, p. 203-227, 2005.

[TIM 06] TIMMERS R., MAROLT M., CAMURRI A., VOLPE G., "Listeners'emotional engagement with performances of a Scriabin etude: an explorative case study", *Psychology of Music*, vol. 34, no. 4, 2006.

[TRA 08] TRAINOR L., "Science and music, the neural roots of music", *Nature*, vol. 453, no. 7195, 2008.

[VAS 02] VASTFJALL D., "A review of the musical mood induction procedure", *Musicae Scientiae*, vol. 6, no.1; SPI, pp. 173-212, 2002.

[VIN 04] VINES B., WANDERLEY M., KRUMHANSL C., NUZZO R., LEVITIN D., "Performance gestures of musicians: what structural and emotional information do they convey?", *Lectures Notes in Computer Science*, pp. 468-478, Springer, 2004.

[VIN 05] VINES B., NUZZO R., LEVITIN D., "Analyzing temporal dynamics in music", *Music Perception*, vol. 23, no. 2, pp. 137-152, 2005.

[VIN 06] VINES B., KRUMHANSL C., WANDERLEY M., LEVITIN D., "Cross-modal interactions in the perception of musical performance", *Cognition*, vol. 101, no.1, pp. 80-113, 2006.

[WAL 86] WALLBOTT H., SCHERER K., "Cues and channels in emotion recognition", *Journal of Personality and Social Psychology*, vol. 51, no.4, pp. 690-699, 1986.

[WAL 98] WALLBOTT H., "Bodily expression of emotion", *European Journal of Social Psychology*, vol. 28, pp. 879-896, 1998.

[WAN 05] WANDERLEY M., VINES B., MIDDLETON N., MCKAY C., HATCH W., "The musical significance of clarinetists' Ancillary Gestures: An Exploration of the Field", *Journal of New Music Research*, vol. 34, no. 1, pp. 97-113, 2005.

[WES 96] WESTERMANN R., STAHL G., HESSE F., "Relative effectiveness and validity of mood induction procedures: analysis", *European Journal of Social Psychology*, vol. 26, pp. 557-580, 1996.

[ZEN 96] ZENTNER M., KAGAN J., "Perception of music by infants", *Nature*, vol. 383, no.6595, pp. 29-29, 1996.

[ZEN 98] ZENTNER M., KAGAN J., "Infants' perception of consonance and dissonance in music", *Infant Behavior and Development*, vol. 21, no. 3, pp. 483-492, 1998.

[ZEN 08] ZENTNER M., GRANDJEAN D., SCHERER K., "Emotions evoked by the sound of music: characterization, classification, and measurement", *Emotion*, vol. 8, no. 4, pp. 494-521, 2008.

Chapter 10

Literary Feelings in Interactive Fiction

10.1. Introduction: emotions and feelings

The spectacular development of emotional interfaces since the publication of Picard's *Affective Computing* [PIC 97], has taken place from a largely communicational perspective. In most cases, this allows the machine to recognize (or in the case of synthetic actors even generate) basic or universal emotions that had originally been described within the framework of interpersonal communication. Behind the idea that considering emotions makes interaction more natural, acceptable or even more efficient, is the almost ethological conception of communication, not too distant from the kind of cognitivism reductionism that received so much support 1980s (see [RAS 91] for an overview). The most obvious indication of this is the desire to find a limited ontology of basic emotions that are universal. Ekman's primitive emotions have fulfilled this role well, although their limitations are also beginning to be recognized. As applications have become more complex and have involved richer semantic context (which might also be termed cultural), wanting to characterize interactions in terms of sadness, fear or disgust inevitably becomes flawed. The emotions evoked by a film, animation or video game are significantly more complex and not necessarily always well defined. The psychology of entertainment [BRY 06] has attempted to describe and analyze this.

Chapter written by Marc CAVAZZA and David PIZZI.

Emotion-Oriented Systems Edited by Catherine Pelachaud

Such issues raise the ongoing debate between the "cognitive" and the "cultural". The emotions found in the deepest structures of the brain [BEL 07] explain our richest cultural experiences, just as Schank and Abelson's 12 primitive emotions [SCH 77] explain the semantics of language. This chapter attempts to provide an alternate view, examining several less orthodox approaches to emotions focusing on linguistic content rather than psychological data [COP 09]. This kind of methodology analyzes textual linguistics [RAS 89], specifically the relationship between linguistics and literature. This involves examination of the complexity of cultural objects, as opposed to research into universal mechanisms. We also attempt to show that practical applications (interactive media) using synthetic actors [NIE 09] can be based on types of emotions that differ hugely from communicational emotions, which will form sections 10.2 and 10.3.

To better appreciate the complexity of emotions in cultural productions, we will initially put aside the dialectic of emotions to focus less on the emotions attributable to the audience and more on those that are found at the heart of a cultural object as a narrative. Despite its relationship with myth, nothing has established that narrative is a vehicle for previously "encoded" emotions that reading (used in the wider sense here, including interpretative) aims to reveal. Emotions therefore do not have a level of status that raises them above the level of the abstract. Even if, according to Agamben "taste is what does not disgust us", this does not mean that disgust (a universal emotion if there ever was one) has anything to do with how we interpret cultural objects or even their digital representations. For example, we can feel disgust when looking at an image or reading a description, but this is a reduced interpretation of a narrative based on referential impressions (direct evocation of the real world or a reference world). Genre is also a factor, although it is true that some (writers and film makers) overtly seek to evoke the primary emotions. Without establishing a hierarchy of values, we should simply say that these genres often deliberately reduce their interpretation to the elicitation of such primary emotions which, at least for the best of them, does not fully express their meaning.

In this chapter we refer to "feelings", whether attributable to a character or not, although this excludes feelings experienced by the reader or the affective content of cultural objects such as narratives. Specifically, we examine the feelings described or evoked by narratives, excluding those of the audience (relating to psychology and therefore not literature) that are not accessible using a linguistic methodology. Furthermore, if fictional characters can feel and express primary emotions, there is nothing to say that a kind of empathy, ignoring any other form of interpretation, sufficient to explain readers' emotions. The role of the audience's empathy towards characters has in this way been questioned by recent research in the psychology of entertainment [SMI 03].

To fully appreciate the complexity of this issue, we need to examine the true nature of emotions in narratives and their linguistic features. To do so this, we focus on an as-yet unparalleled study [RAS 95a] that attempts to create an inventory of emotions in novels, specifically those written in French.

10.2. French novels and the representation of feelings

Gustav Flaubert's famous *Madame Bovary* [FLA 56] is a novel based around the concept of *boredom* – a complex emotion, in this case expressing kind of a middle-class boredom that differs from general boredom relating to the immediacy of existence and whose description is more commonly attributed to Martin Heidegger. As a result, we can therefore say that boredom is not a basic or universal emotion but it represents a culturally indispensable feeling. Compared to disgust, which is a communicative (or universal) emotion, it allows us to approach the theme of this chapter, namely the distinction between nature and culture.

To return to *Madame Bovary*, bourgeois boredom is one brought about by conditions, and not a boredom of class (e.g. Flaubert versus Marx) and it is therefore far from the kind of boredom described by Baudelaire and Huysmans. Nevertheless, boredom is rarely lexicalized in novels as an emotion in itself, as Rastier has demonstrated with four different orthonyms. Boredom should not be interpreted as the novel's overall theme but as a semantic form constructed through *interpretation*[1] that allows us to establish a link between different components of the narrative. For example, Emma Bovary does not simply become bored, rather she experiences a number of complex emotions among which boredom is the framework of survival and the conditions of evolution can be explained. Emotions in the novel are not only far more complex than those described by Ekman, but their analysis reveals a process opposite to that used in current attempts at emotional language processing, which has sought to collate all vocabulary corresponding closely to "universal" emotions in highly emotional contexts (e.g. the processes in Wordnet-Affect) [VAL 05].

This highly reductive approach is largely inapplicable to literary analysis, specifically cultural issues. Indeed, the identification of a feeling from the interpretive process is required at least for lexical analysis in context. These feelings are therefore reconstructed using texts and not the contrary (since they are linked to both content and expression). The nuances of feelings expressed in the novel increases the ability of lexical analysis to highlight more specific semantic classes. However, there are other indicators that confirm the nature and subtleties of these

1 I.e. boredom is not so much an element of immediate representation as a means of organizing and composing a representation.

identified emotions. First, we can say that a given feeling is more easily detectable using its semantic correlates than its direct occurrences [ERL 95]. In the case of boredom, for example, we can say that this phenomenon found in several novels (including *Madame Bovary* and *A Rebours* by Joris-Karl Huysmans [HUY 77]), and that this is manifested across several correlates. We are therefore faced with two basic questions:

– Are correlates simply variations of expressions of canonical (or primitive) emotions?

– Or does their expression correspond to a finer distinction that is more indicative of the real level of description of the feelings expressed in the novel?

This second question encapsulates an issue that has continued to cause debate in social sciences since their collaboration with cognitive sciences. Can all descriptions, including literary descriptions, be reduced to a "cognitive" psychology of the characters or do characters' psychologies in novels allow us to identify literary and aesthetic categories of emotion that are the basis of a real ontology of literary sentiments? If we return to the example of boredom, we could certainly describe it as a cognitive pathology of rationality in terms of lack of motivation or difficulties with decision-making.

A comparison of the boredom expressed by Huysmans and Flaubert cannot be achieved using purely cognitive data (which would reduce the nuances determining the nature of the novel), however, because their differentiation is due to a social context that is reproduced by literary expression. Literary feelings therefore cannot be reduced to primitive cognitive emotions as, although important to identify underneath the expression, the richness of the social content they convey must refine their description. This is what allows us to distinguish the boredom expressed in *Des Esseintes* from that of Emma Bovary.

This is not a new debate; it has existed since the birth of computer science, which, following the lead of artificial intelligence, has sought to represent more complex data than that of the immediate physical world, specifically at a linguistic level.

A recent study of Flaubert's original drafts of *Madame Bovary* [LEC 95] allows us to see the author's exact intentions, which demonstrate the characters' intentions, motivations (i.e. their plans) and emotional states. These are described with the kind of finesse required by their novelistic expression, which are emotions that cannot be simply reduced to a psychological description of a basic emotion. In other words, the subtleties of characters' feelings should not be constructed using universal basic emotions because this characteristic approach towards cognitivism dating from the 1980s has failed, even if it has been instructive.

10.3. *Madame Bovary*: plot and scenes

There are a number of studies examining Flaubert's drafts of *Madame Bovary*, ranging from the genetics of the texts (see [RAD 89] for further details) to the works themselves. Leclerc, for example, has composed an original compilation of Flaubert's annotated notes describing his plans for *Madame Bovary* [LEC 95]. These drafts comprise plans including a detailed map of Yonville, the small market town in which the novel is set, a timeline of the novel's events (which is not to go against innocent users of the kind of logic pioneered by Allen [ALL 83] in artificial intelligence) and different drafts of scenes, providing the work with its title, *Plans et Scénarios*. The most interesting feature of these drafts, however, lies in how the novel's characters are psychologically described. Such a description can be justifiably interpreted as emotions. This allows us to examine how the characters' descriptions differ from the previously cited generic emotional categories. We propose that literary feelings should be analyzed at three differnt levels of descriptions which reflect the various classes used in textual semantics, namely dimensions[2], domains and taxemes[3] [RAS 94].

Dimensions correspond to abstract semantic categories where language overlaps with larger categories describing the world (/*animated*/, /*inanime*/)[4] or categories of expression (subjective dimensions such as /*pejorative*/). In this context, traditional categories with a higher level of genericity correspond to emotions that are described more generally, such as anger or fear. The range of feelings expressed in the novel originates from meso-semantic[5] research into emotions. We will now focus on a more specific description of sentiments, such as honor and boredom, avoiding the dilemma of distinguishing Emma Bovary's boredom from that expressed in *Des Esseintes*.

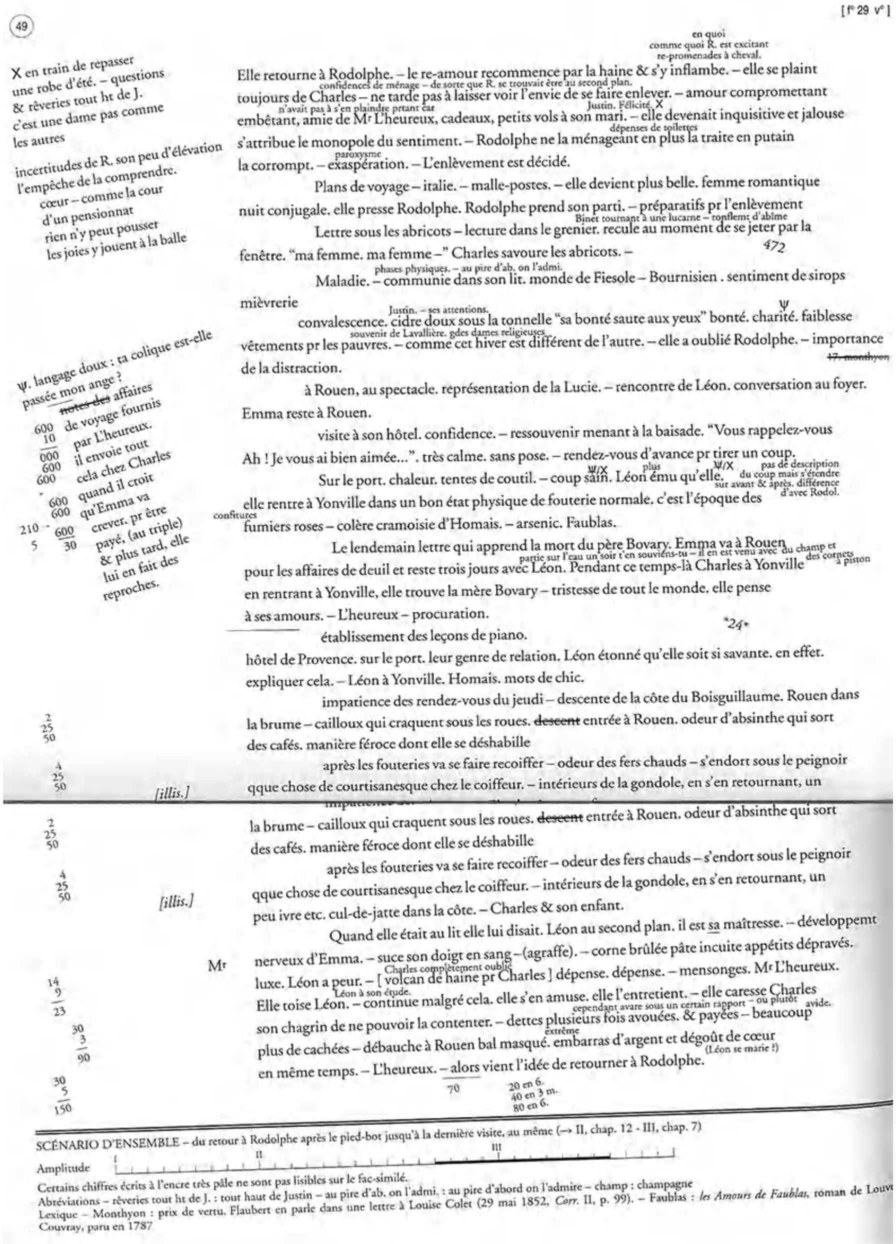
---

2 Note that the term (a semantic) "dimension" is not related to that used in dimensional approaches to emotions, which we will examine later.

3 Dimensions, domains and taxemes refer to the classes of semantic equivalence of increasing specificity that form the basis of descriptions in textual semantics.

4 We will use the normal notation conventions for semantic features [RAS 87].

5 Meso-semantics [RAS 87] relates to an intermediary level of description including domains and classes of meaning corresponding to a cultural aspect of human activity (domains are therefore often referred to in dictionary definitions).



49 [f° 29 v°]

X en train de repasser une robe d'été. – questions & rêveries tout ht de J. c'est une dame pas comme les autres

incertitudes de R. son peu d'élévation l'empêche de la comprendre. cœur – comme la cour d'un pensionnat rien n'y peut pousser les joies y jouent à la balle

en quoi
comme quoi R. est excitant
re-promenades à cheval.

Elle retourne à Rodolphe. – le re-amour recommence par la haine & s'y inflambe. – elle se plaint
confidences de ménage – de sorte que R. se trouvait être au second plan.
toujours de Charles – ne tarde pas à laisser voir l'envie de se faire enlever. – amour compromettant
n'avait pas à s'en plaindre prtant car — Justin. Félicité. X
embêtant, amie de Mr L'heureux, cadeaux, petits vols à son mari. – elle devenait inquisitive et jalouse
dépenses de toilettes
s'attribue le monopole du sentiment. – Rodolphe ne la ménageant en plus la traite en putain
paroxysme
la corrompt. – exaspération. – L'enlèvement est décidé.

Plans de voyage – italie. – malle-postes. – elle devient plus belle. femme romantique
nuit conjugale. elle presse Rodolphe. Rodolphe prend son parti. – préparatifs pr l'enlèvement
Binet tournant à une lucarne – ronflemt d'abîme
Lettre sous les abricots – lecture dans le grenier. recule au moment de se jeter par la
fenêtre. "ma femme. ma femme –" Charles savoure les abricots. – 472
phases physiques. – au pire d'ab. on l'admi.
Maladie. – communie dans son lit. monde de Fiesole – Bournisien . sentiment de sirops
mièvrerie
Justin. – ses attentions. — Ψ
convalescence. cidre doux sous la tonnelle "sa bonté saute aux yeux" bonté. charité. faiblesse
souvenir de Lavallière. gdes dames religieuses
vêtements pr les pauvres. – comme cet hiver est différent de l'autre. – elle a oublié Rodolphe. – importance
~~17. monthyon~~
de la distraction.

ψ. langage doux : ta colique est-elle passée mon ange ? ~~notes des~~ affaires de voyage fournis par L'heureux. il envoie tout cela chez Charles quand il croit qu'Emma va crever. pr être payé, (au triple) & plus tard, elle lui en fait des reproches.

600 10 000 600 600 600 600 600 30 210 5

à Rouen, au spectacle. représentation de la Lucie. – rencontre de Léon. conversation au foyer.
Emma reste à Rouen.
visite à son hôtel. confidence. – ressouvenir menant à la baisade. "Vous rappelez-vous
Ah ! Je vous ai bien aimée...". très calme. sans pose. – rendez-vous d'avance pr tirer un coup.
Ψ/X — plus — Ψ/X — pas de description du coup mais s'étendre sur avant & après. différence d'avec Rodol.
Sur le port. chaleur. tentes de coutil. – coup sain. Léon ému qu'elle.
elle rentre à Yonville dans un bon état physique de fouterie normale. c'est l'époque des
confitures
fumiers roses – colère cramoisie d'Homais. – arsenic. Faublas.
Le lendemain lettre qui apprend la mort du père Bovary. Emma va à Rouen
partie sur l'eau un soir t'en souviens-tu – il en est venu avec du champ et des cornets à piston
pour les affaires de deuil et reste trois jours avec Léon. Pendant ce temps-là Charles à Yonville
en rentrant à Yonville, elle trouve la mère Bovary – tristesse de tout le monde. elle pense
à ses amours. – L'heureux – procuration. *24*
établissement des leçons de piano.
hôtel de Provence. sur le port. leur genre de relation. Léon étonné qu'elle soit si savante. en effet.
expliquer cela. – Léon à Yonville. Homais. mots de chic.
impatience des rendez-vous du jeudi – descente de la côte du Boisguillaume. Rouen dans
2 25 50 la brume – cailloux qui craquent sous les roues. ~~descent~~ entrée à Rouen. odeur d'absinthe qui sort
des cafés. manière féroce dont elle se déshabille
4 25 50 après les fouteries va se faire recoiffer – odeur des fers chauds – s'endort sous le peignoir
[illis.] qque chose de courtisanesque chez le coiffeur. – intérieurs de la gondole, en s'en retournant, un

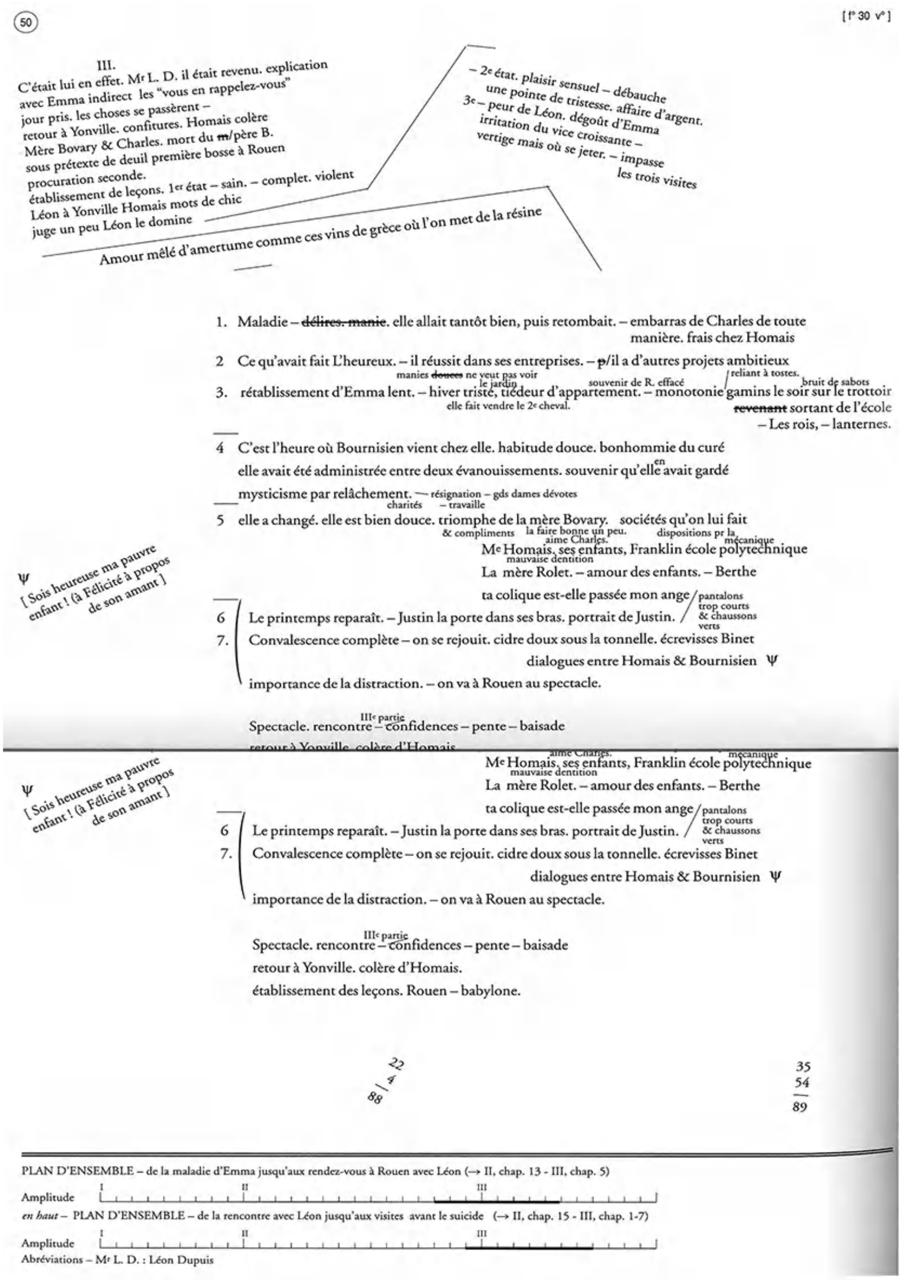
2 25 50 la brume – cailloux qui craquent sous les roues. ~~descent~~ entrée à Rouen. odeur d'absinthe qui sort
des cafés. manière féroce dont elle se déshabille
4 25 50 après les fouteries va se faire recoiffer – odeur des fers chauds – s'endort sous le peignoir
[illis.] qque chose de courtisanesque chez le coiffeur. – intérieurs de la gondole, en s'en retournant, un
peu ivre etc. cul-de-jatte dans la côte. – Charles & son enfant.
Quand elle était au lit elle lui disait. Léon au second plan. il est sa maîtresse. – développemt
Mr nerveux d'Emma. – suce son doigt en sang – (agraffe). – corne brûlée pâte incuite appétits dépravés.
Charles complètement oublié
luxe. Léon a peur. – [volcan de haine pr Charles] dépense. dépense. – mensonges. Mr L'heureux.
Léon à son étude.
14 9 23 Elle toise Léon. – continue malgré cela. elle s'en amuse. elle l'entretient. – elle caresse Charles
cependant avare sous un certain rapport – ou plutôt avide.
son chagrin de ne pouvoir la contenter. – dettes plusieurs fois avouées. & payées – beaucoup
extrême
30 3 90 plus de cachées – débauche à Rouen bal masqué. embarras d'argent et dégoût de cœur
(Léon se marie ?)
30 5 150 en même temps. – L'heureux. – alors vient l'idée de retourner à Rodolphe.
70 — 20 en 6. 40 en 3 m. 80 en 6.

SCÉNARIO D'ENSEMBLE – du retour à Rodolphe après le pied-bot jusqu'à la dernière visite, au même (→ II, chap. 12 - III, chap. 7)

Amplitude I II III

Certains chiffres écrits à l'encre très pâle ne sont pas lisibles sur le fac-similé.
Abréviations – rêveries tout ht de J. : tout haut de Justin – au pire d'ab. on l'admi. : au pire d'abord on l'admire – champ : champagne
Lexique – Monthyon : prix de vertu. Flaubert en parle dans une lettre à Louise Colet (29 mai 1852, *Corr.* II, p. 99). – Faublas : *les Amours de Faublas*, roman de Louvet de Couvray, paru en 1787

Figure 10.1. *The original text and its annotations by Flaubert, taken from Y. Leclerc's Plans et scénarios de Madame Bovary [LEC 95]*



50 [f° 30 v°]

III.
C'était lui en effet. Mr L. D. il était revenu. explication
avec Emma indirect les "vous en rappelez-vous"
jour pris. les choses se passèrent –
retour à Yonville. confitures. Homais colère
Mère Bovary & Charles. mort du ~~m~~/père B.
sous prétexte de deuil première bosse à Rouen
procuration seconde.
établissement de leçons. 1er état – sain. – complet. violent
Léon à Yonville Homais mots de chic
juge un peu Léon le domine

– 2e état. plaisir sensuel – débauche
une pointe de tristesse. affaire d'argent.
3e – peur de Léon. dégoût d'Emma
irritation du vice croissante –
vertige mais où se jeter. – impasse
les trois visites

Amour mêlé d'amertume comme ces vins de grèce où l'on met de la résine

1. Maladie – ~~délires. manie~~. elle allait tantôt bien, puis retombait. – embarras de Charles de toute manière. frais chez Homais
2 Ce qu'avait fait L'heureux. – il réussit dans ses entreprises. – ~~p~~/il a d'autres projets ambitieux
3. rétablissement d'Emma lent. – hiver triste, tiédeur d'appartement. – monotonie gamins le soir sur le trottoir ~~revenant~~ sortant de l'école – Les rois, – lanternes.
manies ~~douces~~ ne veut pas voir le jardin elle fait vendre le 2e cheval. souvenir de R. effacé reliant à tostes. bruit de sabots
4 C'est l'heure où Bournisien vient chez elle. habitude douce. bonhommie du curé
elle avait été administrée entre deux évanouissements. souvenir qu'elle en avait gardé
mysticisme par relâchement. – résignation – gds dames dévotes charités – travaille
5 elle a changé. elle est bien douce. triomphe de la mère Bovary. sociétés qu'on lui fait
& compliments la faire bonne un peu. aime Charles. dispositions pr la mécanique
Me Homais, ses enfants, Franklin école polytechnique
mauvaise dentition
La mère Rolet. – amour des enfants. – Berthe
ta colique est-elle passée mon ange / pantalons trop courts & chaussons verts

ψ [Sois heureuse ma pauvre enfant ! (à Félicité à propos de son amant]

6 Le printemps reparaît. – Justin la porte dans ses bras. portrait de Justin.
7. Convalescence complète – on se rejouit. cidre doux sous la tonnelle. écrevisses Binet
dialogues entre Homais & Bournisien ψ
importance de la distraction. – on va à Rouen au spectacle.

Spectacle. rencontre – IIIe partie confidences – pente – baisade
retour à Yonville. colère d'Homais.

Me Homais, ses enfants, Franklin école polytechnique
mauvaise dentition
La mère Rolet. – amour des enfants. – Berthe
ta colique est-elle passée mon ange / pantalons trop courts & chaussons verts
ψ [Sois heureuse ma pauvre enfant ! (à Félicité à propos de son amant]
6 Le printemps reparaît. – Justin la porte dans ses bras. portrait de Justin.
7. Convalescence complète – on se rejouit. cidre doux sous la tonnelle. écrevisses Binet
dialogues entre Homais & Bournisien ψ
importance de la distraction. – on va à Rouen au spectacle.
Spectacle. rencontre – IIIe partie confidences – pente – baisade
retour à Yonville. colère d'Homais.
établissement des leçons. Rouen – babylone.

22 / 4 / 88

35 / 54 / 89

PLAN D'ENSEMBLE – de la maladie d'Emma jusqu'aux rendez-vous à Rouen avec Léon (→ II, chap. 13 - III, chap. 5)
Amplitude I II III
en haut – PLAN D'ENSEMBLE – de la rencontre avec Léon jusqu'aux visites avant le suicide (→ II, chap. 15 - III, chap. 1-7)
Amplitude I II III
Abréviations – Mr L. D. : Léon Dupuis

Figure 10.1 (continued). *The original text and its annotations by Flaubert, taken from Y. Leclerc's Plans et scénarios de Madame Bovary [LEC 95]*

It has not gone unnoticed that these overlapping categories do not comprise an exhaustive taxonomy and literary feelings are not, and do not seek to be, examples of basic emotional categories such as those described in research in psychology and communication. The response to this paradox lies in the arbitrary nature of the more generic categories corresponding to arbitrary areas in a continuous space composed of affective "dimensions". Dimensional emotional models place emotions on indexed continuous axes that should be viewed as semantic dimensions e.g. /*positive*/ *versus* /*negative*/, /*active*/ *versus* /*passive*/ or, at the extreme end, /*enjoyment*/ *versus* /*unpleasantness*/.

This is the least of the paradoxes at play in the framework of experimental methodologies. For example, questionnaires on emotions often refer to Osgood's differentiators [OSG 57], which have a capacity of semantic description well below that of differential approaches [RAS 87]. Therefore we see how an emotion such as honor could be described as the more generic feeling of "restriction" (not necessarily lexicalized), which is found in the overlap between these dimensions.

We propose to investigate the most specific semantic categories and to move, leading towards a nominalist approach. It is therefore unnecessary to resolve the difficult problem of the kind of generic categories to be used in interaction (or contradictions) between the linguistic and the cognitive. At most, the brief discussion that we have included avoids the main objections to our approach. We will examine some examples of emotions described by Flaubert as being experienced by Emma Bovary. Their psychological nature (in the sense of the character's psychology, based on transitions between different states of mind, rather than a psychological model seeking to describe her aims and obectives) is confirmed by the lack of continuity between their characterization in the drafts and their expression in the novel itself. (The novel itself reproduces the same phenomenon in that the concept of boredom, despite being central to the plot, is represented as a semantic category rather than a true occurence of the term itself.) This demonstrates the coexistence of primary emotions, such as "Leon's fear" or "Emma's disgust" [LEC 95, p. 50] and much more specific emotions such as "irritated by vice" [LEC 95, p. 50].

It is first important to analyze the composite or unique nature of these emotions to confirm their existence as a specific category. Some of these descriptors can lead to such confusion as, for example, that referring to "a happiness mingled with bitterness, like those ill-made wines that taste of resin" (translated from [LEC 95]). Such a description is related to the periphrastic lexicalization of a semanteme corresponding to the more specific level of representation. Examining a larger number of these descriptors highlights their specificity. For example, "feeling hatred for Charles" is only specific in the overlapping of the (basic) emotion and its object. In contrast the fact of linking an emotion with its trigger creates a much more specific emotion as in “emboldened by love” [LEC 95, p. 48] or “irritated by vice” [LEC 95, p. 50].

The creation of hybrid emotions, such as “jealousy-curiosity” [LEC 95, p. 48], is another example of the subtleties of the emotions conveyed in the novel. Other descriptions such as “a feeling of emptiness” [LEC 95, p. 17] can appear to be simple elements of more complex emotions. However, we risk falling into the trap of character psychology. For example, with some feelings there are signs or criteria for the occurrence of a given psychological state. The paragon of this approach are the well-known criteria for clinical depression. If there is an area where clinical psychology and literature meet, then it is easily here. From *acedia*[6] to melancholy, we have seen in both philosophy and literature that a number of preclinical concepts have been replaced by a clinical description of depression as modern psychiatry has developed. For example, a number of terms, such as melancholy, have been given a new, more specific meaning such as “melancholic depression”.

It is therefore all the more surprising to see the term *acedia* appear in a recent study of emotions [BEL 07], a usage that may cause confusion about current opinions in psychology, and which could be compared to the historical perspective of *acedia* given by Agamben in his work *Stanze* [AGA 77]. However, can we class *acedia* as depression? From *acedia* to *saudade*[7], emotions like sadness and boredom are found throughout the history of philosophy and literature. However, our objective is to catalog emotions whose depiction is true to life and coherent with narrative and linguistic representations. The fact that these feelings correspond to a psychological or even clinical reality lies in their pragmatics (i.e. external knowledge about a cultural object, which is not directly involved in semantically describing its content).

Flaubert’s notes give us access to an expression of the emotions felt by Emma Bovary as a character rather than a psychological subject, which is an important starting point. Restructuring this domain would remove the voice of the author and lessen the possibility of anchoring emotional representations in a reference description. This ontology provides us with two things: a means of linking it to the original text’s semantic content (not only in terms of expression); and an experiment

6 Described in the Middle Ages as a torpor or listlessness affecting the well educated rather than the poor, *acedia* is often considered the “ancestor” of depression. Given modern data from psychiatry, it is difficult to compare this feeling with the contemporary (and specialized) description of depression.

7 *Saudade* is a prime example of a culturally specific emotion. A kind of existential sadness about the state of the world and a feeling of loss or absence, it is an important part of Portuguese culture, as found in Fado music and in the work of Fernando Pessoa.

in automatically generating narrative content using the characters' described emotions.

FEELINGS FROM *FLAUBERT* (*PAGE IN "PLANS ET SCENARIOS"*)	DESCRIPTION	OPERATOR
PRIDE-OF-HAVING-A-LOVER *(P. 48)*	*Part and Chapter:* **II, 9** *Page in Novel (Eds. Garnier):* **167** *Narrative Sequence – Motif/Theme:* **"*BAISADES*" (Fling) – Joy of Emma**	*Name:* **Joy-of-Love(EMMA)** *Precondition:* **Affinity(E,R,3)** **Pride(E,3)** **Womanhood(E,3)** *Effect:* **Satisfaction(E,3)**
IRRITATED-BY-VICE *(P. 50)*	*Part and Chapter:* **II, 15** *Page in Novel (Eds. Garnier):* **232** *Narrative Sequence – Motif/Theme:* **OPERA – Meetings with Léon** **NOTE:** Originally for Léon, the feeling is adapted here to Rodolphe.	*Name:* **Regret-Falling-for (EMMA, RODOLPHE)** *Precondition:* **Embarrassment(E,3)** **Affinity(E,C,2)** **Power-over(R,E,3)** *Effect:* **Affinity(E,R,1)** **Anger(E,R,3)** **Affinity(E,C,3)**

Figure 10.2. *The emotions described by Flaubert are used to create a corpus of the main characters' emotions as a prelude to their computer formalization. Planning operators determining the narrative's dynamic take their preconditions from these emotions*

10.4. Interactive fiction and emotional planning

Interactive narration can also be achieved with user involvement paradigms. Historically, the role of the audience in pioneering experiments with interactive cinema involved choosing a possible next event and positioning it outside of the narrative. With regard to the interaction itself, this was more basic. Contemporary research into interactive fiction uses the *Holodeck* [MUR 98] as a model. It is a virtual reality system in which the user takes on the role of one of the narrative's actors and interacts naturally with other (virtual) actors, physically affecting the environment. Based on this model, interactive fiction systems can be classified according to two dimensions: the role of the user in the story (actor or viewer-controller) and the frequency of his/her intervention in the narrative.

The role of the user has a major influence on linguistic interaction. For example, if he/she has to interact as a character, the type and level of language must match the style of dialog expected for the genre in question. With regards the Bovary system, this has been developed using the narrative's central events, such as dialog between Emma and Rodolphe. The user could therefore be in "Rodolphe's skin" and substitute his/her own replies (in the wider sense, without changing interactions' outcomes) to Rodolphe's. It is, however, agreed that this must respect the original style of the novel, i.e. genre conventions.

Interactive fiction aims to reconcile narration (the progression of a story presented in a visual media) and interaction (the possibility for the audience to intervene in narration or to influence it). Without examining the overall history of this technology, we should simply say that this apparent contradiction is solved by the generation of the narrative's event in real time using a basic background narrative. For example, in the Bovary system we have retained the background story by leaving Emma's reaction open to different temptations and deceptions found in the novel, and making it reliant on the dynamic situation that will be influenced by the user's interventions.

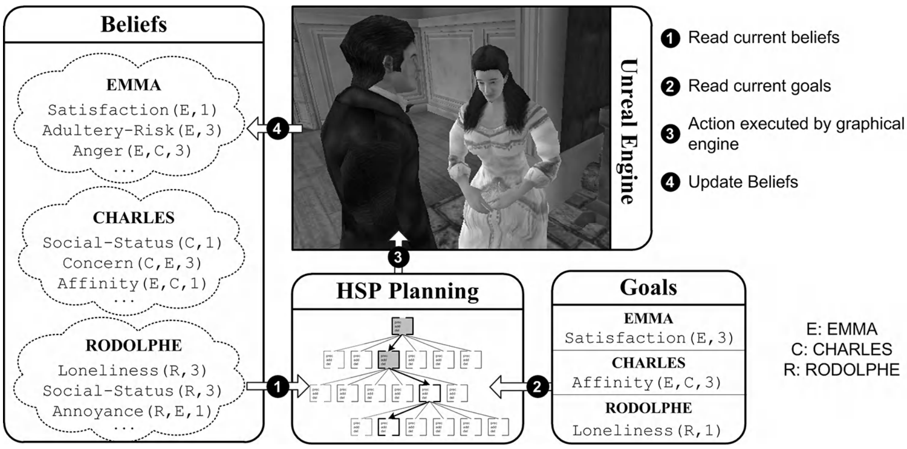


Figure 10.3. *Principles of the interactive fiction system based on Madame Bovary: the characters' feelings guide their actions, which in return modify these emotions. Their actions are represented in the form of animations, which in sequence creates an interactive (animated) movie*

A default generation mode creates a basic story that is very close to the original novel without any influence or intervention with the book's original conditions. From a representational point of view, rather than changing the essential events at decision points, these are dynamic objects whose occurrence and realization is subject to narrative context. The dynamic of these actions and their combination to create a narrative sequence requires the intervention of generative interventions that, in the majority of cases, are based on planning techniques.

If we consider that a plot is composed of an action sequence, it is logical to use planning techniques to produce a series of narrative events. It is at this representational level that there are genuine challenges both in terms of realism (the action sequences generated must be coherent with the narrative plot and not only in terms of generic constraints derived from shared meaning) and performance (the generation must be sufficiently fast so as to not compromise the story's continuity). There are two main approaches in interactive fiction where planning either alters the whole story or focuses on the behavior of a specific character, with as many possible plots as there are characters in the case of the latter.

We have chosen to generate a story on the basis of individual characters' behavior, as established by a previously-developed approach [CAV 02a]. If the generation of these events uses planning, it should not be assumed that the characters' roles can be reduced to a single plot whose fulfillment would constitute a "solution to the problem of the narrative". This is overly optimistic and, in practice, inaccurate. Analysis of the novel itself disproves this conception. If Emma Bovary commits suicide at the end of the novel, then this is not so much a planned solution to the existential problem of boredom (which is an overall aim), but an instantaneous reaction in the face of dishonor.

A weaker approach of formalization is therefore required where planning is only used for its generative capacity, where the aims are dynamic and modifiable and current motivations are again used to drive events. With regard to the plot's content, this is most often adapted to the style of the novel. Rather than dramatic events as found in epics or fairy tales, the content of *Madame Bovary* is based on the psychology of its characters and, specifically, their emotions and feelings. Several major events, whether real or potential, in scenes that we want to model, such as Emma's fleeing Yonville, are based on the characters' emotional state. It is therefore logical to base our representation on the characters' emotions and ensure that planning alters these emotions according to their exchanges with other characters (with the user playing the role of Rodolphe), triggering events that are more compatible with characters' emotional states. Included in the term "emotional planning" [GRA 99] encompasses the entirety of techniques changing virtual agents' (or autonomous agents') feelings, depending on the planner's behavior. Traditional approaches make emotions appear in reaction to the current plan's probability of success. This type of reference framework is reminiscent of ethology.

Moving away from this utilitarian view, we have chosen to combine the representation of emotions in planning and therefore directly use an "ontology" of feelings, as mentioned, to define the planning domain's content. This also corresponds to a representational bias that favors the representation of cultural content over the cognitive modeling of characters' personal states, often ignorant to the novel's narrative plot or its aesthetics.

We have used the emotions described by Flaubert in [LEC 95] as a starting point, specifically with regard to their subtlety or level of nuance. We will consider the planning domain as being defined by a list of emotions that can be associated with an intensity described by a tri-valued variable (LOW, MEDIUM and HIGH). This domain carries generic emotions, such as affinity or the “amount of affection” between characters (e.g. between Emma and Charles or Emma and Rodolphe), that are described in Flaubert’s drafts. In terms of planning, emotions cannot only be represented as facts (e.g. the feeling of loneliness or shame) but also, more innovatively, as operators. In the latter case, specific feelings (e.g. the “regret of yielding to Rodolphe”) are valued less by their existence (or an equivalent value of truth) than by the change of mood and day-to-day life that it causes. These emotions are therefore represented by operators whose pre and post conditions refer to more basic emotions whose character or strength is altered based on the experience of this feeling. This approach indirectly solves the problem of emotional dynamics, i.e. their spontaneous development through time, at least for more complex emotions, by reducing the current emotional state to a number of basic emotions whose intensities are reevaluated throughout the events and interactions between characters.

Formalization includes two distinct phases. The description of the planning domain records the possible states of the world; and characters’ mental states in a uniform way, the latter being defined by their emotions. The planning operators’ descriptions are responsible for developing characters’ mental states as well as the actions that they undertake. We can distinguish two main types of operators: communication operators and interpretation operators. The first type determines the communicative actions of one character towards another. Their preconditions include the emotions felt by the agent producing the utterance. A prime example of this might be that, if Emma feels a strong degree of affinity for Rodolphe and she is therefore ready to take risks (a state where her desire for pleasure dominates her sense of duty or social constraints, although this is not in itself a feeling), she becomes likely to declare her love for him.

This has therefore created determinism for emotion-based actions. Communicative actions are designed to influence the emotions felt by the characters to which they are addressed. However, they cannot predetermine these emotions, in that their impact depends on the recipient’s requirements. The affect of a communicative action is in turn determined by an interpretation operator as part of the recipient’s planning system. For example, Rodolphe reacts to Emma’s declaration based on his emotions at this time, which in turn depend on previous exchanges between the two.

As such, the affinity that he feels can be replaced by other feelings, such as annoyance or fear of commitment, and it is their balance that will determine his reaction to Emma’s declaration. The planning system therefore operates on the basis of characters’ emotions. The interpretation operators ensure that the context and history of their interactions are taken into account and such planning is done step-by-step in real time, with each situation giving rise to a new planning stage.

Each action is visualized as digital animations accompanied by subtitles corresponding to dialogs related to communicative events that are also spoken by characters thanks to a vocal synthesis module. A moment in the story, which appears to the user as a complete animation lasting a few minutes (from two to six minutes in our most recent experiments), can be formally represented in the form of a development diagram that indicates the immediate move by the story towards an ending corresponding to the choice between pleasure and duty that Emma Bovary faces.

10.5. Linguistic interaction and emotions

The overall story is generated using characters' feelings: the intervention of the user will alter the intensity and type of emotion felt by the characters (those of Emma, in this case, as the user adopts the role of Rodolphe). The use of language is the only option that can satisfy both the need for natural communication and for a contribution to the narrative's natural form: a response from Rodolphe to Emma is part of the interaction (when the user takes on the role of Rodolphe) and the story itself, in its theatrical form, as well as its dialog. This contradiction generally remains an unresolved problem since it leads the user to "play" their role as an actor. This can be facilitated when the user is incorporated into the story, as occurs in mixed reality systems where the user's video image is reflected in a mirror, mixed with other characters' virtual images [CAV 04].

In the current case, there is a practical limit to using vocal recognition because current performance in this area does not allow a sufficiently reliable comprehension of long or complex sentences whose conditions of production are not easily controllable in an interactive fiction context. We have tried, above all, to keep the use of language similar to that used in the novel, whose register and style match Flaubert's original drafts. In order to do this, interaction with the prototype is carried out using written language. A written reply can be analyzed at different levels, allowing us to get around the limitations of automatic analysis, specifically for phrases of grammatical complexity such as those found in the novel. The guiding principle of the analysis consists of identifying the emotions contained in the reply in order to link them to those composing the character's emotional state when this reply is received.

Several earlier interactive fiction prototypes have incorporated natural language processing as a form of interaction as written [MAT 04] and spoken dialog or even simple oral interventions [CAV 02b]. Dialog is more complex, however, particularly if interactive dialog seeks to reproduce the narrative's original style (although the user's interventions can take the form of advice expressed in a shared language, even with simple interjections). If oral dialog naturally reproduces the narrative's style, it should not be concluded that written dialog should automatically be unrestricted and closer to simple interventions. We have therefore preserved the hypothesis that dialog should reflect the language register used in the novel and have opted for an appropriate analysis considering the amount of more complex language that the system has to process.

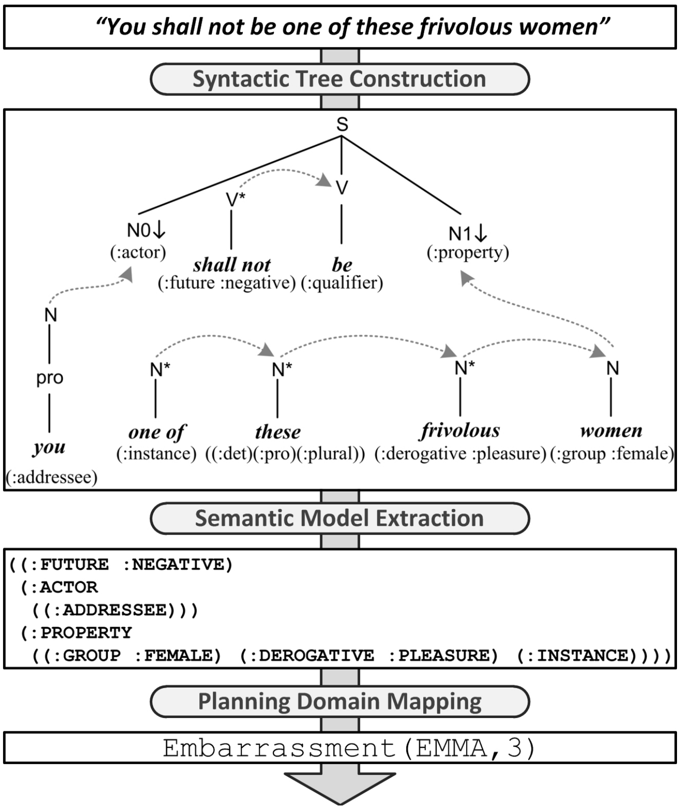


Figure 10.4. *Automatic language analysis for user responses to Emma. This is created using semantic categories based on affective categories (the emotions in the novel)*

The type of language processing that we have just introduced is often described by the term *Emotional Natural Language Processing* (ENLP). There are different approaches to this, the simplest of which being indexing terms according to emotional categories with the most complex in principle preceding more traditional semantic analysis (assembling semantic representations using the sentence's syntactic structure but using emotional semantic traits). It is this more ambitious approach that we have selected and applied. To better learn the speaking style used in *Madame Bovary* as well as the syntactic breadth that it involves, we have collected and manually analyzed all the responses found in the novel, modeled according to the prototype's needs.

Ironically, implementation was carried out in English, since it was the responses from the English translation that were used to define the syntactic coverage required while all theoretical study of the narrative prior to the interactive fiction system's design was carried out using the original texts and their French annotations. The few stylistic variations found between different translations do not have any serious practical impact on the automatic analysis [UZU 05].

The final aim of the analysis is to identify the structures of semantic features whose components correspond to the emotions represented in the planning domain. While the analysis overall is geared towards identifying these categories, however, we have chosen to carry out a syntactic analysis that may underlie the organization of semantic structures. The semantic description and the definition of lexical content are based on Rastier's *unified differential semantics* [RAS 89][8]. Based on textual semantics, semantic descriptions should be able to produce more nuanced categories than those corresponding to the generic emotional markers proposed by various "emotional lexis" that have appeared in recent years [VAL 05].

In the case of semantic lexis supporting automatic analysis of user responses, differential descriptions can be structured around the theme of the novel, identifying opposing themes such as */boredom/ versus /pleasure/*, */illusion/ versus /reality/*, */fear for one's reputation/ versus /accepting risk/*, etc. The use of semantic features allows us to get around the book's lack of lexicalization of central concepts such as boredom (see Rastier's [RAS 95b] comments on non-lexicalization of boredom in *Madame Bovary* and [RAS 89 and RAS 95b] for examples of lexemes included in the term */boredom/*).

The lexical description [CAV 96] should at least take into consideration context, as is common for this genre (the genre gives a certain regularity to context that allows the content at least to reach an initial level of genericity). We can best capture these linguistic phenomena using some of the most famous phrases from the novel: *"la conversation de Charles était plate comme un trottoir de rue, et les idées de tout le monde y défilaient in leur costume ordinaire, sans exciter d'émotion, de rire ou de rêverie"* [Charles's conversation was commonplace as a street pavement,

8 Differential semantics describe the semantic context of linguistic units in context. The term takes its name from the use of systems using oppositions within classes of meaning. In relation to traditional structuralist approaches, this methodology is used on texts rather than abstract language and therefore a number of studies have focused on the relation between linguistics and literature.

and everyone's ideas trooped through it in their everyday garb, without exciting emotion, laughter, or thought] which is a typical case of isotopy[9] on the semanteme /*boredom*/ (with, among others "dull", "everyone", "ordinary"). It should be noted, however, that because there is a possible reserve, in that in interactive narration analysis must be based on elements of dialog whose register may be slightly lower than that of the text, even when this a question of using a specific character's style.

The perspectives opened up by this approach become more obvious when we compare research into corresponding semantic traits to previous approaches to language processing in interactive fiction, which has been dominated by pragmatic analysis. Analysis of responses has traditionally sought to isolate language acts related to narrative functions or even, when dialog is essential to narrative action, identify with them. In some previous experiments, the number of language acts described is variable and can sometimes result in disagreement [CAV 04, MAT 04]. If we consider the following response from the novel:

Comme je resterais ce soir, demain, les autres jours, toute ma vie!

[And I shall remain to-night, to-morrow, all other days, all my life!]

Traditional language act analysis aims to identify a promise and even more so in a novelistic context, a vow of love. We also see that the language act can identify a narrative function, constituting a strong integrator principle. However, this kind of analysis ignores the difficulty of identifying a language act using a literary expression devoid of basic conversational markers that characterize ordinary conversation.

A semantic approach to analysis, on the other hand, can provide a more efficient method of identifying relevant semantic elements as well as linking them to representations from planning. As such, the identification of traits such as /*presence*/ and /*duration*/ constitutes the basis for the narrative functions of eternal love versus abandonment. This principle of analysis developed through previous research [CAV 04] therefore identifies emotions by aggregation of semantic traits, followed by means of identifying themes in differential semantics [RAS 95b]. In emotional natural language processing, similar approaches have recently been described by Basili and Marocco [BAS 06a, BAS 06b] as well as those that use latent semantic analysis [LAN 97] and by Gliozzo [GLI 06] who have defined semantic domains that are equivalent to semantic differential classes [RAS 94] (the term "domain" being used in the generic sense here).

9 An important feature of semantic description, isotopie is defined as the "syntagmatic reoccurrence of the same semanteme". The repetition of a semantic category in several words of one phrase creates an impression of continuity and aids the text's cohesion.

Based on the linguistic description principles set out here, we have carried out an initial implementation of an analyzer that allows us to recognize the emotions evoked by the user's responses. Its covering, specifically syntactic, is indeed limited, but relies on the analysis of all dialog from *Madame Bovary* based on the hypothesis that users' interventions must have some proximity with the book's original style to retain the desired impression for interactive fiction. The aim of this analyzer, which can be characterized as an emotional natural language analyzer, is to reproduce the phrase's semantic structure, including semantic features corresponding to the emotions expressed in the novel, which can be projected on the planning domain to influence the behavior of the characters that the response is aimed at. The development of this semantic structure is guided by syntactic links obtained from lexicalized grammar using a simplified form [CAV 98] derived from *tree-adjoining grammar* (TAG) [JOS 92].

This formalism has several practical advantages, making it a well adapted approach for the problem at hand:

– it can support partial analyses;

– it is well adapted to describing complex and idiomatic structures, and therefore the level of style that we wish to analyze (even without aiming for a complete analysis producing a single phrasal tree head).

The analyzer that we have used has been adapted from a previously developed version [CAV 98] integrating syntactic and semantic analysis, with the final product therefore being a structure of hierarchical semantics. This is preceded by ascending syntactic analysis by combining adjacent lexicalized trees with substitution operations and *furcation* (merging trees, replacing the traditional approach of addition in joined grammar trees) depending on the trees' syntactic category. The combining operations are accompanied by fusing of the semantic structures attached to the trees' main lexical anchors. In this way, the resulting semantic structure is constructed in parallel with syntactic analysis. Through the *furcation* process, semantic features are propagated from the auxiliary tree towards to the tree to which it will be joined. During substitution, the functional links to the nodes that are going to be replaced are used to structure the semantic structure under development (see Figure 10.4). As such, an `actor:` link could give rise to primitively structured semantics.

When syntactic analysis does not produce a single phrasal tree structure, the different semantic structures corresponding to each of the syntactic fragments can be merged to produce a usable structure (usually identical to an ideal structure or a similar hierarchical structure).

In other words, the aim of syntactic analysis is not so much to identify an actant structure that can be assimilated with a narrative (even communicational) function, but to recognize the recurrence of semantic traits whose assemblage can be identified with one of the emotions from the description domain.

The emotional language module functions according to three stages:

– the TAG analyzer produces a semantic structure (based on semantic features) in parallel with the syntactic analysis of the phrase typed on the keyboard by the user as a response to Emma;

– this semantic structure is interpreted not as a logical formula representing the literal (or even derived) meaning of the user's response, but as a resource for extracting semantic themes based on emotional traits (corresponding to the novel's emotions) which could be interpreted in the planning domain; and

– these sub-structures are projected on the planning domain using simple association rules. Some of these sub-structures are found in the operators' trigger conditions in the planning domain. As such, some semantic traits (/*illusion*/, /*selfishness*/, /*prodigality*/) can be projected on Emma's emotional domain (such as embarrassment or shame, as the result of her actions).

This mechanism's effectiveness in the system is illustrated by analyzing the following phrase, adapted from the novel's original text (which we have used as a replica during the experimental development of the system):

> *"Mais cette exaltation délicieuse vous a empêchée de comprendre la fausseté de notre position future."*

[This charming exaltation has prevented you from understanding the falseness of our future situation.]

The semantic complexity of this sentence puts it out of reach of contemporary language analysis techniques: even at best, the effort of processing would be disproportionate given the system's other constraints. However, if one of the aims of analysis is to construct a representation of the sentence's theme using semantic features activated by the words composing it (grouped according to the syntactic relations linking them), then it is possible to analyze this response in real time. Figure 10.5 illustrates the type of semantic representation produced according to this principle. The analysis produces a fairly rich semantic structure, formed by incorporating semantic features around casual primitive traits (e.g. actor, instrument, object, etc.) that involves a certain amount of information redundancy (see Figure 10.5). The advantage of this approach is that it allows us to make relevant inferences with a small role-play using semantic features corresponding to the novel's main themes (e.g. /*illusion*/ *versus* /*reality*/).

```
(:ACTION :CAUSE-NOT-TO)
(:CAUSE
 ((:ROMANTICISM :ILLUSION) :ENDEARING (:DET) (:PRO)))
(:PATIENT
 ((:ADDRESSEE)))
(:OBJECT
 ((:COG-ACTION :KNOWLEDGE)
  (:KNOWLEDGE-OF
   ((:ILLUSION)
    (:ILLUSION-OF ((:SITUATION) (:POSS :ADDRESSEE :SPEAKER))))))
```

Figure 10.5. *The semantic structure produced by analyzing the above phrase*

In the example in Figure 10.5, the roles (`:speaker` and `:addressee`) are performed by the actors Rodolphe and Emma. The most significant sub-structure (`:illusion-of ((:situation (:poss Emma Rodolphe)))`), and the recurrence of the theme /*illusion*/ allows us to make inferences for the planning domain, leading to the creation of emotions such as embarrassment (`embarrassment(Emma, HIGH)`). We have therefore defined a set of inferences that trigger the projection of affective dimensions onto the planning domain. With regard to semantic features describing lexis, these are influenced by differential semantics. The classes describing them have been inspired by lexicographical research into emotions in novels [RAS 94].

10.6. Emma Bovary's virtuality

Having modeled the main narrative actions in terms of an ontology of the novel's original emotions, it is possible to generate alternate fragments of *Madame Bovary* that, while retaining a certain amount of coherence with the initial book, allows us to produce variants. These variants take two forms.

On one hand, static variants producing animations correspond to another possible development of the story, without the normal progression of the original story being influenced by the user. On the other hand, conforming to the aims of interactive fiction, the intervention of the user will modify the progression of the story and therefore *de facto* produce a variant that is compatible with the type of intervention.

In the research that we have described, it is well understood that user intervention occurs from inside the narrative. By temporarily playing the role of Rodolphe, the user is participating in the evolution of Emma's emotions and consequently causes an appropriate reaction, whether immediate or not. We can characterize this intervention using the mixonyme "viewer". This does not convey the user's direct control of the narrative's events, however, such as those resulting from the formulation of orders or direct instructions given to Emma: we therefore find ourselves in a staging reasoning, often described in relation to experiments with video games [BOL 98] but to which interactive fiction cannot be reduced.

One of the natural conditions for modeling is the capacity to regenerate the original development of the chapters represented in the model in the absence of interaction. We have managed to obtain this kind of validation effectively, with minimal variations with regard to non-essential action permutations or the length of events, also determined by the application of three dimensions (the likelihood of such encounters depends, for example, on the dispersal of characters in the initial scene).

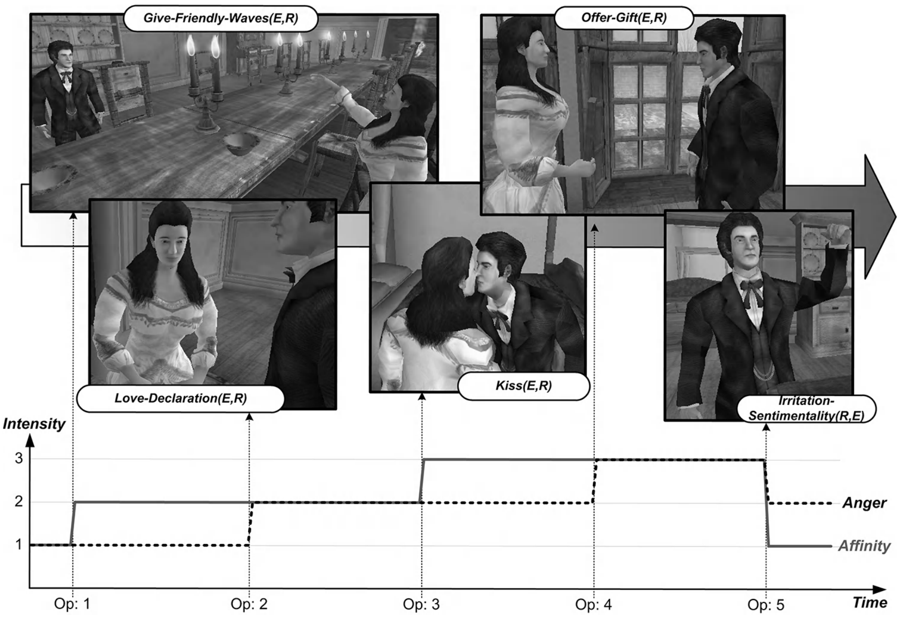


Figure 10.6. *Example of scenes generated by the system, characterized by triggered planning operators and the level of intensity of some emotions essential to the relationship between Emma and Rodolphe*

In the following sections, we will provide an example of a variant of the story generated by the system. The written presentation of this kind of story, even illustrated by screen sensors, can only partially describe the events observed but can include explanatory elements illustrating the development of characters' emotional states. We will examine the original book's narrative context, followed by the sequence of events generated by our system. Rather than describe its functioning step-by-step using the sequence of formal properties in the planning domain, in describing the example the semantic aspects of the development of characters' emotions should be emphasized. More advanced examples can be found in [CAV 07].

After Leon leaves, Emma begins to feel the resurgence of her boredom and regrets that her "only love has gone". She is surprised that she feels an aversion to her daughter Berthe, reflecting the dull life that she shares with her husband. Emma therefore becomes acquainted with Rodolphe Boulanger, whose servant is a patient of her husband. He immediately notices the distance that has developed between Emma and her husband and decides to try to seduce Emma. When her husband Charles is away visiting patients, Rodolphe decides to visit Emma. When she first meets Rodolphe, Emma's is overwhelmed by curiosity. The resulting event is an invitation to conversation (operator 1: *Gives-Friendly-Waves*). The progression of the event (accepted by Rodolphe due to his interest in Emma – basic emotion) leads to an increased reciprocal affinity between the characters (an increased feeling of affinity). The development of the relationship leads to increased interest in Rodolphe on the part of Emma. When this passes a certain point, Emma declares her love for Rodolphe (operator 2: *Love-Declaration*). This level of interest has a negative effect on Rodolphe (his feelings of annoyance increase).

However, this does not affect his basic affinity for Emma and, due to their developing relationship, the two lovers soon kiss (operator 3), in turn giving rise to a new amorous relationship (renewed increased affinity). The establishment of this relationship corresponds with the feeling of pride at having a new lover (as described in the original book), which can be explained by an improved self-image following a state of boredom (precisely defined in this way in the system). The emotions that increase in intensity are pride and risk taking. The latter is a component of the social constraints acting on Emma, one of the novel's other major themes (see Figure 10.7).

Emma's disinhibition enables her to carry out actions that would have previously been impossible, such as offering presents (operator 4: *Offer-Gift*) to Rodolphe for example (a sign of commitment, the novel's infamous *amor nel cor*). Rodolphe's reactions are determined by his main motivations (that we have termed *drivers*, see [PIZ 07] for further information) and Emma's attitude is perceived as intrusive (Rodolphe's annoyance increases) and he therefore expresses this irritation (operator 5: *Irritation-Sentimentality*) leading to temporary relief (his annoyance reduces) but which is the beginning of the end of his relationship with Emma (reduced affinity).

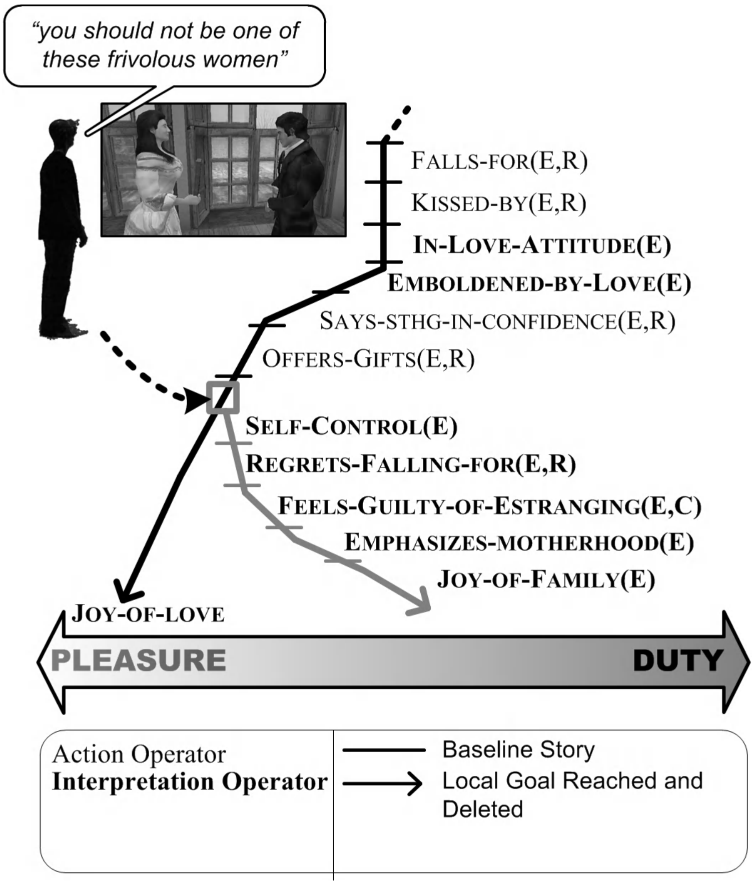


Figure 10.7. *The influence of the user can be represented by the development of the story seen via the main dimensions structuring the novel, i.e. the opposition between duty and pleasure*

This example allows us to make several observations. Emotional representations are found throughout the planning domain (the ontology of emotions leading to certain observations) and its operators (i.e. the action modes influenced by these emotions). The latter represent the traditional component of an interactive fiction system, since there can be no purely emotional system. The acting component has an inevitably arbitrary role in the selection of possible events. We have also tried to relate these actions to the description of the characters' motivations given by Flaubert [LEC 95].

10.7. Conclusion

Interactive fiction provides an opportunity to use various aspects of recent research in emotional interfaces, from the representation of a story's emotional content, as desired by the author him/herself, to the creation of characters that can feel and express emotions (we have not addressed this topic, since it is the subject of Chapter 6). Current research is crossed by a number of dichotomies, the most important of which being the distinction between cognitive and cultural approaches, on which this chapter is based. This has been a recurrent theme since the establishment of cognitive sciences as an interdisciplinary field (on which there is an extended discussion in [RAS 91]).

It is particularly important that each time a new informatics system is developed, that linguistic content (specifically textual [RAS 89] content) is processed or, more generally, cultural content whose interpretative modalities go beyond modeling in the form of universal basic emotions that would involve regenerating parts of the discourse environment. Characters' psychological mechanics (whether real or virtual) are less important than the narrative's discourse environment as this also extends to their emotions.

A number of cognitive theories have been illustrated using narrative examples (the famous "Othello believes that Desdemona loves Cassio"), but it would be absurd to seek to regenerate the notion of jealousy using equations highlighting the relations between characters, predicted by their desires and beliefs about one another. Such formalization certainly has a certain epistemological significance, but isolated from the literary body and narrative that it describes, its power of description is limited to that of a prototype and cannot capture the nuances found in the original context.

The cognitive approach examines universality where the cultural approach only focuses on nuance and diversity. How, then, can we change a situation that sees this first approach lacking critical meaning and the second lacking the ability to produce formalized representations? This challenge lies in descriptive semantics of emotions, which have influenced the approach described by [RAS 94] who produces a highly contextualized content that has enabled him to create computer formalisms that could dynamically create contextual representations. The continuity between semantics and narratology should not be underestimated. Since the time of Greimas [GRE 66] and Barthes [BAR 68], even if their main field of description was that of events, we have extended this to the study of emotions, which are the link between a narrative event's semantics and the entire plane of expression – whether this is textual or another semiotic modality.

This continuity is based in part on the need for textual semantics to extend its reach towards "integrated pragmatics" [RAD 94], in other words to reincorporate contextual elements that play a role in defining and updating semantic content into linguistic description. This therefore leaves "inclusive pragmatics", defined as a body of knowledge without direct interaction with the text itself, and "communicational context", which itself gives way to the theory of emotion that has long dominated research into affective interfaces, bringing it full circle.

10.8. Bibliography

[AGA 77] AGAMBEN G., "Stanze: la parola e il fantasma nella cultura occidentale", *Saggi*, no. 579, Einaudi, Turin, 1977.

[ALL 83] ALLEN J.F., "Maintaining knowledge about temporal intervals", *Communications of the ACM*, vol. 26, no. 11, pp. 832-843, November 1983.

[BAR 70] BARTHES R., *S/Z*, Le Seuil, Paris, 1970.

[BAS 06a] BASILI R., MAROCCO P., "A geometrical approach to literary text analysis", *Proceedings of the LREC 2006 Workshop, Towards Computational models of Literary Analysis*, Genoa, May 22,2006.

[BAS 06b] BASILI R., "Toward computational models of literary analysis", *Proceedings of the LREC 2006 Workshop*, Genoa, May 22, 2006.

[BEL 07] BELZUNG C., *Biologie des Emotions*, De Boeck University, Bruxelles, 2007.

[BOL 98] BOLTER J.D., GRUSIN R., *Remediaztion: Understanding New Medias*, MIT Press, Cambridge, 1998.

[BRY 06] BRYANT J., VORDERER P., *Psychology of Entertainment*, Lawrence Erlbaum Associates, Mahwah, 2006.

[CAV 96] CAVAZZA M., "Semiotique textuelle et contenu linguistique", *Intellectica*, vol. 23, pp. 53-78, 1996.

[CAV 98] CAVAZZA M., "An integrated parser for TFG with explicit tree typing", *Proceedings of the Fourth TAG+ Workshop*, University of Pennsylvania Press, Philadelphia, 1998.

[CAV 02a] CAVAZZA M., CHARLES F., MEAD S.J., "Character-based interactive storytelling", *IEEE Intelligent Systems*, vol. 17, no. 4, pp. 17-24, 2002.

[CAV 02b] CAVAZZA M., CHARLES F., MEAD S.J, "Interacting with virtual characters in interactive storytelling", *Proceedings of the First International Joint Conference on Autonomous Agents and Multi-Agent Systems*, AAMAS'02, Bologna, pp. 318-325, 2002.

[CAV 04] CAVAZZA M., MARTIN O., CHARLES F., MEAD S.J., MARICHAL X., NANDI A., "Multi-modal acting in mixed reality interactive storytelling ", *IEEE Multimédias*, vol. 11, no. 3, 2004.

[CAV 07] CAVAZZA M., LUGRIN J.-L., PIZZI D., CHARLES F., *Madame Bovary on the Holodeck: Immersive Interactive Storytelling*, ACM Multimedia 2007, Augsburg, 2007.

[ERL 95] ERLICH D., "Une méthode d'analyse thématique. Exemples de l'ennui et de l'ambition", in: RASTIER F. (ed.), *L'Analyse Thématique des Données Textuelles: l'Exemple des Sentiments*, pp. 85-103, Didier, Paris, 1995.

[FLA 56] FLAUBERT G., *Madame Bovary*, La Revue de Paris, Paris, 1856.

[GLI 06] GLIOZZO A.M., "Semantic domains and linguistic theory", *Proceedings of the LREC 2006 Workshop Toward Computational Models of Literary Analysis*, Genoa, May 22, 2006.

[GRA 99] GRATCH J., "Why you should buy an emotional planner", *Proceedings of the Autonomous Agents 1999 Workshop on Emotion-based Agent Architectures, EBAA'99*, Workshop 5, May 1, 1999.

[GRE 66] GREIMAS A.J., *Sémantique Structurale*, Larousse, Paris, 1966.

[HUY 77] HUYSMANS J.K., in: FUMAROLI M. (eds), *A Rebours*, Gallimard, Paris, 1977.

[JOS 92] JOSHI A.K., SCHABES Y., "Tree-adjoining grammars and lexicalized grammars", in: NIVAT M. and PODELSKI A. (Eds), *Tree Automata and Languages*, Elsevier Science, Amsterdam, 1992.

[LAN 97] LANDAUER T.K., DUMAIS S.T., "A solution to Plato's problem: the latent semantic analysis theory of acquisition, induction and representation of knowledge", *Psychological Review*, vol. 104, no. 2, pp. 211-240, 1997.

[LEC 95] LECLERC Y. (Ed.), *Plans et Scénarios de Madame Bovary*, CNRS Editions, Paris, 1995.

[MAT 04] MATEAS M., STERN A., "Natural language understanding in façade: Surface-text processing", *2nd International Conference on Technologies for Interactive Digital Storytelling and Entertainment*, TIDSE 2004, pp. 3-13, 2004.

[MUR 98] MURRAY J.H., *Hamlet on the Holodeck: The Future of Narrative in Cyberspace*, MIT Press, Cambridge, MA, 1998.

[OSG 57] OSGOOD C.E., SUCI G.J., TANNENBAUM P.H., *The Measurement of Meaning*, University of Illinois Press, Urbana, 1957.

[PIC 97] PICARD R., *Affective Computing*, MIT Press, Cambridge, 1997.

[PIZ 07] PIZZI D., CHARLES F., LUGRIN J.-L., CAVAZZA M., "Interactive storytelling with literary feelings", *The Second International Conference on Affective Computing and Intelligent Interfaces*, ACII07, Lisbon, 12-14 September 2007.

[RAS 87] RASTIER F., *Semantique Interprétative*, PUF, Paris, 1987.

[RAS 89] RASTIER F., *Sens et Textualité*, Hachette, Paris, 1989.

[RAS 91] RASTIER F., *Sémantique et Recherches Cognitives*, PUF, Paris, 1991

[RAS 94] RASTIER F., CAVAZZA M., ABEILLE A., *Sémantique pour l'Analyse, de la Linguistique à l'Informatique*, Masson, Paris, 1994.

[RAS 95a] RASTIER F. (Ed.), *L'Analyse Thématique des données Textuelles, l'Exemple des Sentiments*, Didier, Paris, 1995.

[RAS 95b] RASTIER F., "La sémantique des thèmes – ou le voyage sentimental", in: RASTIER F. (Ed.), *L'Analyse Thématique des données Textuelles. L'Exemple des Sentiments*, Didier, Paris, pp. 223-249, 1995.

[SCH 77] SCHANK R.C., ABELSON R.P., *Scripts, Plans, Goals, and Understanding: An Inquiry into Human Knowledge Structures*, Lawrence Erlbaum Associates, Hillsdale, 1977.

[SMI 03] SMITH G.M., *Film Structure and the Emotion System*, Cambridge University Press, Cambridge, 2003.

[UZU 05] UZUNER Ö., KATZ B., "Capturing expression using linguistic information", *Proceedings of the Twentieth National Conference on Artificial Intelligence*, AAAI 2005, pp. 1124-1130, 2005.

[VAL 05] VALITUTTI A., STRAPPARAVA C., STOCK O., "Lexical resources and semantic similarity for affective evaluative expressions generation", *ACII2005*, pp. 474-481, 2005.

Chapter 11

The Design of Emotions: How the Digital is Making Us More Emotional

11.1. Representing, interpreting and evoking emotions

Networks and computers have gradually begun to move away from symbolic forms, such as MSDOS, towards the analogy of the "office" and then *Second Life*-type spaces. The significance of these environments is not only that they are evolving from 2D to 3D, but that they allow us to explore ourselves and others and, potentially, build a unique relationship with a digital image [GEN 07a]. However, what is an emotion in these kinds of environments? Can a digital image even evoke emotion? How can an avatar or intelligent virtual agent not only display emotion but induce emotion in users? This chapter will focus on the issue of representing and eliciting emotions.

Most image processing research has focused on real-life images, notably of faces and gestures, a type of photographic mimicry that allows us to correctly interpret the emotions conveyed and, potentially, to share in them. However, we also need to move away from the realism of "theme" or "subject" as it cannot convey the emotional experience of these digital worlds in its entirety. In fact, digital media now present a specific form of emotional expression that we should address.

Chapter written by Annie GENTÈS.

Emotion-Oriented Systems Edited by Catherine Pelachaud

Each medium develops its own esthetic forms and relates to and affects those who use it in original ways. The user is involved in unusual unique interaction with each specific device. It should be noted, however, that this specific esthetic is far from being fully established at the start, because inventors and users of new media tend to copy what they used to do with previous media. For example, it has frequently been highlighted how photography initially imitated forms of painting. Gradually, with the influence of "pioneers", artists, designers and creators, specific esthetic qualities of the new media emerged. "Invention" and "discovery", to use the words of philosopher Pierre-Damien Huyghe [HUY 05], are two stages that meet and reveal the intrinsic qualities of new media. Our research is therefore based in the context of clarifying the original techno-semiotic properties of a still-emerging digital medium, examining what distinguishes it from other media, i.e. how it conveys meaning and emotions via signs and technical formats [JEA 00].

In this chapter, we will attempt to contribute to this characterization of the esthetic of digital media by examining its unique contribution. To better understand the original qualities of 3D interactive interfaces, we will briefly compare them with works from other media, such as painting, photography and film. In particular, we would like to demonstrate how, in digital media, emotion is related to issues of identifying with the machine as an extension and transformation of ourselves. Research into artists has focused not only on the technical properties of media (how they function, their mode of production and reproduction, and distribution) but also on the questions of representing and appealing to the audience. We will examine some of these artworks, which specifically explore the use of emotion in digital media.

We firstly examine some of the distinctions made by Aristotle between experienced and represented emotions and the role of cathartic experiences. Secondly, we will emphasize the role of the device in conveying emotions. We will take the example of photography as it highlights the semiotic process that elicits emotions through the indexical nature of its production. Then we will study creations in digital arts. As such, we will examine three affective modalities in digital devices:

– anthropomorphism using representations of human bodily movements to blur the boundaries between human and machine so as to feel for both;

– real-time interactivity as a mirror of self;

– and empathy as a way to occupy different viewpoints to fully inhabit virtual worlds.

11.2. Emotion, *mimicry* and technical devices

11.2.1. *Representing emotions and catharsis*

Representing emotions and feelings raises difficulties not only in terms of their adequacy for what the artists want us to understand or experience, but also because of the subtlety of affective forms and the expressive power of media. For example, theater, painting, film and now digital forms of expression all explore multiple ways of both representing and provoking emotion in their audiences. We should take a moment to examine this double objective. The art of emotion is both the art of representing emotion and inducing a specific emotional experience.

In *Poetics* [ARI 90], Aristotle distinguishes between feelings, the expression of feelings as they appear "naturally" in real life and the emotions manifested and triggered by fiction. In other words, mimicry – imitation used on the stage and in fiction – should not be confused with daily experience, since the two elicit different emotions. For example, tragedy evokes its own pleasure, "in that which comes with pity and fear through imitation" [ARI 90, p. 10]. The term "catharsis", found in theatre, should also be seen as a means of exercising buried passions – a kind of mechanical emotional purge and a means of representing these passions and gaining critical distance. The two possible translations of *catharsis* as purging or purification convey this subtlety of meaning [DEL 06]. What we would like to highlight here is that these moments of psychological freedom and/or creation of critical distance rely on having a proper space of representation which allows this freedom and distance. It is within the scope of an imaginary space that representations of emotion by *mimesis* take place, allowing us to witness a death without being affected in the same way as if we had really seen it.

11.3. Devices as an alternate source of emotion: photography

Narrating and mimicking comic or tragic situations, staging happiness or sadness contribute to the evocation of an emotion. However, the theme or subject of a piece of art cannot in itself explain the type of emotion that is conveyed. If we take photography as an example, we know that this medium relies both on recognizing a past event and on its evidential value, i.e. the real and non-fictional presence of the photographer with his/her subject [BAR 80]. What moves us is the link to this past time, of which the image is a fragment.

Photography is an art of "vanities" in that it evokes the passage of time and our mortality. For example, Dorothea Lange's photographs are moving when they depict a woman and her children suffering during the Great Depression (specifically, her famous *Migrant Mother*, 1936). In this photograph the mother looks away from the camera; one arm lightly rests on her cheek in an elegant pause, the other holds her baby. She looks like an American Madonna. The subject, both through her features (her thin, drawn face and the tattered clothing) and its religious connotations, is highly evocative. However, it is the making of the portrait which also evokes emotion. The technique itself, the way in which the medium works, has an emotive ability that relies on the production process. In this case, Dorothea Lange had taken a detour to visit a camp of seasonal workers in Nipomo, California and the photograph is evidence of their meeting and coexistence. The perception of the photograph and the emotion that it evokes is based on the image particular semiotics. We will use the term "index" based on Pierce's theory of signs (indexes are signs "whose relation to their objects consists in a correspondence in fact" (W2.56). In Pierce's semiotic typology of modes of meaning [PIE 78], an index joins both the signified and the signifier in one entity:

> "...each of its occurrences is existentially linked to what it indicates; the sign and what it relates to are part of the same existential situation, such as smoke and fire, symptoms and an illness, movement and footsteps" [PIE 78, p. 22].

"*An index is a fragment torn away from the object*". In this one sentence, Daniel Bougnoux points out that the efficiency of an index stems from our very first experience of meaning as children. They are the first signs combined with events with which a child communicates:

> "... during an individual's acculturation, it is these [indexes] which come first, through community and contact (this complex function is what Jakobson has termed 'phatic')" [BOU 91].

This dimension of an index is at the basis of photographic imagery.

> "The first message from a photo is less a representation of the object itself (descriptive function) and more a verification of its existence: its *semiosis* is indicative (there is a physical link between the sign and its cause) or phatic (our way of indicating personal contact with a person in language, whether physical or psychological).".

Photography is therefore a measure of engagement between the artists and their work. There is continuity between the sign (the photo) and reality. Analog media, such as photography and video exploit both the image's shared culture, such as the Virgin Mary and Child in our example, and the acquisition and production of images and therefore rely on a specific semiosis.

It is immediately obvious that digital media do not function in the same way, even if they try to produce realistic representations. It is for this precise reason that artists have attempted to explore the potential of digital devices and their differences with analog media.

11.4. Art and computers: formal beginnings

Fascination with the "method of production" has captured a number of artists interested in the possibilities of computing. The first artists using computers produced work similar to serial art, conceptual art and the minimalist art of the 1960s and 70s by focusing on the exposure of both materiality and concepts. Most modern and contemporary art worked on the autonomy of materials, colors and forms and did not seek to imitate real life, nor to bring any kind of ideological message. Similarly, digital art explores different codes, programs, possible models, images of pixilated screens, etc. and has opened up the possibility of evolution and random permutation. As the art critic Catherine Grenier says [GRE 08], most of these abstract art forms do not seek to evoke emotions. Rather, they teach us to examine digital material, no doubt to remove us from the illusion of photography. A digital image is not an imprint, it is not the testimony of a meeting; rather, it is a calculation based on a model, a construction or reconstruction. The CODeDOC exhibition held by the Whitney Museum in New York is a prime example of this digital art form which focuses on the creative process.

In this exhibition, artists were invited to create digital pieces based on the instruction – which can be taken either literally or metaphorically – to "relate and move three points in space" using several programming languages. Visitors could then see both code sheets and resulting animations. This not only serves an educational purpose, but also takes place in a debate on the nature of the work: does art lie in the performance of the code on the screen or in the writing of the program?

Without entering into this debate, we can see how it is no longer an issue of considering a piece based on its emotional effects. Writing the code with elegance, establishing a successful relationship between computer language and result, and finally, the originality of the graphic animation itself are the focus of attention. To return to Pierce's categories of analysis, the digital work that separates code and its representation functions on a symbolic level. A symbol is "a completely arbitrary sign, graphic symbols for example" [PIE 78, p. 131]. Symbols create distance and reflection:

> "...an index is a pole of attachment and affirmative richness (highly difficult to deny with symbols and icons) and a symbol is the pole of detachment or critical thought" [BOU 91, p. 273].

In the works displayed at CODeDOC, this is a case of linking the pleasure evoked by the animations with an understanding of the process behind it. There is certainly pleasure, but not necessarily emotion (Cicéron, *Doctrine des Officia*).

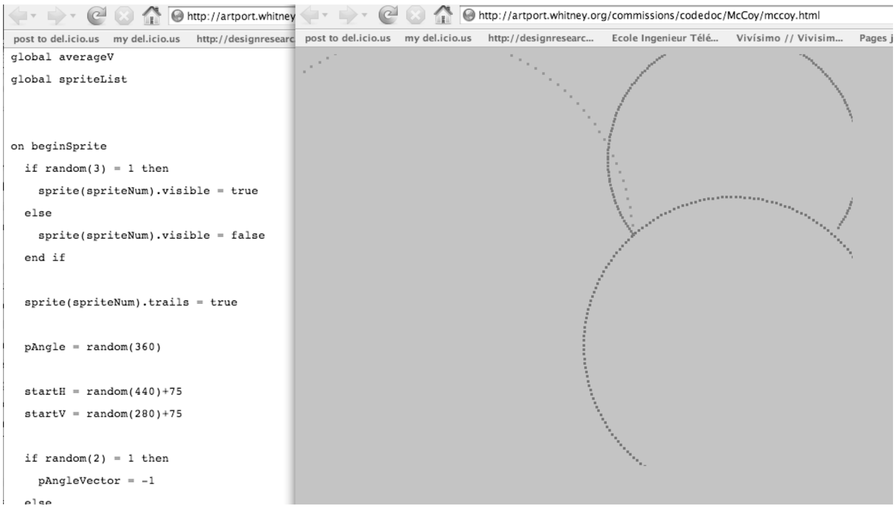


Figure 11.1. *Kevin McCoy, CODeDOC, Whitney Museum, 2003 [COY 03] (source: http://artport.whitney.org/commissions/codedoc/index.shtml)*

This artistic trend does not, however, capture the whole gamut of artistic experiments, as the art critic Anne-Marie Duguet has stated:

> "From the beginning, alongside pieces which are geometrically abstract or inspired by permutational art, appear figurative studies and research into the perception and interpretation of forms, notably thanks to the conversion of images into digital forms. With regard to computer animation, it also includes films by John Whitney which have been entirely generated using computer programs and explore the laws of visual and sound harmonies, or films by Peter Foldès, who uses computers to calculate images between two key images – an interpolation technique with a critical, humoristic and narrative approach, or even animation scenes by John Lasseter, for example." [DUG 02]

In addition to this conceptual approach some artistic experiments tend to reflect less on digital texture and more on our personal relationship with the digital, which can take several forms as the following chapters examine. First, artworks question the limits between human and machine introducing some doubt as to their clear-cut boundaries. Second, works of art play on the indexical quality of interactive interfaces that mimic our own behaviors. Third, in particular in 3D environments, we inhabit not only our own standpoint as in a painting with perspective, but can experience multiple viewpoints. The system uses our capacity for empathy.

11.5. The human behind the mechanics and the mechanics behind the human

Some digital pieces play on our ability to project human qualities onto mechanical objects. Antoine Schmitt[1], a French digital artist, has worked since the beginning of the 1990s on what he considers "algorithmic material"[2], where he explores what can be achieved visually. One of his particularly interesting pieces of work in relation to emotion is the series *Avec détermination* (With determination). This is a collection of 12 pieces: *Standing, Standing 2, Jumping, Not dying, Behaving, Not behaving, Resisting, Stepping, Stepping 2, Pushing, Not moving, Nailed* [SCH 00].

Writing on the subject of this work, Schmitt raises the question of anxiety:

> "Why does it work like that? Perhaps this is the fundamental question. What are the forces behind movement? These movements have a form: which forces for which forms? Some forms have certain qualities (aesthetic, emotional, etc.). What are the qualities of underlying forces? What is their mode of existence? At what level do they act? What is the relationship between the forces within us and those external to us? Where are we positioned among these forces?" [SCH 00]

These pieces use connecting lines and ovals resembling strangely disembodied legs. These forms and features are animated by displacement and force algorithms that are disfigured by random coefficients, producing unexpected changes. Notwithstanding their abstract nature, the programmed movements and their accidents encourage an anthropomorphic interpretation. The series of titles obviously pushes us in this direction, while the choice of red on a black background acts as a distress signal. The significance of these representations is that they rely not on a depiction of distress with realistic facial expressions and context, but on movements that collide with the edges of the "frame" on the screen. The underlying movement models are recognizable but constrained. The feeling of unease produced

1 Graduated from the l'Ecole Nationale Supérieure des Télécommunications in Paris in 1984.
2 From an interview with Sandra Vie, a DEA project on algorithmic art.

by this series of images is a projection of our own difficulties, mistakes and even tragic destinies.

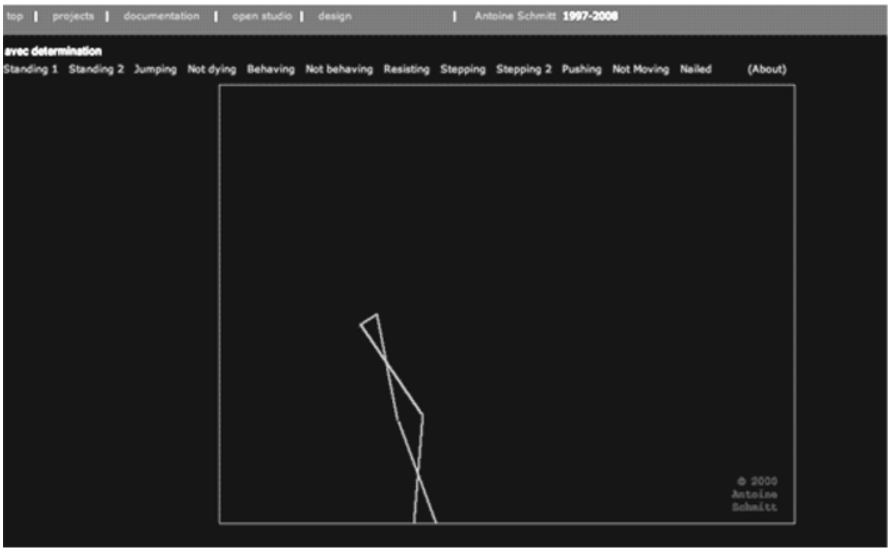


Figure 11.2. *Antoine Schmitt, "Avec détermination" (With determination), September 2000 (source: www.gratin.org/as/avecdetermination/index.html [SCH 00])*

These anthropomorphic interpretations are interesting for two reasons. First, they show that realism is not necessary for pathos. Schematic features, provided that they are animated expressively, are enough to show difficulty, alertness and fatigue because emotion is evoked by mimicking movements and gestures. A comparison with another artist, Turux, who produces similar graphic pieces [TUR 99], shows in contrast a type of art that is not attempting to create a sense of proximity and anthropomorphic projection [GEN 07b]. Turux, as he states on his website, just offers an opportunity for Internet users to play with basic "particles" such as balls, arrows and lines, corresponding to a pure geometrical research on motion. Second, the anthropomorphic projection in Schmitt's work relies on clues. In his work we would not necessarily interpret the movements on the screen as human movement if titles of the art pieces and the presence of articulations indicating the thigh, tibia and foot did not tip us off. The interpretation of Schmitt's work is also therefore guided by textual elements as well as codes and colors.

Anthropomorphism plays a generally poetic and critically important role in this type of artwork [ARM 06]. It is poetic in the sense that is proposes a representation of the relationship between Man and the world. It is also critical because the anthropomorphic forms of representation allow us, indirectly, to approach various political issues (as in La Fontaine's fables).

In effect, the anthropomorphic figure plays on two levels of meaning. First, it relies on the denotative nature of the representation. We see geometric shapes, uncontrolled movement, imperfect models as well as the choreography of simplistic shapes. Second, it holds a surplus of meaning, a depth linking directly back to human experience. The question ultimately posed by *Avec determination* is: are there mechanics behind humans? Or, in the words of Antoine Schmitt, "what relations are there between the forces within us and those outside of us? Where are we among these forces?"

In Antoine Schmitt's work, the role of the spectator is, in contrast, very limited. As she manipulates the mouse, the user provokes counterintuitive effects. When she points upwards, the shafts fall, but when she points down, the shafts rise again. Looking at the screen is contradicted by manipulating it in the same way as the figures' movements on the screen are altered by invisible rules. With emotions such as compassion or embarrassment that "*Avec determination*" can potentially create, feelings of frustration or annoyance at not being able to correctly manipulate the puppets can also arise. This leads us to another dimension of emotion in digital works, linking our gestures with events on the screen, whose mirror-like effect makes us more conscious of our own actions.

11.6. Mirror interaction as an emotional vehicle

The term "interactivity" in informatics spread during the 1980s and refers to dialog between an individual and information provided by the machine where the user enters data to obtain information from the interface. The user's actions therefore have an effect on the machine not only because it provides results, carries out calculations and finds and attaches information, but also because these actions are made visible by the interface. Interactivity is therefore a technical, semiotic (because we see and interpret it) and practical system (we can act based on this visualization) [GEN 10]. Again, the semiotic analysis of interactive systems shows that they rely on indexes of the user's actions. Changes on the screen, and therefore movements, signify the indexical relationship that we have with the screen. There is a relationship of continuity which stems from time (real time) more than space (trace). The device's efficacy lies in the power of suggestion of "reaction" to movements. Using a keyboard interface, mouse or other sensors can construct an image of an intelligent machine but also provide us with a mirror image of ourselves. The "fragment torn away" from reality is a combination of the machine and ourselves.

The emotional force of our relationship with the machine lies in this visible presence of our actions in the machine. The computer screen acts as a mirror to the body. The researcher François Pachet has used this hypothesis of the digital as a reflexive medium to invent musical tools that can reflect musicians' practices. Musical emotion is linked to our collective and personal experiences of sound [CHI 00]. This emotional relationship with music has developed through its growing presence in our daily lives, although this is not specific to the digital. The highly significant hypothesis developed by François Pachet, in the *Continuator*, at the CSL Sony laboratory is that we can develop an emotional relationship with music by receiving a mimetic feedback from our own sound production. François Pachet has attempted to design an interactive musical reflective system that creates a desire to play:

> "…More precisely, we propose to consider the class of interactive systems in which users can interact with virtual copies of themselves, or at least agents that have a mimetic capacity and can evolve in an organic fashion. To make this imitation efficient, there are a number of characteristics that we consider important to define reflexivity in interactive systems." [PAC 06]

His aim is to research what causes pleasure in interactive experiences. Among the most basic elements of this experience, François Pachet examines what he terms the mirroring effect:

> "Similarity or mirroring effect. What the system produces sounds like what the user himself is able to produce. This similarity must be easily recognizable by the user, who must experience the sensation of interacting with a copy of himself. Similarity is not equivalent to mirroring. For instance, a systematic echo or repetition of the phrases played by the user does not induce such a sensation." [PAC 06]

This issue of similarity is important: the machine is not producing an exact copy of what the player has produced. It creates nuances that increase the complexity of the relationship so that the player not only recognizes herself but also has an experience of discovery. The *Continuator* is indeed a mirror, but a vexing one. The motive for this program is therefore not purely narcissistic. The art historian Phay-Vakalis' analysis of the relationship between painting and a mirror better explains what is at work, not only visually but more broadly in this type of sound, texture and image representation [PHA 01]. She takes, for example, Manet's famous work, *Un Bar aux Folies Bergères* (1881-1882). This painting conveys ideas from the early period of photography in which the established myth that painting must imitate nature has lost its hold. In his painting, Manet portrays impossible reflections in the mirror behind the female seller. She appears to be looking at us but, when we examine her back reflected in the mirror, we see that she is looking at a client outside of our field of vision. Gradually the mirror is no longer a trick of the eye but causes us to question what we ourselves are seeing. In the history of art, the mirror is not only represented in paintings. It gets to be embedded in them. The use of mirrors inside the painting allows the viewer to enter into the work itself. The viewer can see her own reflection as in Jacques Monory's 1968 painting *Meutre* (oil on canvas, *bullet impact on mirror*).

> "Since a painting can never match photography exactly, the mirror takes the place of the canvass, and becomes a support and surface. No longer a metaphor of mimesis, but a real work of art, [the mirror] is therefore transformed into an "artistic material" and is gradually confirmed as an "operator of exchange" between the viewers and their environment." [PHA 01, p 133]

François Pachet explores this notion of an "operator of exchange" combining it with notions of rhythm, amplitude and tempo. A musician plays notes on a piano. These notes are interpreted by a program which then proposes a follow up. The system has been tested, both with experienced composers and children. Both groups recognized their own personal "style" in the system's suggestions. In other words, what captivates and what touches in these interactive systems is the personal yet unusual aspect of what we discover about ourselves.

The translation of our input by the machine surprises us because it is partly true to what we have produced but offers us new vision of ourselves. It is therefore not a system designed to test our abilities, nor for aiding success, but to create exploratory environments of ourselves and others. In other words, sensor systems are not only recording and measuring tools; they are perceived as systems that can introduce different forms and proposals that can intrigue, worry, surprise, or even delight by taking and modifying the user's input and returning a "result" that is both loyal to the source and different to what the user expects.

11.7. *Trompe l'œil* versus explicit expression

In interactive digital media there is a strong tension between these two models:

– either they are a reflection, allowing the audience to take a step back and discover the device's mechanisms; or

– they create an imaginary universe without distance that plays on our emotions using an "imitation of the real", specifically in creating perfect synchronization between the user and the machine (essentially the basis for video games).

This *trompe l'œil* utopia, which plays on our desire to go *through the looking glass*, is at the heart of art. The myth of Zeuxis and the birds is often cited as an example of how strong the human desire is to experience imaginary worlds. Pline the Elder records this story of a competition between two painters, Zeuxis and Parrhasius. Zeuxis painted fruits so life-like that birds attempted to eat them. The second painted a curtain that appeared to hide an object. This time, the painter Zeuxis himself was tricked when he tried to lift the curtain to see what it concealed.

> "[Parrhasius] we say, challenged Zeuxis who had painted grapes with such realism that birds tried to peck at them, whilst Zeuxis, full of pride at the birds' judgment, asked him to draw back the curtain to reveal his painting. Recognizing that it was an illusion, he admitted defeat with modest frankness, understanding that whilst he had tricked the birds, Parrhasius had tricked an artist, who was Zeuxis." [PLI 99, XXXV (IV)]

This myth is a reflection of artistic representation but is not confined to the issue of imitating real life. Parrhasius, who prevails, relies on his colleague's curiosity and desire to discover what he cannot see rather than analyze what he does. The painter Zeuxis wants to know what is hidden behind the curtain and therefore no longer only sees the curtain. Lacan has analyzed this scenario to define art as a desire to see beyond forms, materials, colors and lines, i.e. the desire to discover a non-existent object hiding behind a medium of representation [LAC 86]. Bill Viola's film *Reflecting Pool* (1977-1979) literally conveys this desire to submerge oneself in virtual realities. The artist is standing in front of a pool showing his reflection. We can hear the sound of the water and the movement of the trees in the adjacent woods. Suddenly the man takes a running jump into the pool. However, he remains frozen in the air and seems like he could fall into the water at any moment. What follows is a series of tricks of the eye with human shadows that appear on the surface of the water but that do not relate to anyone at the side of the pool. The suspended body above the water gradually disappears. Finally, the man comes out of the water naked. In this work, Bill Viola questions immersive environments by highlighting that they do not project anything significant. A virtual universe is one composed of signifiers that do not relate to any real concepts. The importance of the artist's metaphysical reflection and biblical and religious references highlight that those entering virtual universes seek another version of themselves, a form of rebirth. The psychoanalytical interpretation is that what fuels this desire is the emptiness behind the image.

In *Remediation* [BOL 00], Bolster and Grusin have demonstrated that some media more than others pursue this utopia of being able to make the materiality of objects disappear to give access to imaginary worlds. Digital immersive environments in particular are gradually tending to move away from appliances, objects and devices. The power of emotion relies on our desire to cross over to the other side of the mirror to manipulate "impossible" virtual objects. For example, with avatars of ourselves it seems to be possible to create a double of ourselves and create a new identity that allows us to discover other facets of our personality, as Sherry Turkkle has examined [TUR 97].

Digital artists for the most explore both axes of experience. Emotion is created because the device captures the attention of the user, copies their gestures or anticipates their positions. Then they surprise or even annoy the spectators, by staging media in such a way that makes them aware that the immersion was an illusion. In this way, the artist is hypothesizing about users' expectations or fears when faced with an interactive computer system. These art works rely on people's abilities to read texts and images and give meaning to signs, images and words [SOU 03]. They tap into their *literacy*, i.e. a person's culture or ability to read and interact with the resource. At the same time, they push peoples' ability to give new meaning and open new perspectives. In interactive art works, a number of emotions can be evoked, ranging from annoyance at the slowness of the device's reactions which the user expects to happen in real time (e.g. faced with an interface that needs to be stroked to show its display as in Agnès de Cayeux's *I'm Just Married* [CAY 03]) to fear of a virus (e.g. Jodi's *SOD*, which conveys an interface being destroyed [JOD 99]) to the role of collective online works (e.g. Nicolas Frespech, *Tell me Your Secret* [FRE 01].

The feeling of loneliness, as everyone is alone in front of their computer, is also evoked by some pieces (e.g. Annie Abrahams *Being Human* [ABR 98]) [GEN 07b]. Emotion therefore arises both as much as the subject of works as in how the interface reacts, whether it "behaves" technically and semiotically as we expect, or flouts our expectations. Game designers also use both scenarios of engagement and disengagement though for different reasons. In FPS (*first person shooter*) type games, that mostly rely on immersion, the image and action may give way to decision windows (on the weapon being used), information windows (on the character's energy and their resources) that do not directly contribute to the "flow" but provide a certain amount of distance from the device. The double approach of both absorption and distance corresponds, for the player at least, to a search for moments of dramatic tension (climax) and release (anti-climax).

11.8. Three-dimensional universes: an empathetic experience

Interactive machines elicit a certain amount of emotion. We should, however, distinguish what creates a *remediation of the self*, a specific representation of ourselves, in these 3D environments with avatars and intelligent virtual agents

(IVAs). The work of Catherine Ikam allows us to capture other features of this digital esthetic.

Since the 1980s, with the help of Louis Fléri, Catherine Ikam has produced 3D interactive installations exploring empathy in digital media. She has written, highlighting the importance of synchronization between person and machine, that

> "…the ritual of real time has considerably altered our perception of events. We are entering into the reign of the all-powerful interface: how can we interact with an inanimate object sensitively?" [IKA 99]

She creates installations where faces are projected onto giant screens. These computer-generated "portraits" do not use optics but are constructed from real people using 3D sensors. Faults in the images are not corrected. On the contrary, the method of scanning is highlighted by the presence of black spots where shadows have affected the image processing. Plus the mask displays identical features both on its concave and convex sides.

In 1999, *Elle* was displayed in the *Portrait Réel-Portrait Virtuel* exhibition in *Maison Européenne de la Photographie* in Paris. Portraits were projected onto a screen, although their expressions were indecipherable. Slight movements of the eyes and barely visible smiles are signs of life in the face that floats on the screen. Our movements, or lack thereof, trigger fluctuations in the face's "flight": the face can move forward, backward, turn away or turn back. Our movements "humanize" the face and give the impression of a meaningful relationship, which fascinates us. At the same time, we adapt to the face's reactions and seek a logic or reason for them. We try to establish a relationship between our movement and *Elle*'s movements. Her eyes seem to both follow us and look away while we try to catch her attention. Catherine Ikam emphasizes that:

> "…in some way we try to animate the inanimate by introducing gesture and meaning into human−machine relationships; we are trying to humanize the digital by equally seeking to digitalize the living." [IKA 99]

The work creates a feeling of expectation, a desire for a relationship with the face being displayed. The darkened room, despite the size of the face being projected, creates a feeling of intimacy with *Elle*. However, as in video art, we always remain on the surface of the screen and *Elle* remains far behind this surface when we try to cross it as we come too close to the screen.

In contrast, we understand that our relative positions define us. Each of our perspectives is linked to a movement that alters the piece in some way. This is not like in classical painting with perspective where an ideal point of view defines our place in the composition. We are not in the same position as we are in Bill Viola's work where we are witness to a false/true past. Rather, the work is explored like a landscape. This is essentially the basis of virtual worlds. In *Remediation*, Bolster and Grusin emphasize the fact that the identity of an avatar is constructed by these multiple perspectives:

> "In three dimensional computer graphics, the subject is defined by the perspectives that she occupies in the virtual space. […] but in her quest for immediacy, the subject in virtual space is not satisfied with a single point of view; instead, she seeks out the positions of other participants and objects in that space. She understands herself as a potentially rapid succession of points of view, as a series of immediate experiences derived from those points of view." [BOL 00: 236]

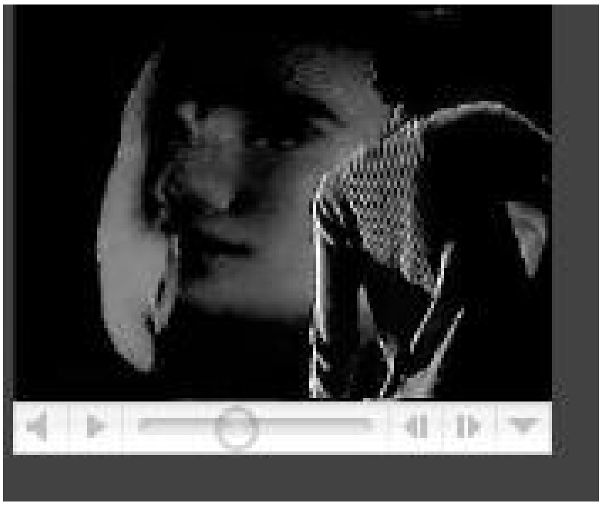


Figure 11.3. *Catherine Ikam, Louis Fléri, Alex, 1996 (source: http://www.ubikam.fr/ [IKA 96])*

This new type of experience contributes to a renewed definition of identity as flexible, fluctuating and fragmented. It also creates a kind of access to knowledge by appropriating successive positions. We can "put ourselves in the place" of someone else, so we can have the same experience and understand what they see and feel. This brings the subject closer to the world that they observe via the intermediary sharing of their point of view. This kind of experience is inherited from cinema. In virtual environments, the avatar functions as a camera controlling our viewpoint. It does not pass via a map or a photograph but occupies a specific space and understands not only what the space is but the perspectives of others within it. This change in standpoints therefore becomes a means of learning and sharing emotions. Note that the metaphor of the camera also applies to the issue of movement. We become a moving camera in a virtual world where all viewpoints are expected and modeled by the creator. We can therefore encounter these viewpoints by moving from one coordinate to another, even exploiting the reverse shot possibilities offered by the device.

We know that video games have successfully benefited from being able to provide low angle, "subjective camera" or more or less broad perspectives. These successful transitions to different points of view constitute the key aspect of our empathetic relationship with technology. What is striking in the work of Catherine Ikam is that she shows that we "listen" to *Elle*, putting ourselves in "her" place.

Based on neurobiological research, Professors Alain Berthoz, Gérard Jorland and Bérangère Thirioux have highlighted that empathy is directly linked to spatial relationships [THI 06]. Research has shown that, far from being an illusion, "the areas of the brain underlying the transition from egocentric reference [from our point of view] to allocentric reference [from the point of view of someone else] and the simulation of an external perspective are different". Empathy is in fact "the ability to change spatial referents". The subject can use allocentric coding, i.e. an external spatial reference, which is independent of his or her own position. This can simulate someone else's perspective.

Thirioux has highlighted that, since the start of the 19th Century, Robert Vischer − the German psychologist and philosopher from whom Lipps, Freud and Husserl have borrowed this notion − introduced the concept of space in a definition of *Einfühlung*, "feeling from within". Vischer proposes a "theorization of empathy to understand the impression−sensation−emotion−content loop". For him, emotions are "resonated" by our feelings, which in turn create a "specific simulation mechanism". "Empathy involves taking others' perspectives of the world into account, a third person perspective". Video games systematically use this capacity to project oneself outside of our own perspective when they make it possible to move from a first person perspective to that of another character in the game.

The specific emotional benefit of virtual worlds and immersive environments seems to be that they not only incorporate others' movement (as in film) but our own movement, providing an allocentric perspective. In effect, it is very different from having an inanimate object that we can interpret (e.g. a painting) but that does not solicit an action (e.g. a painting) and having to adapt our behavior to a living or mechanical system. In this case, the relevance of our reactions increases with our ability to anticipate what the "other one" will do. We therefore need to be attentive to the signs produced by the "system" and put ourselves in its place. This does not mean that interactive environments or 3D forms are more effective than other media at evoking emotions. Rather, these systems achieve this by using our capacity to empathize using our gestures and coordinated reactions on screens. This method of using our abilities to project a perspective of the "other" has an impact on the variety of emotions that we can feel.

11.9. Empathy and identifying emotions

11.9.1. *From empathy to shared emotions*

Empathy is a phenomenon that should not be confused with sympathy. Empathy is the ability to imagine an emotion different from the one we are experiencing ourselves while with sympathy we share the feelings experienced by the person, animal or object. We feel for them. This therefore is a question of contagious emotion.

In the reception of digital works, the separation between empathy and sympathy is not always clear. Putting ourselves in the place of others and changing our viewpoint does not occur without an emotional effect on us. This is already evident with certain devices that encourage attention. For example, the slow motion used in Catherine Ikam's work seems to be fundamental to focus our understanding on our relationship with the system. Our involvement with interfaces, using our gestures and their alterations on the screen, captures our attention. We are not merely analyzing a technical device; the impression of a living system forces us to adopt a maieutic posture, i.e. the use of methods to "deliver" a piece or bring about a situation that reveals the author's intentions, how the device functions and our own performance.

The "behaviors" of objects on a screen are key to our involvement. In his piece, *Les Mains* (Hands), presented at the first Contact Festival, organized by *Le Centre de Création Multimédia*, *Le Cube* at Issy-les-Moulineaux in 2005, Michael Cros exploits empathy and sympathy using an image of hands projected onto a surface with which the viewer can interact. The "virtual" hands felt by our own hands (recorded by a camera) produce new small hands. As explained on Cros' website, (http://www.lesiteducube.com/atelier/michaelcros-lesmains.html):

> "…the virtual hands can be given different behaviors. If they are afraid then they avoid the real hands; if they are gregarious then they will follow the other hands; if they are curious then they will seek contact, offering their palm to be touched to create a small new hand […]. Contact and procreation are accompanied by short sequences of crystalline sounds mixed with low bass."

The users interpret the hands' movement perfectly and touch with the "hands" as they would with a living creature and not a mechanical object.

Figure 11.4. *Michaël Cros, Les mains (Hands), Festival Premiers Contacts, Le Cube, 2003 (http://www.lesiteducube.com/atelier/michaelcros-lesmains.html [CRO 03])*

11.10. Making human−machine interaction and dialog effective

Another level of complexity is achieved when we focus not only on mirror forms but interaction with IVAs. User expectations in relation to IVAs are similar to those of computer-animated cartoons. The 2008 film *Wall-E*, by Pixar, is a perfect example of using scenery and robots' body language to express and evoke love, worry, laughter and nostalgia. Humanoid robots feel emotions, rebel and malfunction (i.e. become crazy). We also expect IVAs to be relevant in the interaction. There must be a minimum amount of equivalency between expected and acted behavior to create an emotional relationship with the IVA. The essential difficulty with this is creating a real-time interaction system. The design must adapt virtual "behaviors" to users' behaviors.

A good number of research projects on IVAs have focused specifically on reducing discrepancies between user expectations and the ability of virtual beings to engage in meaningful dialog. For example, the MyBlog3D[3] project [ANR 07, ANR 10] attempts to increase the expressivity of both avatars and IVAs by recording gestures and facial expressions that are then incorporated into the avatar and dialog models all within a 3D virtual environment (http://www. myblog3D.fr).

Knowing that human−machine interaction and dialog is far from being technically perfect, the interaction designer must not only use IVAs with caution but also imagine and implement situations where the limited interaction with the IVA does not seem to be completely unnatural. As such, IVAs are mostly used to respond to frequently asked questions in short scenes such as welcome pages on websites (e.g. those created by virtuOz). They can also allow web users to simulate their presence when they are inactive on the computer.

In any case, the problem is that of discrepancies in what the user expects of them, a significant and relevant relationship with a given situation, and occasional inadequacy in some interactions. The user's emotional experience relies on this balance between a virtual character and a task they are supposed to complete. A possible strategy involves playing on the potential poetics of a discrepancy between expectations and results. Inadequacy in itself can be a source of reflection. As such, in the ENSAM and LIMSI research projects, it is actually sought for creative purposes [BUI 07].

Frustration, annoyance and anger (if both expressive and linguistic responses are inadequate) can also be interesting responses in an artistic sense, but must be avoided in the framework of service design. Designers must offer situations where behavior is sufficiently simple to create meaning in the interaction with the user. In

3 The ANR project (2007-2009), carried out by the Institut Telecom and e-maginer.

the example developed for Myblog3D, the interaction with the "IVA recruiter" must teach the aspiring candidate how to behave appropriately in a professional interview. The IVA does not adapt to the user, but places the user in a situation where they have to adapt to the IVA to gain a perspective of themselves. The fundamental question here is how do we adjust our speech, posture, intonation, etc., to communicate with the IVA recruiter? This device therefore relies entirely on the "mirror effect" that we have described in the previous examples.

11.11. Conclusion: "revenge of the emotions"

Catherine Grenier, in her analysis of contemporary art, uses the phrase "revenge of the emotions" [GRE 08], highlighting how artists today are reintroducing pathos into their work. This relies primarily on eliciting emotions, as in Proust's *Madeleine*, which evoke special memories. For example, portraits are not necessarily direct and photographic. They are indirect and perceptible in that we recognize ourselves in objects, animals and avatars. This revival of emotions is also visible in scientific and artistic research that questions our relationship with technology. The emotion found in virtual worlds, 3D representations and interactive devices is strongly linked to this recognition of our feeling of "togetherness" with technology and the expression of human behavior in digital media.

If we examine the etymology of the term "emotion", it therefore seems that it is perfectly suited to the analysis of digital esthetics. Indeed, emotion is literally that which creates motion. As we have seen, the interactive esthetic − specifically in 3D worlds, reintroduces an active relationship with space and moves the user. This type of empathy from digital devices elicits emotions, notably when it puts us in a given situation and questions us using this. Research undertaken by Sherry Turkkle into virtual worlds has highlighted the strong emotional dimension created between the user and his or her avatar on the Internet. This relationship is based on the machine-specific indexical semiotics that tests our existence in the technical world. As far as emotions are concerned, the kind of encounter we can expect with IVAs reinforces the feeling of a technical reality, via the dialogical dimension, which concerns us not only in terms of what it can offer us but also its limits, particularly the fact that the imagination found in media still falls short of our desires.

11.11. Bibliography

[ARI 90] ARISTOTLE, *Poétique*, Le Livre de Poche, Paris, 1990.

[ARM 06] ARMENGAUD F., "Anthropomorphisme", *Encyclopaedia Universalis*, Paris, 2006.

[BAR 80] BARTHES R., *La Chambre Claire. Notes sur la Photographie*, Gallimard, Paris, 1980.

[BOL 00] BOLSTER J.D., GRUSIN R., *Remediation. Understanding New Media*, MIT Press, Cambridge, MA, 2000.

[BOU 91] BOUGNOUX D., "L'efficacité iconique", *Nouvelle Revue de Psychanalyse*, pp. 267-280, 1991.

[BUI 07] BUISINE S., AOUSSATA., MARTIN J.C., "Embodied creative agents: a preliminary social-cognitive framework", in: PELACHAUD C., MARTIN J.-C., ANDRÉ E., CHOLLET G., KARPOUZIS K. and PELÉ D. (Eds). *IVA*, *7th International Conference*, pp. 304-316, Springer, Paris, 2007.

[CHI 00] CHION M., *L'Audio-vision*, *Son et Image au Cinéma*, Nathan, Paris, 2000.

[DEL 06] DELAUNAY A., "Catharsis", *Encyclopaedia Universalis*, Paris, 2006.

[DUG 02] DUGUET A.-M., *Déjouer l'Image. Créations Électroniques et Numériques*, Jacqueline Chambon, Paris, 2002.

[EKM 93] EKMAN P., "Facial expression and emotion", *American Psychologist*, pp. 384-392, 1993.

[GEN 07a] GENTES A., "Second Life, une mise en jeu des médias ", in: A. DE CAYEUX *et al.* (Eds), *Second Life, un Monde Possible*, pp. 111-121, Les Petits Matins, Paris, 2007.

[GEN 07b] GENTES A., "L'intime à l'épreuve du réseau", *Communication et Langages*, vol. 152, pp. 89-106, 2007.

[GEN 10] GENTES A., "De la traduction comme médiation et création interlinguistique des interactions humaines", *Revue des Interactions Humaines Médiatisées (RIHM)*, pp. 4-28, 2010.

[GRE 08] GRENIER C., *La Revanche des Émotions*. *Essai sur l'Art Contemporain*, Le Seuil, Paris, 2008.

[HUY 05] HUYGHE P.D., *L'Art au Temps des Appareils*, L'Harmattan, Paris, 2005.

[IKA 99] IKAM C., FLERI L., *Portraits Réel/virtuel*, Editions de la Maison Européenne de la Photographie, Paris, 1999.

[JEA 00] JEANNERET Y., *Y a-t-il (vraiment) des Technologies de l'Information*, Presses Universitaires du Septentrion, Savoirs Mieux, Villeneuve d'Asq, 2000.

[LAC 86] LECAN J., *Le Séminaire*, Livre VII, L'Ethique de la Psychanalyse, Le Seuil, Paris, 1986.

[PAC 06] PACHET F., *Enhancing Individual Creativity with Interactive Musical Reflective Systems, Musical Creativity*, Psychology Press, Hove, 2006.

[PHA 01] PHAY-VAKALIS S., *Le Miroir dans l'Art, de Manet à Richter*, L'Harmattan, Toulouse, 2001.

[PIE 78] PIERCE C.S., *Ecrits sur le Signe*, Le Seuil, Paris, 1978.

[PLI 99] PLINE, *Histoire Naturelle*, Gallimard, Paris, 1999.

[SOU 03] SOUCHIER E., JEANNERET Y., LEMAREC J. (eds), *Lire, Écrire, Récrire. Objets, Signes et Pratiques des Médias Informatisés*, Bibliothèque Publique d'Information, Centre Pompidou, Paris, 2003.

[THI 06] THIRIOUX B., BERTHOZ A., JORLAND G., *L'Empathie comme Physiologie du Changement de Point de Vue: Une Origine Philosophique Vischerienne*, Collège de la Cité des Sciences et de l'Industrie, Paris, 2006.

[TUR 97] TURKLE S., *Life on the Screen. Identity in the Age of the Internet*, Touchstone, New York, 1997.

Websites of the artists referred to in the text

[ABR 98] ABRAHAMS A., *Being Human/Etant Humain*, http://www.bram.org/, 1998.

[CAY 03] DECAYEUX A., *I'm Just Married*, http://www.justm.net, 2003.

[FRE 01] FRESPECH N., *Tell me Your Secret*, http://enbalyon.free.fr/frespech/index.html, 2001.

[IKA 96] IKAM C., *Elle*, http://www.ubikam.fr/, 1996.

[JOD 99] JODI (alias Dirk Paesmans and Joan Heemskerk), *SOD*, http://sod.jodi.org/, 1999.

[SCH 00] SCHMITT A., *Avec Détermination*, http://www.gratin.org/as/avecdetermination/index.html, 2000.

[TUR 99] TURUX (aliasLia), http://www.turux.org/, 1999.

List of Authors

Antonio CAMURRI
InfoMus Lab-Casa Paganini, DIST
University of Genoa
Italy

Alice CAPLIER
GIPSA-lab/DIS
Grenoble INP
France

Marc CAVAZZA
School of Computing
Teesside University
Middlesbrough
UK

Andy CHRISTEN
NEAD
CISA
University of Geneva
Switzerland

Chloé CLAVEL
ICAME-SOAD
EDF R&D
Clamart
France

Emotion-Oriented Systems Edited by Catherine Pelachaud

Géraldine COPPIN
E3 Lab
University of Geneva
Switzerland

Laurence DEVILLERS
LIMSI-CNRS
Orsay
France

Annie GENTÈS
Institut Télécom
Télécom ParisTech
LTCI-CNRS
Paris
France

Donald GLOWINSKI
InfoMus Lab-Casa Paganini, DIST
University of Genoa
Italy

Didier GRANDJEAN
NEAD
CISA
University of Geneva
Switzerland

Sylwia Julia HYNIEWSKA
Institut Télécom
Télécom ParisTech
LTCI-CNRS
Paris
France

Valérie MAFFIOLO
Orange Labs
Lannion
France

Jean-Claude MARTIN
LIMSI-CNRS
Orsay
France

Radosław NIEWIADOMSKI
Institut Télécom
Télécom ParisTech
LTCI-CNRS
Paris
France

Magalie OCHS
CNRS
Télécom ParisTech
LTCI-CNRS
Paris
France

Catherine PELACHAUD
CNRS
Télécom ParisTech
LTCI-CNRS
Paris
France

David PIZZI
School of Computing
Teesside University
Middlesbrough
UK

Gaël RICHARD
Institut Télécom
Télécom ParisTech
LTCI-CNRS
Paris
France

David SANDER
E3 Lab
University of Geneva
Switzerland

Ioana VASILESCU
LIMSI-CNRS
Orsay
France

Index

A

acted emotion, 82-83, 98, 158

action
  tendency, 16, 21, 233
  unit (AU), 110, 125, 177

affective loop, 220

affective neuroscience, 8, 33-35, 40, 48, 60, 63

amygdala, 8-10, 36, 40, 43-53, 57-62

annotation, 84-93, 96-101, 159, 162, 191, 194-200, 205- 208, 222

appraisal, 5, 8-11, 14-21, 48, 51-52, 59, 91- 96, 142, 173, 177-178, 197, 220, 232

arousal, 4-7, 12-13, 19-20, 50, 55, 88, 123-125, 171-172, 175, 197, 224, 250

autonomy, 218, 223, 228-229, 303

C

classification, 86, 92, 97, 112-118, 122- 130, 139-140, 151-164, 192-200, 205- 208, 222-223

cognition, 5-7, 18-20, 39-40, 218

complex emotion, 7, 80-81, 86, 89, 92- 96, 198, 221, 273, 279, 283

conversational agent, 79, 88-99, 108, 169-170, 174, 181, 223, 235

D

descriptor, 88, 140-143, 146-154, 157, 160

discrete emotion, 36, 48, 170-171, 183-184, 196-197

E

emotion, 3- 21, 33-44, 48-52, 55-60, 64, 79-89, 94-100, 107-108, 115, 118, 121-130, 139-141, 147-152, 155-156, 159-162, 169-183, 191-208, 218-238, 248-249, 252-255, 260-263, 272-274, 278-279, 283-284, 287, 292, 295, 299-301, 304-308, 311, 314-315, 318
  induction, 81-83, 100

emotional
  corpus, 99, 204
  intelligence, 217

Emotion-Oriented Systems Edited by Catherine Pelachaud

reaction, 18, 55, 121, 220, 232, 248-254

state, 5, 10, 16, 55, 80, 83-85, 92-96, 101, 107-108, 122, 128-129, 139, 144, 159, 170-176, 179-184, 192-195, 198-199, 203-204, 229-230, 237, 250, 274, 282-284, 291

empathy, 226-228, 235-236, 272, 300, 305, 312-315, 318

engagement, 219-220, 223, 230, 236, 260, 302, 311

expressive gesture, 255-259, 263

extraction, 112, 115-116, 140, 222

F, I

Facial Action Coding System (FACS), 109-113, 118-121, 124-125, 176, 222

imitation, 301, 308-309

L, M

learning algorithm, 97-98, 118, 142, 154, 160

micro-expression, 177-178, 182-184

mimesis, 301, 309

motivation, 274

multimodality, 128-131, 181, 253

P, R

perception, 4, 35-36, 40, 44, 48, 59, 80, 82, 85, 93-94, 98-99, 115, 128, 159, 191-195, 198-208, 220, 223, 238, 249, 253- 255, 302-304, 312

prefrontal cortex, 43-44, 50, 53-57

recognition, 6, 10, 36-37, 48-49, 59, 89, 107-108, 111-122, 125-131, 139-143, 146-149, 151-152, 155-162, 191-194, 202-208, 218, 221-225, 230, 237, 249, 252-254, 284, 318

regulation, 36, 42, 56, 57, 173

S, V

social interaction, 263

spontaneous emotion, 100, 121, 124, 128, 143, 158, 221

valence, 12-13, 19-21, 49-50, 53, 57-59, 90, 94-96, 123, 158, 171-178, 184, 196-197, 224-225, 250