HAL
open science

Modélisation et détection des émotions à partir de données expressives et contextuelles

Franck Berthelon

To cite this version:

Franck Berthelon. Modélisation et détection des émotions à partir de données expressives et contextuelles. Autre [cs.OH]. Université Nice Sophia Antipolis, 2013. Français. NNT : 2013NICE4151 . tel-00917416v2

HAL Id: tel-00917416

https://theses.hal.science/tel-00917416v2

Submitted on 6 Feb 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Université de Nice Sophia-Antipolis

THÈSE

pour obtenir le grade de

Docteur en sciences de l'Université de Nice Sophia-Antipolis

Spécialité : **Informatique**

préparée au laboratoire **I3S**

dans le cadre de l'École Doctorale **EDSTIC**

présentée et soutenue publiquement
par

Franck Berthelon

le 16/12/2013

Titre:
Modélisation et détection des émotions à partir de données expressives et contextuelles

Directeur de thèse: **Peter SANDER**

Jury

M. Frank FERRIE,	Rapporteur
M. Claude FRASSON,	Rapporteur
M. Ladjel BELLATRECHE,	Examinateur
M. Nhan LE THANH,	Examinateur
M. Peter SANDER,	Directeur de thèse

Remerciements

L'élaboration de ce document n'aurait pu se réaliser sans l'aide de nombreuses personnes que je souhaite tout particulièrement remercier par le biais de ces quelques lignes.

En premier lieu, j'adresse mes plus sincères remerciements à mon directeur de thèse Peter Sander. En plus d'avoir accepté d'encadrer ma thèse, il a toujours su m'encourager et m'orienter dans les moments difficiles. Grâce à son soutien et sa confiance inconditionnelle, j'ai pu progresser toutes ces années afin d'arriver au terme de mes travaux. Il restera pour toute ma carrière un exemple à suivre.

J'exprime tous mes remerciements à Frank Ferrie et Claude Frasson qui ont eu la gentillesse de rapporter mon manuscrit de thèse.

Je remercie toute mon équipe de recherche qui m'a acceuilli et fourni un cadre de travail idéal pour mener à bien ma recherche.

Je remercie chaleureusement mes amis Nadia Cerezo, Khalil Riad Bouzidi, Viet Hoang Vu et Vincent Caraccio avec qui j'ai partagé mes doutes, mes échecs, mais aussi mes réussites. Sans eux, je n'aurais jamais pu achever ce travail. Je pense également à tous ceux que je n'ai pas cités, et qui m'ont encouragé pendant ces années.

Je remercie enfin, toute ma famille et plus particulièrement mes parents Henri et Patricia Berthelon, ainsi que mon amie Cyrielle Claret, qui n'ont jamais douté de mes capacités à réussir et m'ont toujours apporté soutien et réconfort.

J'adresse un remerciement tout particulier à mon père qui, il y a 21 ans de cela, a initié ma passion pour l'informatique.

Résumé

Nous proposons un modèle informatique pour la détection des émotions basé sur le comportement humain. Pour ce travail, nous utilisons la théorie des deux facteurs de Schachter et Singer pour reproduire dans notre architecture le comportement naturel en utilisant à la fois des données expressives et contextuelles. Nous concentrons nos efforts sur l'interprétation d'expressions en introduisant les Cartes Emotionnelles Personnalisées (CEPs) et sur la contextualisation des émotions via une ontologie du contexte émotionnel(EmOCA). Les CEPs sont motivées par le modèle complexe de Scherer et représentent les émotions déterminées par de multiple capteurs. Les CEPs sont calibrées individuellement, puis un algorithme de régression les utilises pour définir le ressenti émotionnel à partir des mesures des expressions corporelles. L'objectif de cette architecture est de séparer l'interprétation de la capture des expressions, afin de faciliter le choix des capteurs. De plus, les CEPs peuvent aussi être utilisées pour la synthétisation des expressions émotionnelles.

EmOCA utilise le contexte pour simuler la modulation cognitive et pondérer l'émotion prédite. Nous utilisons pour cela un outil de raisonnement interopérable, une ontologie, nous permettant de décrire et de raisonner sur les philies et phobies pour pondérer l'émotion calculée à partir des expressions. Nous présentons également un prototype utilisant les expressions faciales pour évaluer la reconnaissance des motions en temps réel à partir de séquences vidéos. De plus, nous avons pu remarquer que le système décrit une sorte d'hystérésis lors du changement émotionnel comme suggéré par Scherer pour son modèle psychologique.

Abstract

We present a computational model for emotion detection based on human behavioural expression. For this work, we use the two-factor theory of Schachter and Singer to map our architecture onto natural behavior, using both expressive and contextual data to build our emotion detector. We focus our effort on expression interpretation by introducing Personalized Emotion Maps (PEMs), and on emotion contextualisation via an Emotion Ontology for Contex Awareness (EmOCA). PEMs are motivated by Scherer's complex system model of emotions and represent emotion values determined from multiple sensors. PEMs are calibrated to individuals, then a regression algorithm uses individual-specific PEMs to determine a person's emotional feeling from sensor measurements of their bodily expressions. The aim of this architecture is to dissociate expression interpretation from sensor measurements, thus allowing flexibility in the choice of sensors. Moreover, PEMs can also be used in facial expression synthesis.

EmOCA brings context into the emotion-modulating cognitive input to weight predicted emotion. We use a well known interoperable reasoning tool, an ontology, allowing us to describe and to reason about philia and phobia in order to modulate emotion determined from expression. We present a prototype using facial expressions to evaluate emotion recognition from real-time video sequences. Moreover, we note that, interestingly, the system detects the sort of hysteresis phenomenon in changing emotion state as suggested by Scherer's psychological model.

Table des matières

Remerciements iii

Résumé v

Abstract vi

Table des matières vii

Table des figures xi

Liste des tableaux xv

Introduction **1**

1 Les interactions Homme-Machine et les émotions 1

2 Contexte applicatif 3

3 Fondements théoriques 4

4 Axes de recherche 7

5 Principaux résultats 8

6 Structure du manuscrit 10

1 L'*informatique affective*, une nouvelle interaction homme-machine **13**

1 Introduction 13

2 L'émotion, une notion complexe 14

2.1 Vers une définition commune 14

2.2 Caractéristiques des émotions 16

2.3 Positionnement 17

3 Théories sur l'activation de l'émotion 18

3.1 L'approche psycho-évolutionniste 18

3.2 L'approche périphéraliste 19

3.3 L'approche béhavioriste 20

3.4 L'approche cognitiviste 21
3.5 Positionnement 24
4 Modèles de description 25
4.1 Les modèles catégoriels 25
4.2 Les modèles dimensionnels 27
4.3 Ontologies des émotions 29
4.4 Modèle de Scherer 33
4.5 EmotionML 34
4.6 Positionnement 35
5 Expression et captation des émotions 36
5.1 Catégorisation des expressions 37
5.2 Les réponses physiologiques 37
5.3 Les réponses psychologiques 42
5.4 Les réponses comportementales 44
5.5 Positionnement 45
6 Observation et validation des émotions 45
6.1 Ellicitation des émotions 46
6.2 Identification de l'émotion 46
7 Les détecteurs émotionnels 47
7.1 Présentation d'une partie des détecteurs existants dans le domaine 48
7.2 Positionnement 50
8 Conclusion 51

2 Un modèle en trois couches pour la détection des émotions 53
1 De la théorie des deux facteurs au modèle informatique 55
1.1 Architecture générale 56
2 Présentation du prototype 61
2.1 Module de capture des expressions faciales 62
2.2 Module de génération des Carte Emotionnelle Personnalisées (CEPs) 64
2.3 Module de catégorisation et de contextualisation 66

2.4 Module de synthétisation des expressions faciales 67
3 Conclusion 69

3 Les variations émotionnelles sous l'angle des sytèmes complexes 71
1 Les Cartes Émotionnelles Personnalisées 72
1.1 Utilisation et objectifs 72
1.2 Modèle dynamique 73
1.3 Apprentissage et annotation des états émotionnels 76
2 Reconnaissance des émotions 78
3 Apprentissage automatique 82
4 Algorithme de génération d'expressions émotionnelles 83
5 Conclusion 84

4 Catégorisation et contextualisation par une ontologie 87
1 Contexte émotionnel 88
2 Choix technologique 89
3 Modélisation de EmOCA 90
3.1 Modélisation de la contextualisation 90
3.2 Modélisation de la catégorisation 92
4 Exploitation de la base de connaissances 93
4.1 Implémentation 93
4.2 Interrogation de la base de connaissances 96
5 Conclusion 100

5 Résultats et discussion 103
1 Evaluation des performances de détection émotionnelle 104
1.1 Méthodologie 104
1.2 Résultats 109
1.3 Discussion 114
2 Contextualisation des annotations 114
2.1 Hypothèse initiale 116
2.2 Résultat 116

2.3 Discussion 120
3 Observation de l'hystérésis 121
3.1 Observation 121
3.2 Discussion 123
4 Synthetisation des expressions émotionnelles 123
4.1 Méthodologie 124
4.2 Résultats 125
4.3 Discussion 127
5 Conclusion 128

Conclusion **131**

Bibliographie **137**

A Algorithme de détection des émotions **145**

B Ontologie des émotions **147**

C Requête SPARQL de catégorisation et de contextualisation des annotations brutes **151**

D Les six émotions de Ekman **155**

Table des figures

1 Application de l'hystérésis pour les émotions "calme" et " colère". . . 5

2 Modèle des émotions de Scherer basé sur la théorie des catastrophes et l'hystérésis. 6

1.1 Comparaison des architectures simplifiées des détecteurs existants par rapport à notre proposition 14

1.2 Représentation graphique de la théorie psycho-évolutionniste 19

1.3 Représentation graphique de la théorie périphéraliste 20

1.4 Représentation graphique de la théorie de Cannon-Bard 21

1.5 Représentation graphique de la théorie des deux facteurs 22

1.6 Représentation graphique de la théorie Component Process Model (CPM) 24

1.7 Les 6 émotions basiques de Ekman. Le lecteur est invité à identifier les émotions exprimées ; réponses dans l'annexe D 26

1.8 La roue des émotions de Plutchik 27

1.9 Modèle bi-dimensionnel de Russell 28

1.10 Ontologie des émotions du groupe GRIHO 31

1.11 Ontologie des émotions HEO 32

1.12 Ontologie des émotions Emotion Ontology (EMO) 33

1.13 Oculomètre tobii 41

1.14 Extrait de la base de données des unités d'actions de Kanade [Kanade 00] 43

1.15 Représentation des trois échelles du questionnaire Self-Assessment Manikin (SAM) 47

2.1 Présentation des différentes sources d'informations émotionnelles à partir de la théorie des deux facteurs. 54

2.2 Vue d'ensemble de notre modèle en trois couches 54
2.3 Architecture globale du système de détection des émotions 57
2.4 Détail du module de traitement du signal 58
2.5 Détail du module de CEP 60
2.6 Détail du module de catégorisation et contextualisation 61
2.7 Interface graphique du module de capture des expressions faciales . . 62
2.8 Extraction des contours de la bouche 63
2.9 Interface graphique du module de génération et d'exploitation des CEPs. Ici, nous pouvons observer une projection sur la commisure droite de la bouche et le sourcil gauche de la carte d'éveil. Le dégradé de couleur représente l'intensité de l'éveil avec en bleu un état calme et en rouge un état éveillé. 64
2.10 Interface graphique du module de catégorisation et de contextualisation des émotions 67
2.11 Interface graphique du module de synthétisation des expressions faciales 68

3.1 Fonctionnement global des CEPs 75
3.2 Représentation des séquences contenues dans une CEP expérimentale 77
3.3 Schématisation de l'algorithme K-nearest neighbors (KNN). Les classes A et B pouvant représenter deux émotions distinctes. 79
3.4 Schématisation de l'algorithme KNN modifié 80
3.5 Schématisation du processus de détection des émotions dans les CEPs 81
3.6 Schématisation du processus d'apprentissage automatique 83
3.7 Principe général de fonctionnement de la synthétisation des émotions 84

4.1 Logique d'activation d'un trait de caractère à partir d'une base de connaissances 90
4.2 Architecture globale de Emotion Ontology for Context Awareness (EmOCA) 91
4.3 Utilisation des coordonnées cartésiennes et sphériques sur le modèle de Russell 93
4.4 Schéma global de l'exécution de la requête SPARQL Protocol And RDF Query (SPARQL) 96
4.5 Détermination du trait de caractère activé 97

4.6 Pondération de l'annotation émotionnelle brute 98

4.7 Sélection du (ou des) terme(s) émotionnel(s) compatible(s) avec la pondération 100

5.1 Images extraites de la vidéo d'apprentissage représentant les 6 expressions faciales de Ekman 105

5.2 Images extraites de la vidéo d'apprentissage représentant les variations d'intensité pour la joie 106

5.3 Répartition des annotations pour une séquence représentant l'émotion joie 107

5.4 CEPs de valence et d'éveil pour les expressions faciales 107

5.5 Récapitulatif de la méthodologie de test 109

5.6 Déviation pour $K = 1$ 110

5.7 Déviation pour $K = 6$ 110

5.8 Déviation pour $K = 15$ 111

5.9 Déviation moyenne en fonction de la valeur de K 111

5.10 Annotation automatique d'une séquence vidéo regroupant trois expressions de la joie 112

5.11 Annotation automatique d'une séquence vidéo regroupant trois expressions de la colère 113

5.12 Annotation automatique d'une séquence vidéo regroupant trois expressions de la colère 113

5.13 Présentation des deux phases du questionnaire 115

5.14 Capacité de reconnaissance des émotions de nos participants 117

5.15 Différence d'annotation en fonction du contexte pour les deux groupes 118

5.16 Résultats globaux de l'impact du contexte émotionnel sur l'identification des émotions 119

5.17 Pourcentage d'accord de nos participants en fonction du contexte . . 120

5.18 Observation d'une hystérésis lors de la détection de la joie sur les deux axes du modèle de Russell 122

5.19 Observation du chemin de détection sur une vidéo symétrique exprimant la joie 123

5.20 Méthodologie d'évaluation des performances de l'algorithme de synthétisation des expressions faciales 124

5.21 Tracés des deux capteurs de sourcils 125
5.22 Tracés des deux capteurs de la bouche gauche et droite 126
5.23 Tracés des deux capteurs de la bouche haut et bas 127

D.1 Emotions correspondant aux six émotions d'Ekman 155

Liste des tableaux

1.1 Tableau récapitulatif des liens entre fonctions, sous-sytèmes organiques et composants émotionnels [Scherer 05] 23

1.2 Résumé des émotions basiques [Belzung 07] 26

1.3 Tableau comparatif d'un panel représentatif des détecteurs d'émotions 49

5.1 Tableau récapitulatif des vidéos utilisées pour l'analyse des performances 106

Introduction

1 Les interactions Homme-Machine et les émotions

Depuis la première interface graphique le *SketchPad* [Sutherland 64] du début des années 60, les interactions homme-machine tiennent une place prépondérante dans l'évolution de l'informatique. Nos systèmes de plus en plus complexes nous obligent à réfléchir aux interfaces de communication avec ce monde virtuel. Deux voies de réfléxions ont été envisagées, d'une part l'Interaction Homme-Machine (IHM) systémique où l'interface est étudiée dans sa globalité et d'autre part l'IHM "à pointe fine" qui se concentre sur les nouvelles méthodes d'interaction.

Cette seconde direction, en forte croissance ces dernières années, a produit des interactions d'un nouveau genre. Parmi les exemples les plus démocratisés, nous pouvons citer les écrans tactiles qui révolutionnent nos smartphones et tablettes, la *wii* [1] de Nintendo, première console guidée par des mouvements naturels au travers de la *wiimote*, ainsi que le module *kinect* [2] de Microsoft permettant d'interpréter les postures humaines comme commande pour les consoles ou ordinateurs.

Outre ces produits déjà commercialisés, la recherche continue son avancée vers des axes très variés. Récemment, Google a présenté ses travaux sur les *Google Glass* [3] basés sur la réalité augmentée et le contrôle par oculométrie de cette interface. Nous pouvons également citer les travaux de Pranav Mistry du MIT avec *Sixthsense* [Mistry 09], proposant une interface portable s'adaptant au monde physique et contrôlable par gestes. Ou encore, les systèmes Brain Computer Interface (BCI), fréquemment basés sur l'Électroencéphalographie (EEG) qui sont souvent utilisés pour les personnes à fort handicap. Ce type de périphérique permet d'obtenir des temps de réponse très courts voire même d'anticiper certains mouvements de l'utilisateur [Wessberg 00] en laissant les membres libres. Néanmoins, les systèmes les

[1] http://fr.wikipedia.org/wiki/Wii
[2] http://fr.wikipedia.org/wiki/Kinect
[3] http://www.google.com/glass/start/

plus avancés utilisent des capteurs très invasifs (IRMf, électrodes intra-craniennes) malgré l'existence des casques EEG non invasifs mais moins performants.

Enfin, le domaine que nous visons pour nos recherches est celui de l'*informatique affective* [Picard 95]. Cet axe de l'IHM "à pointe fine" a pour but d'intégrer les émotions comme nouvelle interaction entre l'homme et la machine. Quotidiennement, dans nos interactions homme à homme, les émotions tiennent une part importante dans notre mode de communication [Goleman 99]. Bien que cette interaction soit non-verbale, nous sommes de formidables détecteurs émotionnels, ce qui nous permet de moduler notre communication en fonction du ressenti de notre interlocuteur. Fournir aux ordinateurs la capacité d'interpréter ou de simuler les émotions peut améliorer notre manière de communiquer avec les systèmes informatiques. Le domaine se divise donc en deux parties possédant des bases théoriques communes. D'un côté, la génération d'avatars possédant la capacité d'exprimer des émotions essentiellement au travers d'un visage virtuel [4] et, d'un autre côté, la détection des émotions. Cette deuxième facette du domaine est fortement liée à la notion d'apprentissage au sens large c'est à dire l'acquisition de nouvelles connaissances [Jraidi 12], mais également l'amélioration de certaines aptitudes. Les pilotes d'avions peuvent être un bon exemple de détection des émotions. Dans les phases de simulation de crash aérien, il est intéressant d'identifier des émotions très intenses pouvant modifier la prise de décision du pilote et donc de lui faire commettre des erreurs.

La notion d'émotion est très familière, nous avons tous la capacité de la définir. Cependant, cette notion est plus complexe que ce qu'il paraît. Les définitions intuitives que nous pouvons proposer ne permettent pas de cerner le sujet de manière précise. De plus, de nombreuses ambiguïtés existent. En particulier, sur des termes tels que : affect, humeur ou sentiment. Les liens entre ces concepts ainsi que leurs limites ne sont pas évidents. La culture joue également un rôle important dans la compréhension de l'émotion. Par exemple, l'"ijirashii" est un terme japonais définissant l'émotion ressentie lorsque l'on voit une personne louable franchir une difficulté. C'est pourquoi nous devons définir précisément le type d'émotion que nous traitons dans nos travaux. Pour cela, nous pouvons nous baser sur des études effectuées sur les définitions [Kleinginna 81] ou encore sur les définitions du réseau d'excellence HUMAINE [5][Cowie 05]. Nous pouvons définir l'émotion comme une réponse à la fois corporelle (neurologique, physiologique, ...) et cognitive (sociale, mémoire, connaissances, ...), rapide, déclenchée par un évènement extérieur (évènement, objet) ou intérieur (souvenir), qui capte l'attention de la personne.

[4] http://www.cam.ac.uk/research/news/face-of-the-future-rears-its-head
[5] http://emotion-research.net/

2 Contexte applicatif

L'utilisation d'un détecteur d'émotion dans les logiciels est très variable. Il existe de nombreuses applications pouvant profiter d'un tel outil et ceci afin de modifier dynamiquement le comportement de l'application ou de récolter des informations sur l'état de l'utilisateur. Par exemple, dans le domaine des jeux vidéo la détection des émotions commence à atteindre le grand public. Microsoft propose au travers de la XBox One une analyse du rythme cardiaque pour l'identification d'émotions. Ce type d'application peut permettre de modifier l'environnement virtuel du jeu en fonction des émotions (modification de la météo, du comportement des avatars, difficulté, ...) mais aussi pouvoir identifier la validité d'un scénario en déterminant les émotions ressenties chez l'utilisateur. D'autres contextes applicatifs peuvent être imaginés comme dans le domaine médical pour identifier la dépression ou des troubles liés au stress. Dans le domaine de l'apprentissage également, nous pouvons imaginer l'emploi d'un tel détecteur afin d'améliorer la productivité de l'apprentissage et obtenir des informations sur l'état d'écoute de l'auditoire. Tous les domaines que nous venons de citer imposent leur propres contraintes. Bien que le sujet reste la détection des émotions, le choix des capteurs, l'environnement des utilisateurs et les informations pertinentes à identifier sont spécifiques. Cependant, nous essayons de montrer comment fournir un système adaptable à ces différentes contraintes en proposant une modélisation commune et une implémentation adaptée au contexte applicatif.

Dans le cadre de nos travaux, nous avons sélectionné le type d'émotion en fonction du contexte applicatif de notre réalisation. En effet, dans le cadre des IHM, il existe un grand nombre de situations à forte charge émotionnelle. Cependant, toutes ne sont pas facilement reproductibles dans le cadre expérimental. Les situations de frustration ou de stress des utilisateurs face à des situations d'incompréhension sont légion, mais limitent le panel d'émotions analysables. Une situation courante en informatique où la charge émotionnelle peut être très importante est le jeu vidéo. En effet, les jeux vidéo sont des logiciels conçus pour optimiser l'immersion de l'utilisateur dans l'environnement virtuel. Que ce soit l'univers dans lequel il évolue ou le personnage qu'il incarne, le joueur peut ressentir des émotions par un phénomène d'empathie au travers des avatars qu'il incarne ou qu'il rencontre. Il est alors possible de susciter un éventail d'émotions plus variées en fonction des évènements proposés au joueur. De plus, les jeux vidéos sont de plus en plus multi-joueurs et communautaires, ce qui augmente encore la capacité du jeu à être un vecteur émotionnel fort.

Le processus d'apprentissage est une autre situation où l'émotion joue un rôle important. Maintenir l'apprenant dans un état émotionnel adapté peut améliorer

l'assimilation des informations. Des travaux [Chaouachi 12] montrent que certains rythmes cérébraux liés aux émotions permettent de distinguer les phases où l'apprenant est capable d'assimiler l'information, des phases où il ne l'est pas.

Pour notre réalisation, nous avons sélectionné un autre type d'application combinant ces deux situations, à savoir les "Serious Games". Une définition communément admise dans le domaine est la suivante :

> "serious games, that is, (digital) games used for purposes other than mere entertainment" [Susi 07]

Bien qu'il existe des "Serious Games" visant des objectifs marketing ou publicitaire, une grande partie de ces jeux ont pour vocation l'apprentissage. Dans cette catégorie, nous retrouvons les simulations (simulateur de vol pour les pilotes de ligne) ou les jeux ludo-éducatifs.

Ces applications ont la double particularité d'avoir une forte capacité d'immersion provoquant des émotions importantes, ainsi que la nécessité d'un retour sur l'état émotionnel du joueur.

C'est en partant de ce contexte, ainsi que des contraintes que cela implique, que nous avons conçu notre modèle de détection des émotions. C'est pourquoi nous proposons de détecter les émotions à court terme, via des capteurs à faible coût, dans le cadre des systèmes multi-modaux, ceci afin de s'adapter aux contraintes dues à l'hétérogénité des situations rencontrées dans les "Serious Games" à but d'apprentissage.

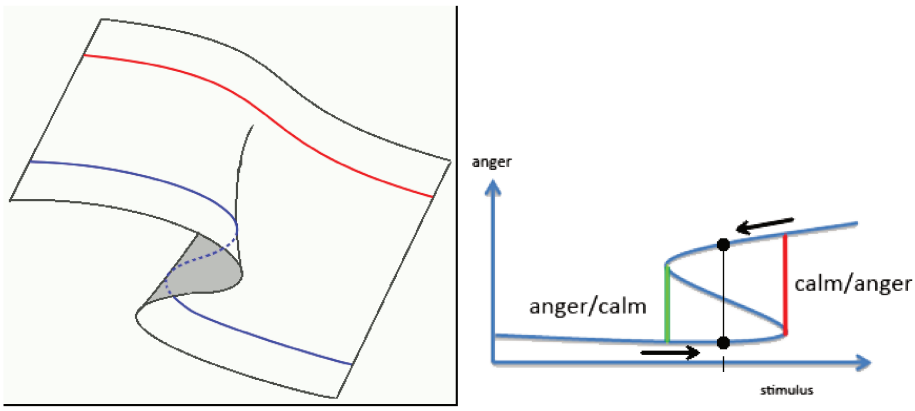
3 Fondements théoriques

Les fondements théoriques de notre réflexion se portent sur les travaux de Schachter et Singer concernant le processus émotionnel, ainsi que les travaux de Scherer sur la modélisation des émotions au travers des systèmes complexes.

La théorie cognitive des deux facteurs, proposée par Schachter et Singer, est l'une des premières théories à intégrer le contexte et les connaissances dans le processus de déclenchement émotionnel. Cette approche se décompose en deux parties. Premièrement, la phase de perception du stimulus, permettant de déclencher les réponses réflexes adaptées à la situation rencontrée, engendrant ainsi un état dit d'hyper-éveil, suivi de la seconde phase d'interprétation des réponses émotionnelles, associées au contexte et aux connaissances de l'individu, permettant d'identifier l'émotion ressentie. Ce courant de pensée s'est vu développer au cours des années par les travaux de Schachter [Schachter 79] et de Philippot [Philippot 01], proposant

tous deux des théories mettant en avant l'évaluation et les connaissances de la personne expérimentant une émotion. La facette cognitive de l'émotion se base sur des connaissances tangibles de l'utilisateur. Nous proposons de concevoir un système capable de raisonnner sur ces dernières, afin d'identifier de manière précise, le ressenti de l'individu. Nous utiliserons pour cela une ontologie nous permettant d'effectuer ce travail.

La seconde théorie fondatrice de notre travail est proposée par Klaus R. Scherer et son modèle des émotions basé sur les systèmes complexes. Dans le cadre des émotions, un changement continu du stimulus peut engendrer un comportement discontinu de notre système. Ces discontinuitées sont appelées *catastrophe* par Thom [Thom 72]. Un système ne subissant pas de changement tendra vers un état stable. Si l'on provoque un changement, le système essaiera de conserver son état initial jusqu'au point de rupture, où le changement exercé est trop important pour atteindre un nouvel état stable. Pour Scherer, une des catastrophes fait écho avec les émotions. L'hystérésis permet de modéliser l'hypothèse suivante : le passage d'une émotion A à une émotion B en fonction d'un stimulus continu se produira à deux instants différents pour le passage de A vers B ou de B vers A. Cette catastrophe permet alors de définir un état futur en fonction des états précédents. Dans le cas de l'hystérésis, deux états sont possibles pour un même stimulus, la sélection de l'état adapté sera effectué grâce à la mémoire de l'évolution du système observé. La figure [Fig. 1b] présente un exemple d'application de l'hystérésis pour le passage de l'état calme vers la colère, en rouge, et de la colère au calme, en vert.



(a) Hysteresis

(b) Hysteresis appliqué au changement Calme/Colère

FIGURE 1 – Application de l'hystérésis pour les émotions "calme" et " colère".

En partant de cette hypothèse, Scherer propose une représentation des émotions comme un système complexe [Scherer 00, Scherer 09] à partir de l'hysteresis [Fig. 2]. Ce modèle n'est qu'une proposition où l'hysteresis est le centre de la modélisation.

Cette surface émotionnelle est disposée dans un espace tri-dimensionnel adapté à la théorie de René Tom, à savoir un espace dit de contrôle à 2 dimensions définissant les composantes de l'émotion (puissance et obstruction) et une dimension "résultat" caractérisant les changements. Sur cette surface, les émotions sont disposées en fonction de leur définition dans l'espace de contrôle.

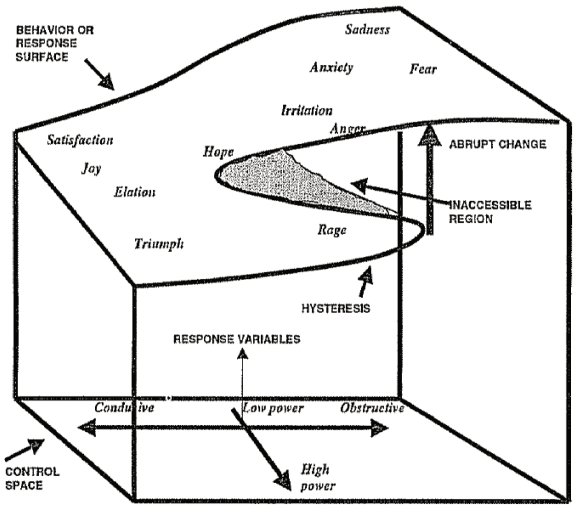


FIGURE 2 – Modèle des émotions de Scherer basé sur la théorie des catastrophes et l'hystérésis.

Enfin, dans le domaine de la détection des émotions deux approches cohabitent, d'une part, les systèmes unimodaux centrés sur une unique source de données, comme le texte, et les systèmes multimodaux, combinant les informations de plusieurs capteurs, souvent physiologiques, comme le rythme cardiaque et la sudation cutanée. Bien que les systèmes unimodaux trouvent leur intêret dans des contextes applicatifs précis tels que la recherche d'émotion dans les textes ou dans des enregistrements téléphoniques, notre hypothèse est que la qualité de la détection est directement liée à la variété des modalités prises en compte. David Marr définit une catégorisation des expressions émotionnelles [Marr 82]. Parmi ces catégories, les expressions dites physiologiques (rythme cardiaque, sudation cutanée, ...) sont plus difficilement falsifiables que les expressions psychologiques (expressions faciales, vocales, ...) ou comportementales (postures). De plus, Cannon [Cannon 27] a montré la non-spécificité des réponses émotionnelles. Par exemple, l'augmentation du rythme cardiaque se retrouve dans l'expression de plusieurs émotions, telles que la joie, la peur ou la colère.

4 Axes de recherche

La détection des émotions est une tâche difficile. Afin de comprendre cette difficulté, nous pouvons faire l'analogie avec la classification dans l'imagerie terrestre. En effet, lorsque l'on souhaite identifier certains objets à partir d'images satellitaires, il est facile de récolter un ensemble de données représentant cet objet, de les comparer et ensuite d'aller vérifier sur le terrain la pertinence du résultat. Ici, l'émotion est une information impalpable et très variable d'un individu à un autre. La vérité "terrain" n'existe pas, car il n'est pas possible de déclencher une même émotion de manière répétée sur un individu. Certaines études sondent les sujets afin d'identifier des schémas communs. Nous souhaitons proposer une alternative en concevant un détecteur fonctionnant sur les mêmes bases que le cerveau humain.

Après avoir étudié l'existant, nous avons identifié des problématiques pour lesquelles nous voulions proposer une réponse. Dans le chapitre 1, nous détaillerons les travaux préalablement effectués dans les domaines que nous visons. C'est à partir de cette étude, que nous avons constaté deux éléments fondateurs de notre réflexion. D'une part, il existe une très grande quantité de connaissances à ce sujet sans alignement précis et d'autre part, concernant la détection des émotions, la temporalité lors de l'apparition d'une expression est rarement prise en compte. Ces deux constats nous ont poussés à réfléchir sur la manière de modéliser les émotions dans le but d'améliorer la qualité de la détection.

Tout d'abord, deux types de modèles existent pour décrire une émotion. Ils sont souvent considérés comme antagonistes de par l'approche continue de l'un et discrète du second. Il est alors nécessaire de faire un choix entre ces deux approches même s'il n'existe aucun consensus dans ce domaine. La question ici est : Comment concevoir un système capable d'unifier les modèles de description des émotions ? Cette problématique peut également se traduire par un besoin d'interopérabilité dans la description des émotions. L'unification des modèles de description a pour objectif de fournir un réponse en adéquation avec les attentes des utilisateurs. Par exemple, une annotation continue est difficilement compréhensible par un utilisateur qui aura tendence à préférer un mot correspondant à l'émotion ressentie. Cependant, l'observation de la valence peut être une information suffisante pour modifier dynamiquement le comportement de l'application.

Un autre aspect difficile à appréhender lors de la détection des émotions est la diversité des sources de données. Il est possible d'utiliser des données dynamiques correspondant aux expressions de la personne mais aussi des données statiques correspondant au contexte ou aux connaissances du sujet sur le stimulus déclencheur de l'émotion. Dans un premier temps, la grande variété des données dynamiques pose une problématique dans la conception de notre système. Comment concevoir

un système capable de s'abstraire des capteurs utilisés ? Le choix des capteurs utilisés lors de la détection est soumis à des contraintes utilisateur, comme le niveau d'invasivité ou le coût de ces derniers. Il est nécessaire de concevoir un détecteur capable de s'adapter à différents capteurs et d'utiliser leurs informations pour l'interprétation des émotions. Il est également intéressant dans un contexte réel de rendre le système robuste en l'absence d'un ou plusieurs capteurs, sans pour autant réentrainer le système. Dans un second temps, l'utilisation des données statiques reste problématique : comment utiliser les données contextuelles dans la détection des émotions ?

Pour cette thèse, notre objectif est double. D'une part, nous souhaitons proposer un modèle des émotions adapté à la détection, répondant aux problématiques précédentes, dans un contexte temps réel. D'autre part, nous cherchons à expérimenter de nouvelles voies, comme la modélisation proposée par Scherer, afin d'observer les émotions sous l'angle du changement émotionnel.

5 Principaux résultats

Le principal apport de nos travaux se trouve dans la réalisation d'un prototype de détecteur émotionnel. Nous avons basé son élaboration à partir du modèle de Scherer n'ayant jamais été implémenté. L'intérêt de ce modèle réside dans l'approche des émotions par les systèmes complexes permettant l'émergence de comportements non prévisibles, ainsi que l'identification de certains phénomènes définis par le terme de "catastrophe" au sens de René Tom. On peut ainsi avoir une vision continue des émotions, ce qui nous permet de les observer sous l'angle du changement. Une seconde partie de notre travail a consisté à coupler cette modélisation avec des informations sur le contexte de l'expérience émotionnelle.

Nous avons également construit notre système afin qu'il soit capable d'utiliser des capteurs variés dont la seule contrainte est de fournir une information numérique continue, par exemple, un grand nombre de capteurs physiologiques ou une caméra, en considérant les déplacements de points caractéristiques. En revanche, le texte ne peut être utilisé pour capter l'émotion dans notre système. Cette possibilité est permise par l'architecture modulaire de notre outil afin de pouvoir utiliser simultanément plusieurs capteurs dont le traitement est indépendant. Nous utilisons aussi cette capacité pour nos Carte Emotionnelle Personnalisées (CEPs) afin de permettre la description des émotions sur l'ensemble des modèles émotionnels dimensionnels.

Nous avons aussi travaillé sur l'évaluation des performances de notre prototype. La comparaison avec d'autres détecteurs est rendue difficile d'un côté, par l'impossibilité d'adapter les jeux de données disponibles [Healey 02, Picard 01], en raison de

l'absence d'information sur l'intensité émotionnelle, et d'un autre côté, par le paradigme d'observation des émotions, qui n'est pas quantifiable par les méthodes rencontrées dans la littérature. Nous proposons donc un protocole d'évaluation adapté à notre outil, ainsi que des indicateurs de performances permettant de quantifier les capacités de notre prototype. Nous utilisons pour cela la notion de déviation qui est la différence entre une annotation manuelle d'une expression faciale et l'annotation continue issue de notre système. A partir de notre jeu de test (une vidéo sur laquelle un acteur mime des expressions faciales), et après avoir paramétré notre algorithme de détection, nous proposons une analyse basée sur une vidéo de 1232 images, dont 163 ont été utilisées pour l'apprentissage. Le taux de déviation obtenu en moyenne sur l'ensemble des six émotions exprimées est de 1,44% pour l'axe de valence et 2,1% pour l'éveil. L'émotion ayant subi la plus forte amélioration après paramétrage de notre algorithme est la peur en passant d'une déviation de 20,95% pour la valence et -7,97% pour l'éveil à respectivement 1,62% et -1,42%. Ces mesures nous permettent de montrer que notre approche permet de détecter des émotions et ceci de manière précise.

Notre outil a été utilisé sur une vidéo, non exploitée pour l'apprentissage contenant 5757 images. Nous avons obtenu une identification du changement émotionnel pour l'ensemble des 17 séquences émotionnelles contenues dans la vidéo. Cependant, nous avons obtenu deux erreurs d'annotations concernant toujours le "dégoût" qui s'est vu être annoté comme une émotion proche de la "peur" ou de la "colère" de faible intensité.

Le dernier résultat impactant les capacités de détection vise l'intérêt de la contextualisation de l'émotion. Nous avons donc réalisé un questionnaire permettant de comparer notre système à l'être humain, dans le cas d'une annotation avec et sans prise en compte du contexte, afin d'identifier l'impact de celui-ci sur l'expérience émotionnelle. Á partir des réponses de 10 participants, nous avons extrait le taux moyen de similitude entre les annotations avec contexte et celles sans, ainsi que le taux d'accord entre les participants. Nous avons pu en tirer la conclusion que le contexte est utilisé pour identifier l'émotion ressentie, mais aussi qu'il n'existe pas une définition unique d'une phobie ou philie.

Nous souhaitons également mettre en avant l'observation du phénomène d'hysteresis lors de la détection d'une émotion. Ce comportement permet d'apporter un indice expérimental aux théories utilisées par Scherer, pour l'élaboration de son modèle que nous utilisons afin de développer nos CEPs.

Enfin, de par la construction de notre système, nous avons pu développer un algorithme de synthétisation des expressions à partir des données d'apprentissage inclues dans nos cartes. Les résultats obtenus ne permettent pas de rivaliser avec

les outils dédiés, mais montrent une certaine faisabilité pouvant être grandement améliorée, afin d'exploiter cette possibilité.

Les résultats nous permettent cependant de mettre en évidence l'intérêt d'utiliser un modèle hybride pour la détection des émotions. Nous avons pu obtenir des capacités de détection comparables aux systèmes existants tout en fournissant un système capable de s'adapter à la fois à une grande variété de capteurs, mais aussi à la manière d'exprimer les émotions.

6 Structure du manuscrit

Afin de décrire les travaux menés pendant cette thèse, nous diviserons ce manuscrit en cinq chapitres. Nous commencerons par une revue des travaux de notre domaine afin de mettre en perspective notre réalisation face aux voies empruntées par la communauté. De plus, le domaine abordé étant fortement inter-disciplinaire, un grand nombre de connaissances seront explicitées dans ce chapitre. Des notions de psychologie, médecine ou encore de traitement du signal seront utilisées pendant l'explication de nos travaux.

Nous continuerons ensuite avec une présentation générale de notre modèle et de son implémentation. Dans ce chapitre, nous montrerons comment nous avons réalisé notre modèle en partant de la théorie des deux facteurs de Schachter et Singer. Une description des différents modules et des données transitant entre eux sera spécifiée afin d'obtenir une vision globale de notre réalisation. Nous terminerons cette partie par une présentation de l'application complète pour fournir une vision concrète de ce que nous avons réalisé.

A partir de la présentation de notre modèle, nous développerons les deux principaux modules de notre système dans deux chapitres distincts. Dans un premier temps, nous détaillerons le fonctionnement de ce que nous appelons les CEPs. Nous en donnerons une définition formelle, puis nous continuerons sur les méthodes de création et d'utilisation de ces dernières. Nous concluerons cette partie par des fonctionnalités permises par l'utilisation de ces cartes, en particulier, sur la notion de confiance attribuée à la détection et aux capacités d'apprentissage induites par cette information, puis la possibilité d'utiliser celles-ci en sens inverse pour synthétiser les émotions.

Dans un second temps, nous présenterons le module de catégorisation et de contextualisation des émotions en décrivant la construction de Emotion Ontology for Context Awareness (EmOCA) [6]. Nous préciserons l'intérêt de l'utilisation d'un

[6] http://ns.inria.fr/emoca

tel outil et dans quelle mesure nous souhaitons le développer.

Après avoir détaillé l'ensemble de notre réalisation, nous présenterons les résultats que nous avons obtenus ainsi que les méthodologies mises en place pour leur quantification. Nous discuterons ensuite autour de ces résultats afin d'identifier les avantages et inconvénients de notre système, pour terminer sur les perspectives que nous envisageons pour l'avenir du projet.

Nous concluerons ce document par un retour sur l'ensemble de nos travaux et une mise en perspective de notre réalisation au sein du domaine de l'informatique affective.

Chapitre 1

L'*informatique affective*, une nouvelle interaction homme-machine

1 Introduction

Cette première partie a pour but de faire un tour d'horizon des connaissances sur les émotions et des techniques informatiques mises en oeuvre pour leurs captations. En effet, comme nous le verrons par la suite, cette notion paraît simple, mais nous en avons tous une vision personnelle, ce qui rend complexe sa définition. De plus l'*informatique affective* est un domaine reliant de nombreuses disciplines, c'est pourquoi nous ne prétendons pas proposer un état de l'art exhaustif, mais plutôt une sélection de travaux montrant l'évolution et les directions du domaine. La variété des champs disciplinaires et l'ancienneté de l'étude de ce phénomène nous fait remonter depuis les travaux de Darwin de 1872 jusqu'à nos jours, tout en traversant des domaines variés, tels que la psychologie, la médecine ou encore la philosophie.

Comme indiqué précédemment, l'objectif de nos travaux est de réaliser un détecteur émotionnel adaptable aux différents contextes applicatifs, ainsi que d'explorer de nouvelles sources de données permettant d'affiner la détection. Nous proposons ici de rassembler les connaissances nécessaires à la compréhension de nos travaux. Les détecteurs émotionnels proposés dans le domaine utilisent une structuration similaire. La figure 1.1 montre une architecture simplifiée de ces derniers comparativement aux travaux que nous avons menés.

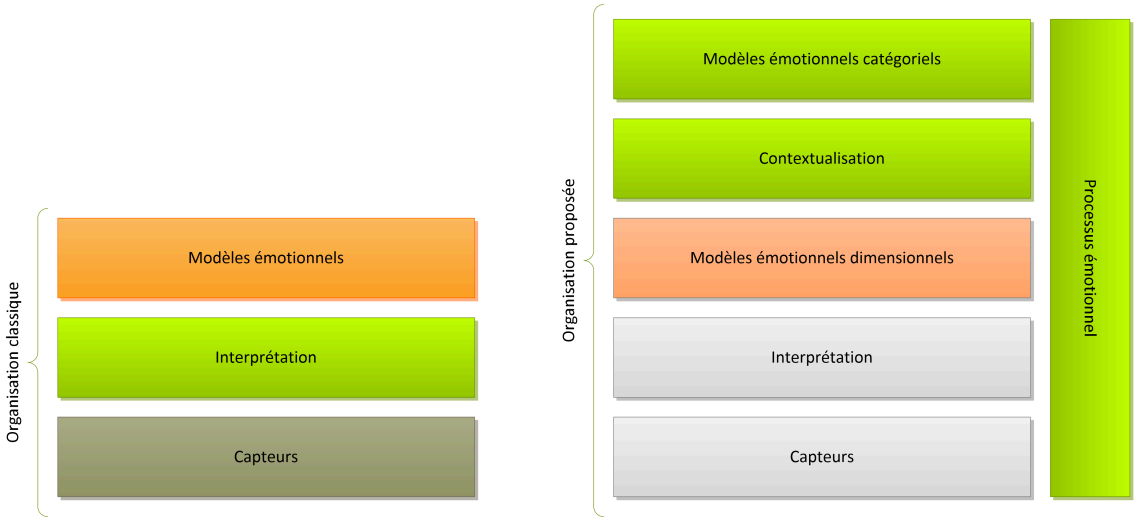


FIGURE 1.1 – Comparaison des architectures simplifiées des détecteurs existants par rapport à notre proposition

Nous définissons cette structure en trois parties, une première centrée sur les capteurs exploités, une seconde sur l'interprétation des données captées et une dernière permettant de décrire l'émotion identifée. Nous proposons dans ce document d'exploiter cette architecture en la complétant par la prise en compte du contexte émotionnel, mais également en calquant le fonctionnement de notre système sur le processus naturel d'apparition des émotions. Ces deux architectures seront utilisées dans ce chapitre comme structure des connaissances décrites.

Nous profiterons de cette première partie pour détailler les émotions sous trois aspects, d'abord les définitions, ensuite les théories d'expressions de l'émotion et enfin les modélisations existantes. La seconde partie se concentrera sur l'expression des émotions et les méthodes de captation couramment employées. La troisième partie présente les impacts comportementaux sur l'expérience émotionnelle, ainsi que les méthodes d'observation et d'élicitation des émotions à des fins expérimentales. Pour terminer, nous présenterons les travaux réalisés depuis la définition de l'*informatique affective* par Rosalind Picard en 1995 jusqu'à nos jours.

2 L'émotion, une notion complexe

2.1 Vers une définition commune

Le terme émotion que nous utilisons couramment est chargé de sens très varié. Il est difficile de donner une définition précise, mais il est également compliqué de le différencier de notions proches telles que les sentiments ou les traits de caractère.

Ce terme provient du latin "emovere" indiquant un mouvement. Cette étymologie montre que l'émotion n'est pas une notion passive, mais au contraire emplie d'action. Les ouvrages non spécialisés ne proposent pas de définition suffisamment explicite pour avoir une bonne appréhension de cette notion. Prenons comme exemple la définition du dictionnaire Larousse [1] :

> "Trouble subit, agitation passagère causés par un sentiment vif de peur, de surprise, de joie, etc..."

Un lien de cause à effet est directement introduit entre l'émotion et le sentiment. De plus, les termes qui caractérisent les sentiments sont souvent employés pour décrire les émotions, créant ainsi une confusion entre ces deux notions. Afin de palier ce manque de clarté, nous avons sélectionné une définition communément acceptée. Cette dernière est issue de l'analyse de 140 définitions provenant de toutes langues, cultures et domaines [Kleinginna 81] :

> "Emotion is a complex set of interactions among subjective and objective factors, mediated by neural/hormonal systems, which can
>
> a) give rise to affective experiences such as feelings of arousal, pleasure/displeasure ;
>
> b) generate cognitive processes such as emotionally relevant perceptual effects, appraisals, labeling processes ;
>
> c) activate widespread physiological adjustments to the arousing conditions ; and
>
> d) lead to behavior that is often, but not always, expressive, goal-directed, and adaptive."

De cette définition, trois points nous intéressent plus particulièrement :

- modifie le comportement du sujet (d)
- impact important sur l'ensemble du corps (c)
- l'espace minimum de définition des émotions composé des axes éveil et valence (a)

[1] http://www.larousse.com/fr/dictionnaires/francais/emotion

Le premier montre l'intérêt de détecter les émotions pour analyser le comportement d'un utilisateur. Le second valide la possibilité de détecter les émotions au travers des conséquences corporelles de l'émotion. Enfin, le dernier point propose une manière de décrire l'ensemble des émotions en fonction de l'éveil et de la valence d'une émotion. Même si cette définition est riche et décrit de manière précise les émotions, la différenciation avec les sentiments ou les traits de caractère reste encore difficile.

Le Réseau Européen d'Excellence HUMAINE [2] permettant de regrouper des travaux sur les émotions propose une définition des émotions permettant de compléter celle de Kleinginna et Kleinginna [Cowie 05] :

> "We [will] consider emotion in an inclusive sense rather than in the narrow sense of episodes where a strong rush of feeling briefly dominates a person's awareness. ... emotion in the broad sense pervades human communication and cognition. Human beings have positive or negative feelings about most things, people, events and symbols. These feelings strongly influence the way they attend, behave, plan, learn and select."

Celle-ci vient appuyer l'importance de l'émotion dans le processus de communication et d'attention d'un individu face à l'ensemble des situations qu'il peut rencontrer. Mais l'aspect qui nous semble important est la différenciation entre deux types d'émotions, d'une part les émotions au sens strict définies comme des états *épisodiques* dominant la conscience de l'individu, d'autre part, les émotions *envahissantes* au sens plus large permettant d'inclure des concepts d'humeur ou de sentiment ayant un impact important sur le comportement de l'utilisateur.

Ces deux définitions nous fournissent un point de départ pour notre réflexion sur les émotions en proposant un contexte général à la communication entre individus ou celle, entre des individus et un système artificiel.

2.2 Caractéristiques des émotions

Bien que les définitions des émotions proposées précédemment fournissent un cadre strict de ce concept, des ambiguïtés ne sont pas levées. Nous pouvons identifier des similitudes ou des confusions avec des termes tels que sentiments, humeurs, traits de caractère ...

Dans un livre sur la nature des émotions [Ekman 94], Paul Ekman et Richard J. Davidson ont proposé un questionnaire sur les caractéristiques des émotions à

[2] http://emotion-research.net/

plusieurs psychologues spécialisés dans le domaine des émotions. Parmi ces psychologues, nous pouvons noter des figures emblématiques de la théorie des émotions comme Frijda ("action tendency" [Frijda 87]), Lazarus (Théorie de l'évaluation [Lazarus 91]), Watson, Clark et Panksepp (Les quatres systèmes [Panksepp 82]). Un résumé de leur réponse est proposé par C. Belzung [Belzung 07] :

- la vitesse de déclenchement (rapide pour les émotions, lente pour l'humeur),
- la fréquence d'apparition (faible pour les émotions et élevée pour l'humeur),
- l'intensité subjective (forte pour les émotions, faible pour l'humeur),
- l'aptitude à identifier les éléments déclencheurs (facile pour les émotions et difficile pour l'humeur),
- la durée (quelques secondes ou minutes pour les émotions et plusieurs heures ou jours pour l'humeur),
- l'effet sur l'attention (respectivement fort et faible),
- la présence d'une composante physiologique et comportementale (respectivement présente et absente),
- la propriété (les émotions permettent l'action alors que l'humeur modifie les processus cognitifs comme la créativité et la flexibilité),
- l'intentionnalité (les émotions sont au sujet de quelque chose : "j'ai peur du lion", "je suis fâché contre X" alors que cette spécificité d'être à propos de quelque chose est absente de l'humeur).

En plus de caractériser la différence entre humeur et émotion, ces spécificités permettent d'approfondir la notion d'émotion. Dans le cadre de la détection des émotions, les notions de vitesse de déclenchement et de durée délimitent l'émotion dans le temps. La majorité du temps, nous ne ressentons pas d'émotion, cet état est souvent nommé "émotion neutre" même si cette dernière ne répond pas aux caractéristiques définies précédemment.

2.3 Positionnement

Dans nos travaux, nous nous appuierons sur ces définitions afin de limiter notre champ d'action, ainsi que pour définir des contraintes fortes. En particulier, les caractéristiques des émotions déterminent un temps de calcul de l'émotion, inférieur au temps de déclenchement de celle-ci, afin de pouvoir suivre l'évolution du processus. De plus, en délimitant précisément cette notion, des interactions émergent,

permettant d'influencer la détection par l'état émotionnel, au sens plus large, de l'utilisateur. Tout au long de nos travaux, nous prendrons comme référence ces trois éléments de définition à savoir la définition de Kleinginna, la distinction de Cowie entre l'émotion *envahissante* et *épisodique* et la liste de caractéristiques de Ekman.

3 Théories sur l'activation de l'émotion

Le fonctionnement interne du processus émotionnel est complexe. Les théories suivantes proposent de répondre aux questions sur le fonctionnement de l'émotion, depuis l'apparition du stimulus jusqu'au ressenti de l'émotion en proposant des schémas cognitifs et physiologiques de l'activité émotionnelle. Nous avons sélectionné les théories les plus marquantes depuis les théories évolutionnistes de Darwin jusqu'aux théories modernes de Scherer ou Philippot. Une majorité de ces théories sont maintenant rejetées par la communauté, mais elles montrent une importante évolution du domaine, ainsi que la difficulté à appréhender ce concept jusqu'à nos jours, où plusieurs théories cohabitent encore.

3.1 L'approche psycho-évolutionniste

Cette première approche est une des plus anciennes (1872). Le premier à avoir émis cette hypothèse fut Charles Darwin. Père de la théorie de l'évolution, il a voulu appliquer sa théorie aux émotions après avoir compris l'importance des émotions dans la survie des espèces. La réflexion proposée [Darwin 02] par ce courant est la suivante : les émotions existent dans une grande variété d'espèces, la sélection naturelle conserve donc cette capacité. Si elle est conservée, c'est que l'émotion possède une fonction adaptative, c'est-à-dire, qui sert à définir le comportement approprié en fonction d'un stimulus 1.2. De plus, pour les théoriciens psycho-évolutionnistes, le nombre d'émotions est restreint de par leur origine (sélection naturelle). Enfin, la seule réponse fournie concernant le mode de transmission des émotions est qu'elles sont innées. Il existe donc des schémas neuronnaux spécifiques pour chaque émotion, qui seront activés lors de l'apparition d'un stimulus.

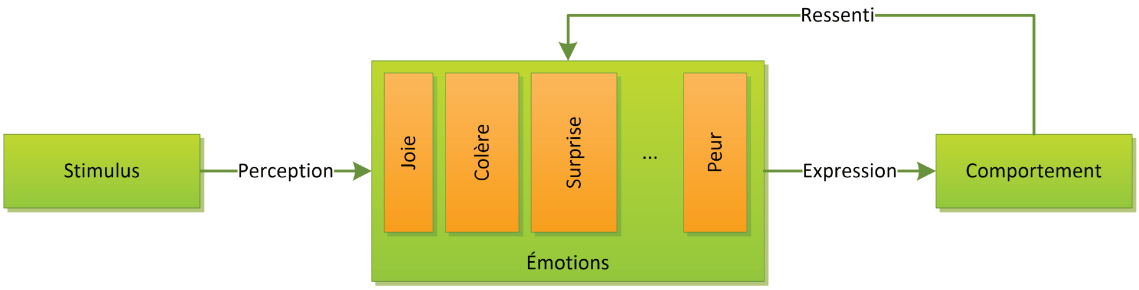


FIGURE 1.2 – Représentation graphique de la théorie psycho-évolutionniste

Cette idée de l'origine évolutionniste des émotions a été reprise plus récemment dans plusieurs travaux. Tomkins propose dans sa "théorie de l'amplification" [Tomkins 80] des précisions sur le nombre des émotions, qu'il determine au nombre de 9 (positive : intérêt, surprise, joie et négative : angoisse, peur, mépris, honte, dégoût, colère), et sur le moment du ressenti de l'émotion qui se produit par la perception de l'expression émotionnelle.

3.2 L'approche périphéraliste

Cette seconde théorie (1884) indépendemment défendue par le psychologue William James et le physicien Carl Lange, propose une vision des émotions centrée sur les retours corporels et physiologiques. William James résume cette théorie par la phrase suivante [James 84] :

> "My thesis on the contrary is that the bodily changes follow directly the PERCEPTION of the exciting fact, and that our feeling of the same changes as they occur IS the emotion."

Il complète cette description par quelques exemples [James 84] :

> "we feel sorry because we cry, angry because we strike, afraid because we tremble, and not that we cry, strike, or tremble, because we are sorry, angry, or fearful, as the case may be."

Le stimulus reste le déclencheur du processus émotionnel mais son interprétation par le système nerveux central n'a pour seul but que de déclencher des réflexes. C'est la perception de ces retours qui déclenche l'état emotionnel en fin de processus.

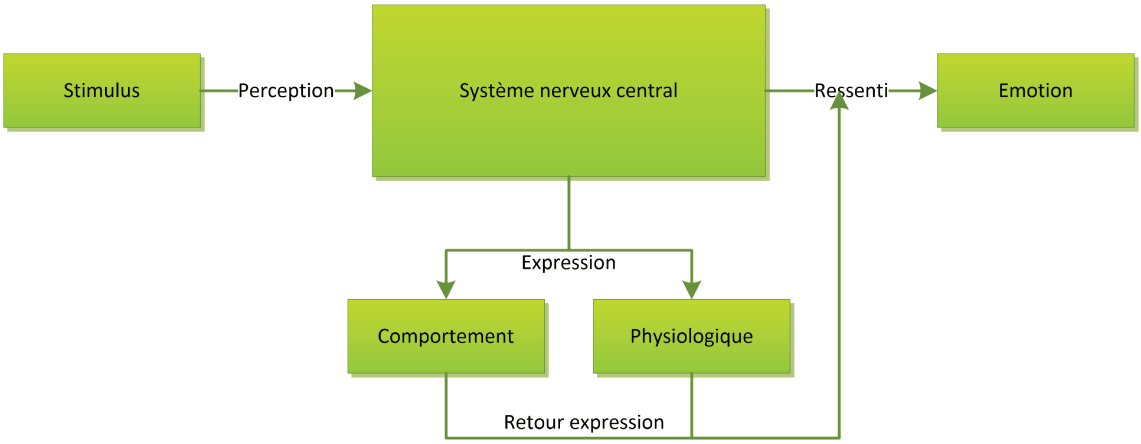


FIGURE 1.3 – Représentation graphique de la théorie périphéraliste

Cette théorie fut vivement critiquée par Cannon W.B. [Cannon 27], partisan du béhaviorisme. Les deux principales critiques se situent sur la non spécificité des réponses physiologiques et sur la présence d'émotions même chez des individus ayant subi une section haute de la moëlle épinière, supprimant la perception des réponses corporelles.

3.3 L'approche béhavioriste

La vision béhavioriste des émotions (1927) vient s'opposer aux théories périphéralistes en replaçant le système nerveux certral comme élément principal de l'expérience émotionnelle. Ce courant fondé en partie par W.B. Watson propose d'analyser les processus psychologiques avec des méthodes issues de la biologie. Ce dernier définit l'émotion comme "an hereditary pattern-reaction involving profound changes of the systems". Cette définition proche de la vision psycho-évolutionniste diffère par l'acceptation de la non spécificité des réponses car l'expérience émotionnelle n'est pas une conséquence de ces réponses. Dans ce courant de pensées, Cannon W.B. et Philip Bard élaborent une théorie des émotions remettant au centre du processus émotionnel le système nerveux central. La génération des réponses émotionnelles et le ressenti de l'émotion ne se produisent pas dans les mêmes centres nerveux.

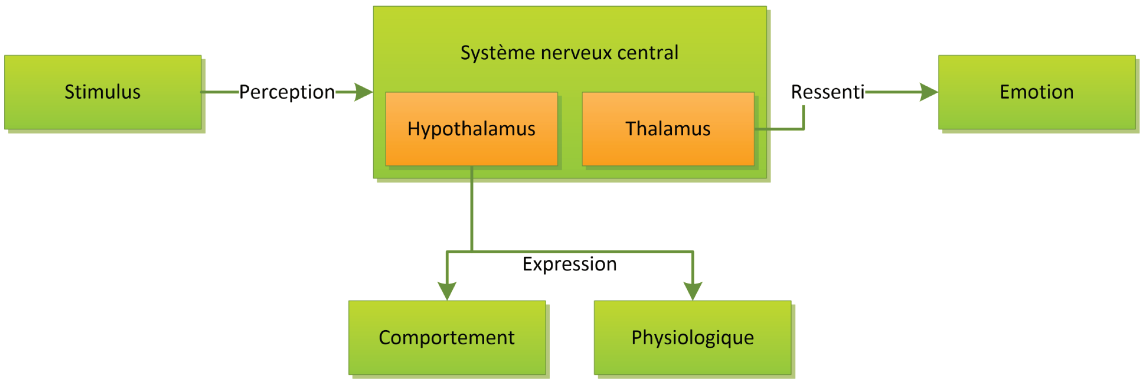


FIGURE 1.4 – Représentation graphique de la théorie de Cannon-Bard

Les critiques émisent par Canon [Cannon 27] envers les théories périphéralistes, ne sont pas applicables à cette théorie du fait de la séparation de l'expérience et de l'expression émotionnelle. En revanche, de nombreux auteurs s'accordent à dire que l'émotion n'est pas qu'un processus automatique, mais qu'il y a une part importante de connaissance du stimulus impactant l'expérience émotionnelle. L'émotion ne peut pas être qu'un réflexe conditionné à l'image de l'expérience de Skinner [Skinner 38]

3.4 L'approche cognitiviste

Les approches cognitivistes (1962) plus modernes se concentrent sur l'intégration de la cognition dans le processus émotionnel. La majorité des ces théories suppriment l'égémonie d'une source unique des émotions. Les notions de connaissances et de réflexions prennent une part importante dans cette vision des émotions fréquemment dénigrée dans les précédentes théories, montrant souvent un processus automatique du déclenchement inné.

La théorie des deux facteurs

La plus ancienne théorie cognitiviste (1962) fut élaborée par S. Schachter et J. E. Singer sous le nom de "Two-factor theory of emotion". Ces travaux tendent à prouver que les retours physiologiques sont importants dans le processus émotionnel, mais restent non spécifiques, et qu'il n'existe pas de schéma préétabli pour les émotions, mais plutôt un rôle important de la cognition dans la détermination des émotions.

Les auteurs ont réalisé une expérience [Schachter 62] afin de montrer le rôle de la cognition ainsi que des retours physiologiques dans l'émergence de l'émotion, en injectant de l'adrénaline à trois groupes de personnes. Cette substance a pour effet

d'augmenter les signes de l'éveil en entrainant une augmentation du rythme cardiaque, de la pression artérielle et de la transpiration cutanée, ainsi qu'une dilatation des bronches et des pupilles. Les différents groupes ont eu un niveau d'information différent sur cette substance, soit en étant complètement informés des effets, soit mal informés ou encore sans aucune information. A la suite de cette injection, les participants se sont vus exposer à un stimulus émotionnel représenté par un comédien jouant la joie ou la colère.

Les résultats obtenus montrent que le groupe informé ayant une justification à son état d'éveil, a eu un comportement normal, alors que les deux autres groupes ont eu une réaction plus importante face au stimulus.

Cette expérience montre donc l'importance du niveau d'éveil dans l'apparition de l'émotion, mais que l'interprétation cognitive des réponses physiologiques est essentielle dans l'expérience émotionnelle.

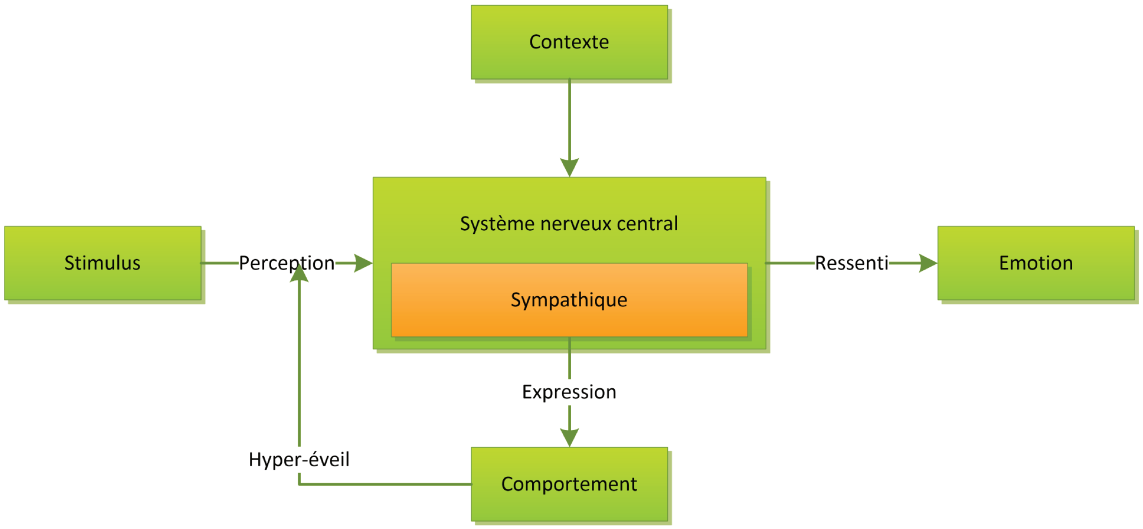


FIGURE 1.5 – Représentation graphique de la théorie des deux facteurs

La figure 1.5 est une représentation schématique de cette théorie. La perception d'un stimulus déclencheur d'émotions provoque une réaction au niveau du système sympathique activant ainsi des réactions reflexes en accord avec la situation rencontrée. Cette réaction positionne le sujet dans un état dit d'hyper-éveil dans lequel il est capable de ressentir une émotion. La détermination de l'émotion s'effectue alors en fonction de deux paramètres qui sont d'une part les réactions réflexes et d'autre part le contexte et les connaissances de l'individu sur le stimulus rencontré. L'expérimentation présentée, ainsi que la théorie qui en découle montre l'importance du contexte et des expressions corporelles dans la détermination de l'émotion.

Théorie de l'évaluation, Component Process Model (CPM)

Une définition de la CPM (2000) et des émotions est proposée par Klaus R. Scherer :

> "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" [Scherer 05]

L'émotion est présentée comme le résultat d'un processus découpé en cinq fonctions principales, chacune étant reliée à un sous-système organique, en réponse à l'évaluation d'un stimulus pouvant être externe (évènement naturel) ou interne (souvenir). La notion d'évaluation peut également être de deux types, d'une part l'évaluation intrinsèque sans prise en compte des besoins ou des buts et d'autre part, transactionnelle avec prise en compte des besoins ou des buts.

Le tableau 1.1 [Scherer 05] résume les différentes fonctions du processus émotionnel ainsi que les liens avec les sous-systèmes organiques et les composantes émotionnelles.

Emotion function	**Organismic subsystem and major substrata**	**Emotion component**
Evaluation of objects and events	Information processing (CNS)	Cognitive component (appraisal)
System regulation	Support (CNS, NES, ANS)	Neurophysiological component (bodily symptoms)
Preparation and direction of action	Executive (CNS)	Motivational component (action tendencies)
Communication of reaction and behavioral intention	Action (SNS)	Motor expression component (facial and vocal expression)
Monitoring of internal state and organism-environment interaction	Monitor (CNS)	Subjective feeling component (emotional experience)

Note : CNS = central nervous system ; NES = neuro-endocrine system ; ANS = autonomic nervous system ; SNS = somatic nervous system.

TABLE 1.1 – Tableau récapitulatif des liens entre fonctions, sous-sytèmes organiques et composants émotionnels [Scherer 05]

Chacune des fonctions définies par Scherer gère une composante importante de l'émotion et se situe sur un ou plusieurs sous-systèmes organiques. Á la différence des autres théories cognitives, la CPM précise certaines distinctions. En particulier, les réponses corporelles et les expressions motrices sont séparées en deux fonctions distinctes. De plus, même si la tendance à l'action est liée à l'évaluation [Arnold 60] , celles-ci ne font pas parties de la même fonction.

Comme présenté dans le schéma 1.6, au moment de la perception du stimulus, l'ensemble des sous-systèmes organiques est activé et synchronisé afin d'aboutir à l'expérience émotionnelle via la fonction de monitoring.

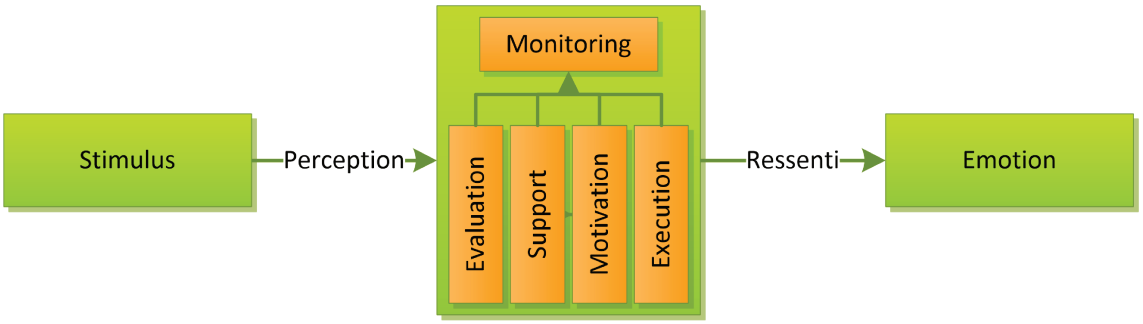


FIGURE 1.6 – Représentation graphique de la théorie CPM

3.5 Positionnement

Concernant les théories de l'expression émotionnelle, nous avons opté pour les théories cognitivistes et en particulier celle de Schachter et Singer. Ce choix est dû d'une part aux critiques existantes sur les théories plus anciennes, à savoir :

- l'aspect culturel des émotions pour les théories psycho-évolutionnistes
- la non-spécificité des réponses émotionnelles pour les théories périphéralistes
- la modification du ressenti émotionnel lors du changement d'expression pour les théories béhavioristes.

D'autre part, la théorie des deux facteurs est la première à intégrer le contexte au sein de l'interprétation de l'expérience émotionnelle. Cet aspect de l'expérience émotionnelle admis de nos jours, nous semble être un des changements importants dans la vision moderne des émotions. Nous proposons pour nos travaux de reproduire la synergie présentée dans la théorie des deux facteurs entre l'expression corporelle et le contexte émotionnel. Par la suite, la théorie de l'évaluation de Scherer reprendra cette notion dans une description plus structurée et détaillée du processus émotionnel. C'est pourquoi nous prendrons comme base la théorie de Schachter et Singer,

suffisante pour le moment dans le cadre de nos travaux, mais également compatible avec les travaux de Scherer.

4 Modèles de description

Comme nous l'avons expliqué, il existe un grand nombre de théories sur les émotions. Une partie de celles-ci décrivent également des modèles de description des émotions. Parmi ceux-ci, nous pouvons en définir deux catégories principales, les modèles dits catégoriels et les modèles dits dimensionnels. La grande différence entre ces deux catégories est le fait que les premiers considèrent qu'il existe un socle d'émotions dites primaires, communes chez tous les peuples [Izard 77, Ekman 92, Plutchik 01, Kemper 81]. Les seconds, considèrent que les émotions sont définies dans un espace continu, décrit par des composantes variables des émotions, telles que la valence ou l'éveil. Dans le domaine, ces deux approches sont souvent présentées comme antagonistes, en opposant le monde discret des modèles catégoriels au monde continu des modèles dimensionnels. Cependant, de nouvelles approches ont émergé plus récemment, exploitant des bases de connaissances (ontologies) ou des théories mathématiques pertinentes pour la représentation des émotions.

4.1 Les modèles catégoriels

Parmi les modèles des émotions, les modèles catégoriels sont les plus nombreux. En effet, de part leur construction simple, de nombreux chercheurs ont proposé leur vision des émotions. Ces types de modèles sont fortement liés à la vision évolutionniste des émotions. Le principe même de la définition d'un thesaurus des émotions commun aux êtres dotés de la capacité de ressentir les émotions, implique que celles-ci sont universelles. La notion d'universalité des émotions est définie par Ekman. Pour cela, il a réalisé une expérience à grande échelle afin de prouver qu'il existe un sous-ensemble minimal d'émotions communes. Pour valider son hypothèse, il utilise des images de visages occidentaux exprimant une certaine émotion et présente ces images à des populations n'ayant eu quasiment aucun contact avec des occidentaux, dans le but d'observer la reconnaissance des émotions.

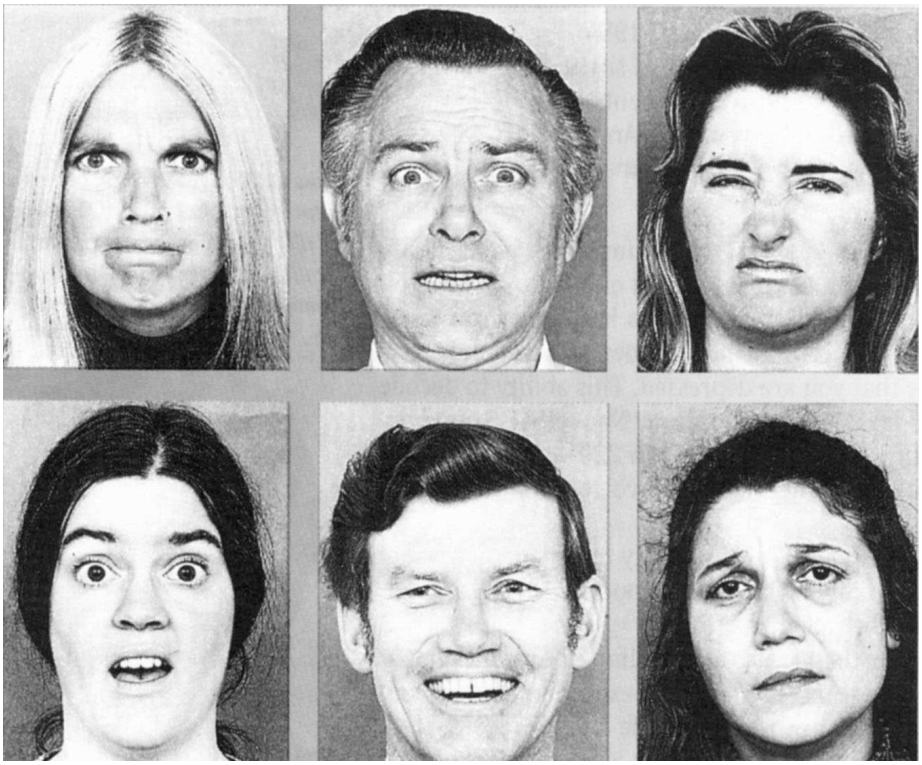


FIGURE 1.7 – Les 6 émotions basiques de Ekman. Le lecteur est invité à identifier les émotions exprimées ; réponses dans l'annexe D

Les résultats de ses observations montrent qu'il existe au moins 6 émotions communes à l'ensemble de l'humanité [Ekman 94]. Il étendra par la suite ces six émotions de base à 16 en ajoutant une plus grande variété d'émotions positives [Ekman 99]. D'autres travaux [Izard 77, Kemper 81] sont basés sur la notion d'émotion basique, afin d'étudier les liens avec le développement humain.

Parmi ces modèles, de nombreuses propositions existent [Izard 77, Tomkins 80, Ekman 92], mais le but ici n'étant pas de fournir une liste exhaustive de tous ces modèles, nous citerons le tableau récapitulatif suivant :

Auteurs	**Émotions**
Izard (1977)	Joie, surprise, colère, peur, tristesse, mépris, détresse, intérêt, culpabilité, honte, amour
Plutchik (1980)	Acceptation, colère, anticipation, dégoût, joie, peur, tristesse, surprise
Kemper (1981)	Peur, colère, dépression, satisfaction
Ekman (1992)	Colère, peur, tristesse, joie, dégoût, surprise

TABLE 1.2 – Résumé des émotions basiques [Belzung 07]

Enfin, le modèle le plus abouti dans cette catégorie est celui de Plutchik, nommé "modèle circomplexe". En plus de définir un ensemble de huits émotions basiques, il propose la notion de combinaison et d'intensité émotionnelle. La figure 1.8 montre une représentation graphique de ce modèle.

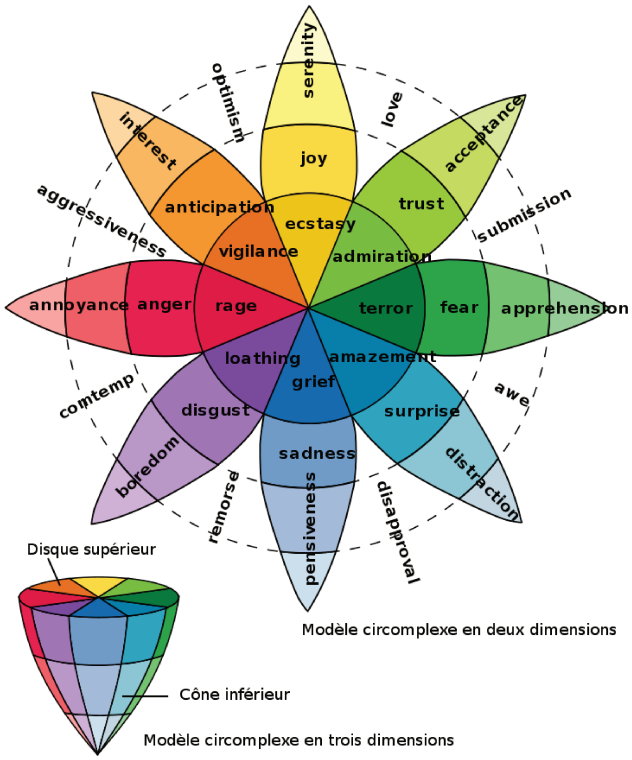


FIGURE 1.8 – La roue des émotions de Plutchik

Le disque supérieur représente les émotions basiques les plus intenses définies par Plutchik dans le tableau précédent 1.2 où les émotions opposées sont antagonistes (la joie et la tristesse ne peuvent pas être liées). L'intensité est traduite par l'axe centrale et diminue vers la pointe du modèle. Le mode de combinaison établi est limité car la combinaison d'émotions antagonistes n'est pas permise, et seules deux émotions primaires peuvent être combinées. Nous obtenons donc trois nouveaux groupes d'émotions nommés dyades primaires et secondaires, en associant respectivement deux émotions primaires voisines ou distantes de deux émotions.

4.2 Les modèles dimensionnels

Il existe plusieurs modèles dimensionnels permettant de représenter les émotions. Ils se différencient par un nombre variable de dimensions allant de 2 à 5. Un des modèles les plus fréquemment utilisé est celui de Russell [Anttonen 05]. Il est composé de deux dimensions, une première représentant la valence et une seconde l'éveil. Une grande

partie des modèles dimensionnels utilisent ces deux dimensions. En effet, la notion de valence permet de cibler deux types d'émotions, les positives et les négatives. La notion d'éveil permet quant à elle de donner une intensité de l'émotion.

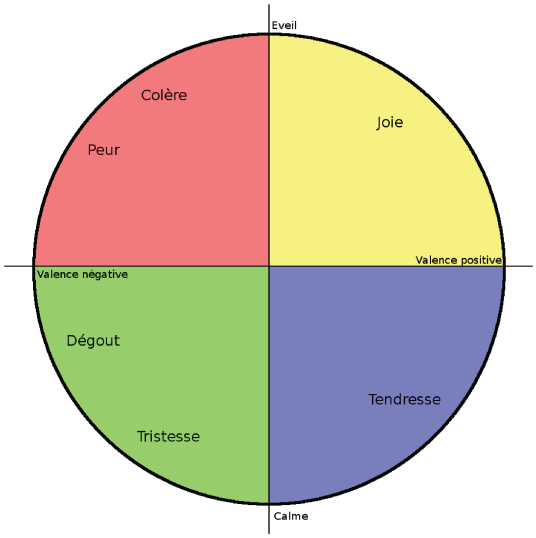


FIGURE 1.9 – Modèle bi-dimensionnel de Russell

Ces axes sont un point de départ permettant de discriminer les émotions. Mais certaines se retrouvent trop proches dans ce modèle, en particulier la peur et la colère ayant toutes deux une valence négative et un éveil important. C'est pourquoi, les modèles utilisant celui-ci comme base, ajoutent de nouvelles dimensions permettant de discriminer d'une manière plus importante les émotions.

Le modèle de Scherer [Fontaine 07] pour sa part, ajoute deux nouvelles dimensions, le contrôle et l'imprévisibilité. Le premier de ces axes permet de séparer les émotions dans lesquelles, la personne qui les ressent est en position de dominance ou de soumission. Nous placerons sur cet axe des émotions avec un fort niveau de contrôle, telles que la fierté et la colère, et celles avec un faible niveau de contrôle, comme la tristesse et la honte. Le deuxième axe permet de dissocier la surprise des autres émotions. En effet, celle-ci peut avoir une valence variable, mais elle peut également être impliquée dans d'autres émotions, telles que la peur. Nous pouvons donc opposer sur cette dimension, des émotions, telles que peur (imprévisibilité forte) à stress (imprévisibilité faible). D'autres modèles dimensionnels existent, créés en fonction des besoins applicatifs. Nous pouvons citer celui proposé par la société Emotiv [3] commercialisant le premier casque EEG à bas coût, ayant pour objectif de détecter les émotions, composé de quatre dimensions (excitation, engagement/ennui, méditation et frustration).

[3] http://www.emotiv.com/

4.3 Ontologies des émotions

Présentation générale

Plus récemment, des travaux ont été effectués sur une modélisation des émotions via les ontologies. Ce nouvel outil informatique est fortement utilisé dans le domaine du web sémantique [Berners-Lee 01]. La structuration et la définition des données issues du web sont au centre des problématiques de ce mouvement. L'ambition de structurer la quantité et la diversité des données issues du web a obligé la création d'outils spécifiques à cette tâche. La notion d'ontologie [Gruber 95] informatique est alors apparue pour répondre à ce besoin. Ces outils reposent sur deux concepts, d'une part l'identification unique des ressources via la notion d'URI (Uniform Resource Identifier) [Masinter 05] structurées sous la forme de triplet (sujet, prédicat, objet) et d'autre part, un vocabulaire définissant la sémantique des données.

L'intérêt de développer des ontologies du domaine est d'une part la possibilité d'organiser les connaissances, mais aussi de lier ces informations avec d'autres domaines. Par exemple, FOAF est une ontologie définissant les personnes et les interactions entre elles. Le concept de "Giant Global Graph" défini par Tim Berners-Lee [Berners-Lee 07] est la possibilité de pouvoir naviguer au travers des données en parcourant les relations les liant, sans restriction. Cet objectif visé par le domaine n'est réalisable que par la définition d'ontologies liées entre elles et permettant de communiquer via la sémantique des connaissances.

La notion d'ontologie repose sur la logique de description. Cette logique permet de décrire la sémantique des objets qui nous entourent. Pour cela, chaque logique de description définit un ensemble d'opérateurs tels que la subsomption(relation hiérarchique) ou l'inclusion permettant de décrire plus ou moins d'objets. Cette capacité est décrite comme étant son niveau d'expressivité. Les connaissances sont alors représentées d'une part via une ontologie décrivant la structure des données et d'autre part via des instances constituant les données compatibles avec la structure définie par l'ontologie. L'ensemble, constitué de l'ontologie et des individus, représente la base de connaissances. En plus des capacités d'interopérabilité et d'organisation des connaissances, un raisonnement est réalisable sur la base de connaissances. Ce raisonnement se traduit par la capacité d'inférer sur les connaissances décrites. Des exemples de déduction peuvent être la subsomption (inclusion ou sous-concept), la cohérence de la base de connaissances ou la vérification des relations.

L'implémentation d'une ontologie se réalise à partir de langages de descriptions tels que RDF Schema (RDFs) ou Web Ontology Language (OWL) et Resource Description Framework (RDF) pour les individus. Le choix d'un langage s'effectue en fonction du niveau d'expressivité nécessaire à la modélisation du domaine, mais

ces langages et les bases de connaissances qui en découlent, restent compatibles entre eux. C'est cette interopérabilité qui permet d'envisager le "Giant Global Graph".

Les propositions du domaine

Dans le domaine de l'*informatique affective*, l'utilisation des ontologies est envisagée pour décrire l'émotion ainsi que les interactions entre les différentes connaissances du domaine. Au moment de la rédaction de ces lignes, il existe peu d'ontologies sur les émotions. Nous en avons référencé quatre [Obrenovic 05, Grassi 09, Hastings 11, Smith 04, Grenon 04] spécialisées dans la description ou l'annotation des détections émotionnelles. D'autres ontologies telles que "Smiley Ontology" [4], définissent la notion d'émotion, mais avec une vision hors de nos préoccupations.

Les premiers travaux [Obrenovic 05] effectués dans cette voie proposent de décrire les émotions au travers de leurs expressions. Il devient possible de décrire les émotions directement depuis les capteurs utilisés pour la détection. Cette approche se base sur le fait qu'il existe une expression universelle des émotions. Il est possible d'inscrire ces travaux dans la vision d'Ekman, dans laquelle il montre que les expressions faciales d'un groupe restreint d'émotions sont universelles. Un tel système peut mener à la constitution d'une base de connaissances de toutes les règles de détection des émotions. Cependant, la définition de ces règles reste un sujet complexe et en constante évolution. De plus, les travaux de Cannon tendent à montrer que les réponses émotionnelles ne sont pas spécifiques. Les théories modèles se basent sur le contexte et les connaissances de l'utilisateur afin d'expliquer cette non-spécificité. C'est pourquoi une partie des chercheurs ayant proposé cette ontologie ont développé une nouvelle ontologie [López 08] prenant en compte le contexte de la détection émotionnelle 1.10. L'usage des ontologies permet alors de rendre compatible ces deux visions des émotions en alignant les concepts, c'est à dire en décrivant la sémantique des concepts similaires entre deux ontologies distinctes.

[4] http://smileyontology.com

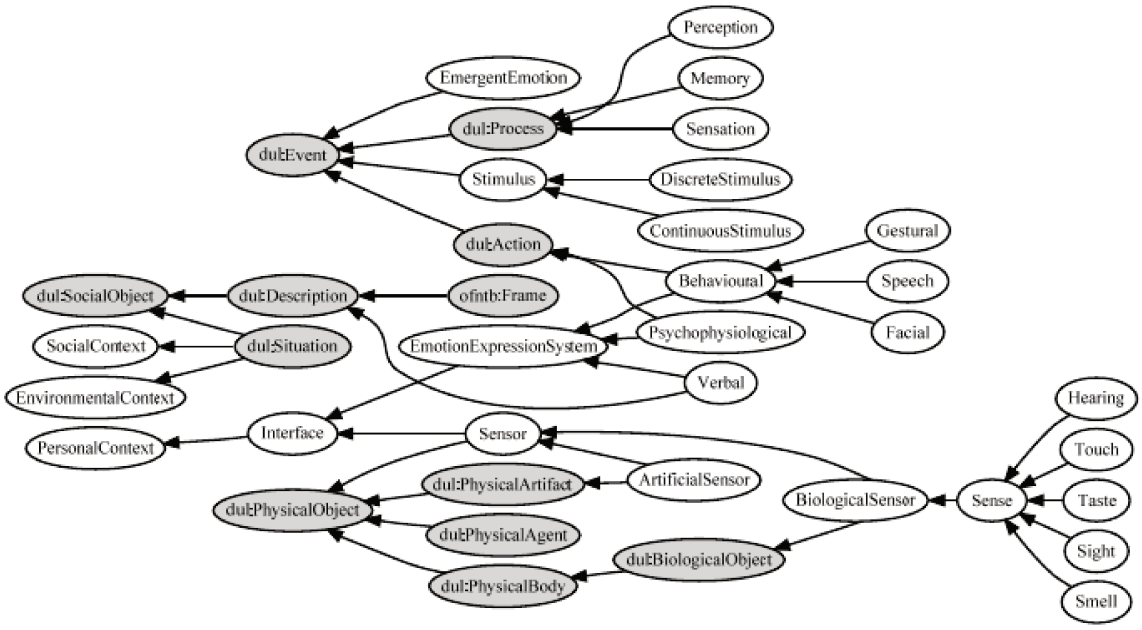


FIGURE 1.10 – Ontologie des émotions du groupe GRIHO

Une autre manière de décrire les émotions est de compiler l'ensemble des connaissances du domaine au sein d'une même structure qu'est l'ontologie.Cette approche se retrouve dans l'ontologie Human Emotions Ontology (HEO) [Grassi 09], visant le domaine de l'annotation des données multimedia. La figure 1.11 représente un schéma de cette ontologie. L'émotion peut être définie suivant une grande variété de modèles émotionnels. De plus, il est possible d'ajouter à l'annotation le mode de détection ainsi que les modalités utilisées. Les bases théoriques de cette conception viennent du langage EmotionML [5] du World Wide Web Consortium (W3C). Ce langage basé sur une syntaxe Extensible Markup Language (XML) permet de décrire les émotions. Défini par le W3C, ce langage a pour objectif de standardiser la description des émotions dans le cadre d'annotation manuelle, de détection automatique ou de synthétisation des émotions. La mise en oeuvre d'une ontologie autour de ce langage permet de compléter l'utilisation d'un tel langage avec des capacités de raisonnement liées à la base de connaissances.

[5] http://www.w3.org/TR/emotionml/

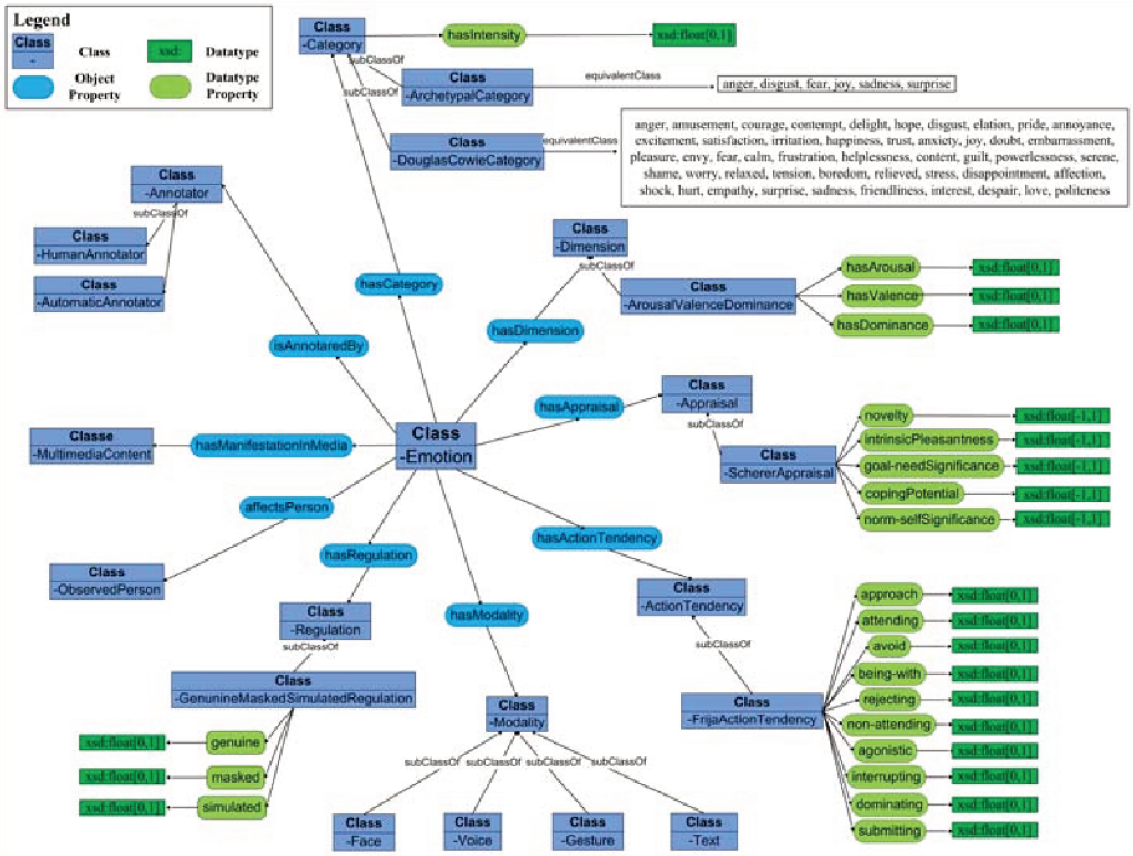


FIGURE 1.11 – Ontologie des émotions HEO

L'ontologie la plus récente du domaine dont une représentation est fournie par la figure 1.12, vise les mêmes objectifs de désambiguïsation et de catégorisation des émotions mais sur un aspect plus médical. Emotion Ontology (EMO) [Hastings 11] tend à distinguer les aspects intemporels des émotions, des détections des états émotionnels ponctuels. Il est alors possible de décrire les caractéristiques persistantes des émotions comme la tendance à l'action mais aussi une expérience émotionnelle. Cette notion de temporalité absente de la définition standard des ontologies est permise grâce à l'utilisation de Basic Formal Ontology (BFO) [Smith 04, Grenon 04] structurant les données en fonction de leur persistance dans le temps.

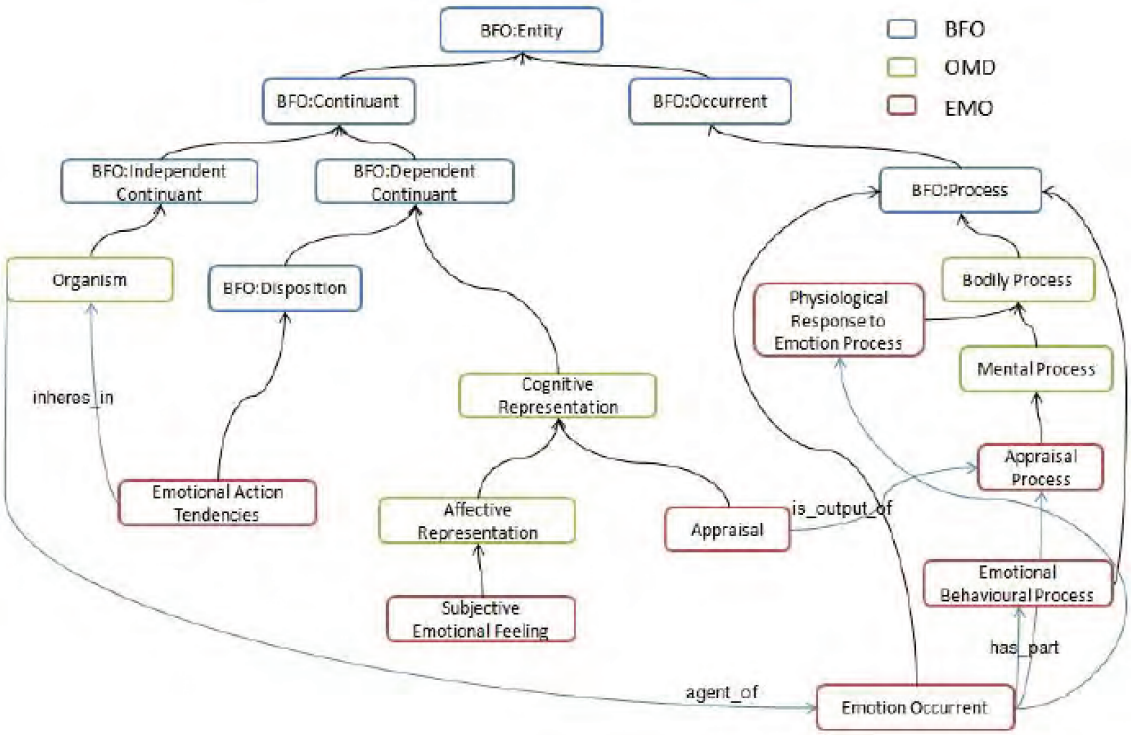


FIGURE 1.12 – Ontologie des émotions EMO

4.4 Modèle de Scherer

Pour terminer notre présentation des modèles émotionnels, nous souhaitons présenter un modèle en marge de ce qui est communément utilisé. Scherer propose une nouvelle vison des émotions basée sur la notion d'hysteresis [Scherer 00, Scherer 09]. Cette notion issue de la théorie des catastrophes de René Tom [Thom 72] permet de modéliser une discontinuité dans les systèmes complexes. Une définition des systèmes complexes peut être la suivante :

> "Un système complexe est un ensemble constitué d'un grand nombre d'entités en interaction qui empêchent l'observateur de prévoir sa rétro-action, son comportement ou son évolution par le calcul." [6]

La théorie de René Tom permet d'analyser les discontinuités dans les systèmes complexes. Une des conclusions de ses travaux est qu'il existe un nombre défini de catastrophes (discontinuités) dans les systèmes possédant un nombre de paramètres limités. Pour les systèmes à quatre paramètres, il en existe 7. Des travaux ultérieurs ont montré qu'il en existe 11 pour les systèmes à cinq paramètres, puis une infinité

[6] http://fr.wikipedia.org/wiki/Système_complexe consulté en Mai 2013

pour les autres systèmes. La catastrophe qui a particulièrement interessée Klaus R. Scherer se nomme l'hysteresis [Fig. 1a]. Cette catastrophe permet de définir deux états différents pour des paramètres identiques. Le changement d'état s'effectue en fonction de la situation précédente.

Pour Scherer, cela explique des situations telles que le passage du calme à la colère. Sur la représentation de ce changement [Fig. 1b], on observe la présence de deux chemins pour la même modification. Cela s'explique lors du passage du calme vers la colère comme une forme de tolérance face au stimulus en augmentation. Pour le changement de la colère au calme en revanche, la situation est différente. Le sujet déjà dans une situation de stress n'acceptera pas un stimulus de même intensité pour revenir au calme mais d'une intensité plus faible.

4.5 EmotionML

Le dernier chainon manquant dans l'utilisation des émotions, est la possibilité de décrire l'émotion de manière standardisée. Le w3c [7] fournit, depuis peu, un nouveau langage de description nommé EmotionML permettant de decrire les émotions. Basé sur la notion de langage balisé, il trouve son intérêt pour annoter des données ou pour faire transiter des émotions entre applications, de manière standardisée.

La structure simple de ces documents, comportant dix balises différentes, permet néanmoins de décrire très largement tout type d'émotion dans des cas d'utilisation très variés. Nous décrirons ici rapidement les différentes balises utilisables afin de montrer l'utilisation d'un tel langage de description.

La balise racine de ce langage est <emotionml>, elle est le conteneur d'une ou plusieurs balises <emotion> décrivant chacune une émotion. De plus chacune de ces balises peuvent contenir un élément <info> permettant de décrire des méta-données sur le contexte de l'analyse.

L'élément <emotion> contient l'ensemble de la définition d'une émotion détectée, pour cela il est nécessaire de définir le vocabulaire que l'on souhaite utiliser pour la décrire. Ce système permet donc de faire référence à un ou plusieurs modèles des émotions. Il existe quatre variétés de vocabulaire :

- <category> faisant référence aux modèles catégoriels
- <dimension> faisant référence aux modèles dimensionnels
- <appraisal> faisant référence au CPM

[7] http://www.w3.org/TR/emotionml/

- <action-tendency> permettant de décrire le comportement de la personne

Dans certains cas, il est possible de compléter ces derniers par la création d'un vocabulaire personnalisé. Chacune de ces balises contient les attributs "name", "value" et "confidence", sauf dans le cas de l'élément <category> qui ne contiendra pas l'attribut "value". Le premier, renvoit à un terme du vocabulaire faisant référence dans le cas de :

- <category> à l'émotion ressentie
- <dimension> au nom d'une dimension dans un modèle
- <appraisal> à un terme du vocabulaire d'évaluation
- <action-tendency> à un comportement

L'attribut "value" permet de donner une note comprise entre 0 et 1, décrivant le niveau de la propriété associée. Enfin, l'attribut "confidence" donne une indication sur la confiance à donner à cet élément, par une représentation numérique comprise entre 0 et 1.

La balise <emotion> peut être complétée par l'attribut "expressed-through", permettant de définir la ou les modalités d'expressions utilisées pour la capture. Cet attribut peut prendre des valeurs parmi "voice", "text", "face" ou "body". Il est cependant possible d'être cet ensemble de valeur en fonction des besoins.

Le dernier aspect de la description, est la possibilité d'associer cette détection à un stimulus. La notion de référence permet donc de lier la capture et le stimulus qui est représenté par une URI. Il devient donc possible d'annoter tous types de média, même les videos grâce à la notion de "timestamp" qui est incluse dans la balise <reference>.

4.6 Positionnement

Parmi la variété de modèles existants, nous n'en voyons pas pouvant être adaptés à une grande variété de contextes. Les modèles catégoriels ont l'avantage d'être compréhensibles par l'homme mais avec des nuances faibles voire inexistantes. Le problème est inversé pour les modèles dimensionnels. L'unification des connaissances par les ontologies est une solution intéressante mais ne valorisant pas suffisamment la complémentarité de ces deux modèles. Enfin, le modèle de Scherer bien que très intéressant n'a à notre connaissance jamais eu d'implémentation concrète. Il n'y a donc aucune certitude sur la capacité d'un tel modèle à décrire la réalité de l'émotion.

5 Expression et captation des émotions

Le point de départ de l'expérience émotionnelle est notre environnement. En effet, l'émotion est déclenchée par un stimulus interne (souvenir) ou externe (environnement). L'observation de ces stimulus étant trop variée et complexe, la seconde voie communément admise dans le domaine est l'observation des expressions émotionnelles. Afin d'informatiser ces expressions, de nombreux capteurs ont été étudiés afin d'extraire une information émotionnelle. L'utilisation de capteurs pour la détection des émotions implique d'analyser certains paramètres afin de déterminer l'utilisabilité de ces derniers :

- invasivité : niveau de restriction de la liberté de mouvement impliqué par le capteur (exemple : bracelet/scanner)
- prix : l'investissement pour l'acquisition d'un tel capteur (exemple : caméra/scanner)
- fiabilité : pertinence des données issues du capteur (exemple : casque EEG/scanner)

A partir de ces paramètre et du contexte applicatif, nous pouvons déterminer le niveau d'acceptation d'un capteur. Par exemple, dans le cadre des jeux vidéo, un casque EEG est un capteur intéressant. Un joueur est capable d'accepter un tel niveau d'invasivité avec la promesse d'une amélioration de l'immersion en revanche il ne pourrait pas accepter un coût trop important dans l'achat d'un capteur. Dans le cadre d'un suivi pour personnes âgées, le coût du capteur n'est plus un problème en revanche le niveau d'invasivité doit être le plus faible possible et la fiabilité accrue afin de ne pas activer des alarmes sans raison.

Bien que le choix de capteurs puisse se faire en fonction des paramètres précédents, tous ne sont pas adaptés à la détection des émotions. C'est pourquoi de nombreux travaux étudient des capteurs potentiellement utilisable pour cette tâche. Certains capteurs ont déjà été éprouvés en partant du rythme cardiaque jusqu'à l'Imagerie par Résonance Magnétique fonctionnelle (IRMf) [Kassam 13].

Dans cette partie, nous exposerons les travaux détaillant la mise en oeuvre de ces capteurs ainsi que des techniques permettant d'extraire des données pertinentes afin de fournir des indices sur l'état émotionnel de l'utilisateur.

5.1 Catégorisation des expressions

Un point commun de toutes les théories sur les émotions est qu'il existe une phase d'expression des émotions. Les travaux de David Marr [Marr 82] sur la trichotomie des émotions fournissent une première catégorisation de ces expressions. Les 3 modalités proposées sont :

- Physiologique (ex. rythme cardiaque)
- Psychologique (ex. voix)
- Comportementale (ex. recul)

Ce découpage proposé par David Marr permet d'identifier les trois modes d'expression des émotions. Les expressions dites physiologiques, comprennent la plus grande variété de capteurs. De plus, ces capteurs existent déjà car largement utilisés dans le domaine médical. L'utilisation de cette modalité pour l'observation comporte un avantage important qui est la difficulté de falsification de l'émotion par ce biai. Cet avantage contraint donc l'utilisation de techniques d'ellicitation de l'émotion afin de capter les indices pertinents.

Les deux autres modalités ont pour avantage l'utilisation de capteurs externes (caméra ou microphone) et sont facilement simulables. Le jeu d'acteur faisant passer les émotions au travers de ces modalités, il est plus facile dans le cadre d'une expérimentation d'observer les émotions de manière contrôlée. De plus, l'être humain qui est le meilleur détecteur émotionnel à ce jour n'utilise que ces deux modalités pour capturer l'émotion de son interlocuteur.

5.2 Les réponses physiologiques

Nous commencerons par présenter les capteurs physiologiques. Parmi ces derniers, il existe une grande variété d'expressions des émotions mettant en cause des réactions physiologiques. L'énumération complète de ces capteurs ne rentrant pas dans le cadre de ce document, nous nous limiterons aux capteurs les plus couramments utilisés dans le domaine.

Rythme cardiaque

Le rythme cardiaque est le capteur le plus évident lorsque l'on pense à l'expression des émotions. En effet, lors du ressenti d'une émotion, nous sentons rapidement la modification surtout lors d'une forte peur, colère ou joie. L'analyse du comportement du muscle cardiaque peut être effectuée à plusieurs niveaux. La manière la plus

simple d'obtenir une information relative aux modifications de notre rythme cardiaque est le nombre de pulsation par minute. Cependant d'autres capteurs peuvent fournir des informations plus précises comme l'Électrocardiogramme (ECG) ou le photopléthysmographe.

L'avantage majeur de ce capteur est sa faible invasivité en particulier pour le photopléthysmographe ou le capteur de rythme cardiaque. De plus, ces capteurs sont fortement utilisés dans le cadre sportif ce qui les rend peu couteux et facilement utilisables par un non spécialiste. Un autre avantage réside dans la bonne connaissance des algorithmes de traitement du signal rendant les mesures de ce capteur fiables et faiblement bruitées.

Concernant la détection des émotions, le principal rôle de ce capteur est de fournir une information sur l'éveil de l'utilisateur. Cette capacité est facilement compréhensible si l'on considère l'utilité de l'augmentation du rythme cardiaque. En effet, l'augmentation de notre rythme cardiaque est provoquée par l'effort physique. Dans le cadre des émotions, l'activité n'ayant pas eu lieu, cette augmentation est provoquée par anticipation de l'organisme à un effort. Lors d'une situation de colère, le corps se prépare à un potentiel affrontement physique. De même pour la peur, l'organisme se prépare à prendre la fuite. L'augmentation du rythme cardiaque permet alors de fournir aux muscles les apports nécessaires pour répondre rapidement à un potentiel effort physique. C'est ce qui correspond à la tendance à l'action dans la théorie de Klaus R. Scherer. Nous pouvons ainsi faire le lien entre l'éveil du sujet et son rythme cardiaque.

Cependant, l'exploitation de cette information uniquement pour la dimension de l'éveil est restrictif. Les travaux de Daniela Palomba [Palomba 97] sur l'évolution du rythme cardiaque lors d'une expérience émotionnelle mettent en évidence la possibilité d'extraire des informations sur la valence du sujet. Les conclusions de ses travaux tendent à montrer que la décéleration du rythme cardiaque est plus importante lors de l'expérience d'émotions négatives.

Sudation et température cutanée

Ces informations relativement simples à capturer sont fortement utilisées en *informatique affective*. La société affectiva [8] basée en partie sur les travaux de Rosalind W. Picard du MIT, développe un bracelet permettant d'extraire l'éveil émotionnel à partir de ces données.

Ce capteur possède de nombreux avantages parmi lesquels un faible niveau d'invasivité ainsi qu'une réactivité importante face au changement émotionnel. Ces deux

[8] http://www.affectiva.com

informations fortement inter-dépendantes sont très visibles dans des conditions de stress, mais se révèlent pertinentes pour identifier le niveau d'éveil d'un individu.

Activité cérébrale

Plus récemment, les recherches se sont orientées sur l'analyse du système nerveux central. Notre cerveau est l'élément unificateur de l'expérience émotionnelle. Que ce soit lors de la captation du stimulus, du déclenchement des expressions ou du ressenti émotionnel, toutes ces actions transitent par le cerveau. Cependant, bien que l'observation de l'activité cérabrale semble la solution optimale pour effectuer une détection fiable, de nombreuses difficultés existent pour sa mise en oeuvre. Il est en effet, difficile de dissocier les informations liées au processus émotionnel de l'activité globale du cerveau.

Dans ce domaine, deux grandes variétés de phénomènes sont observées, d'une part l'activité électrique et d'autre par le flux sanguin. Ces deux approches ont pour objectif de mettre en évidence les zones du cerveau en activité à un instant précis. En effet, depuis de nombreuses années, les scientifiques ont découpé le cerveau en zones correspondant à certaines fonctions de l'organisme. Par exemple, le lobe occipital situé à l'arrière du crâne est le siège de la vision. La principale différence entre les deux approches de l'exploration cérébrale se situe au niveau de la fiabilité des informations obtenues, du niveau d'invasivité des capteurs utilisés et du coût de ces derniers.

Les IRMfs sont utilisés afin d'observer le flux sanguin à l'intérieur de notre cerveau. Une concentration plus forte dans une zone montre une activation de cette dernière. L'information obtenue est extrèmement précise mais nécessite un environnement médicalisé pour son exploitation. Le niveau d'invasivité ainsi que le coût d'un tel système limite son utilisation au cadre expérimental. Cependant, de récents travaux [Kassam 13] ont montré un lien entre le ressenti émotionnel et l'activation de certaines zones cérébrales. Une expérience a même été menée afin de détecter l'émotion d'un utilisateur à partir des imageries médicales et d'un algorithme de classification.

Néanmoins, dans le domaine de l'*informatique affective* c'est la seconde approche qui est le plus souvent étudiée en raison de son coût plus faible et de sa facilité de mise en oeuvre. L'EEG utilise des électrodes afin de capturer les modifications de potentiel électrique au niveau du cuir chevelu. Cette méthode moins précise que l'IRMf permet également de capturer l'activation des zones du cerveau grâce à une disposition des électrodes de manière standardisée sur le crâne. L'engouement pour ces capteurs est tel que plusieurs sociétés se sont lancées dans le développement d'EEG à faible coût et simple d'utilisation. Par exemple, le casque Epoc de la société Emotiv s'oriente sur

l'utilisation de ce capteur pour les jeux vidéo dont une fonctionnalité, est la détection des émotions. Ce casque se démarque de par son coût très faible, une connexion sans-fil et une structure rigide des capteurs, ne nécessitant pas de connaissances pour son installation.

Dans le cadre des émotions, l'exploitation d'un EEG s'effectue principalement par une étude du spectre des fréquences. L'utilisation d'un algorithme simple comme la transformée de fourrier permet de transformer le signal de l'espace temps à l'espace des fréquences. Ceci permet d'observer l'activation des rythmes cérébraux. Il en existe cinq principaux couvrant une plage de fréquence de 4 à 50 Hz. A chacun de ces rythmes est associé un comportement pour la personne les activant. Par exemple, une activation du rythme β, correspondant aux fréquences entre 12 et 45Hz, est associée à un état de concentration. L'étude de l'activation de ces rythmes et de leur synchronisation est une voie empruntée par la communauté pour la détection des émotions. L'utilisation des rythmes cérébraux a permis l'identification de relations entre l'activité cérébrale et l'apprentissage, en se basant sur une analyse émotionnelle [Heraz 09].

D'autres outils d'analyse sont employés en particuliers, l'observation du potentiel évoqué P300. Cette onde apparaît 300ms après la présentation d'un stimulus visuel à un sujet. L'analyse de son amplitude est utilisée afin d'identifier les émotions ressenties. Cependant, ce phénomène est difficile à observer. Il est nécessaire d'employer des capteurs performants et souvent plus difficiles à manipuler. C'est pourquoi ce type d'analyse se limite au champ expérimental. Il existe néanmoins des travaux exploitant cette information afin de lier la personnalité au ressenti émotionnel. Par exemple, une étude [Mardaga 09] montre une différence d'amplitude de cette onde face à un stimulus plaisant, en fonction de certains traits de caractères.

Oculométrie

C'est une technique permettant le suivi des mouvements oculaires, grâce à des lumières infra-rouges se reflétant sur les pupilles de l'utilisateur, et une caméra infra-rouge qui capture cette reflexion. Ce système permet à la fois de capter les mouvements oculaires que l'on appelle saccade, les points de fixation, c'est-à-dire les zones sur lesquelles le regard s'arrête, et la dilatation pupillaire.

FIGURE 1.13 – Oculomètre tobii

Ce type de capteur peut être utilisé de deux manières différentes. La première est le suivi du mouvement des yeux, ce qui peut donner des informations sur le centre d'intérêt de la personne lors de la perception de l'émotion. Un stimulus visuel fournit un nombre varié d'informations à un instant donné, il est alors intéressant d'observer quelle zone occupe l'attention de l'utilisateur. Dans ce cadre là, l'oculomètre est utilisé pour éliminer des données parasites. Les informations que nous pouvons étudier sont les points de fixation essentiellement. Nous pouvons également étudier les saccades. En nombre élevé, cela peut révéler une incompréhension du stimulus.

Le deuxième aspect de l'étude des données de l'oculomètre est l'étude des émotions via la dilatation pupillaire. Cette modification est provoquée par des émotions [Partala 03] et d'une manière plus évidente par la modification de l'intensité lumineuse, mais aussi par l'augmentation de la charge de travail cognitif [Kahneman 66]. L'origine de cette réaction n'étant pas spécifique à l'émotion, l'interprétation de celle-ci est difficile.

Ce capteur peut donc se révéler très utile dans le cadre expérimental, pour la validation et la détection des émotions.

Rythme respiratoire

Le rythme respiratoire est un des capteurs plébiscités par l'*affective computing group* du MIT [Healey 98]. Ce capteur permet d'identifier le nombre de pulsations effectuées dans le temps (généralement 60 secondes). L'existence d'une relation entre la modification du rythme respiratoire et le changement émotionnel a été démontrée [Boiten 94, Homma 08]. Cependant, comme pour le rythme cardiaque les modifications respiratoires ne sont pas uniquement liées à l'émotion. Par exemple l'activité sportive peut être une cause d'augmentation du rythme cardiaque. Il reste néanmoins intéressant d'analyser ce signal particulièrement pour l'identification de l'éveil, qui est une des principales composantes de l'émotion.

5.3 Les réponses psychologiques

La seconde modalité d'expression des émotions que nous présentons est psychologique. Cette catégorie rassemble les expressions que nous utilisons pour transmettre ou détecter les émotions de notre entourage. Les deux principaux axes de recherches utilisant cette modalité se fixent sur les expressions faciales et vocales. Nous divisons l'expression vocale en deux sous-catégories qui sont la tonalité et le langage. L'utilisation de ces expressions pour la détection des émotions possède de nombreux avantages, mais aussi quelques difficultés. Parmi les avantages importants, nous pouvons identifier la facilité d'interprétation des différentes expressions. L'être humain est dans ce cas le meilleur annotateur possible du fait de son expérience acquise lors de la communication entre deux personnes. Cette facilité d'interprétation est contrebalancée par la facilité d'expression des émotions au travers de cette modalité. En effet, le jeu d'acteur est basé sur l'imitation des émotions au travers du visage et de la voix. Cette dernière caractéristique est intéressante dans le cadre expérimental, mais se révèle un indicateur non fiable dans une situation réelle.

Expressions faciales

Les expressions faciales représentent le vecteur de communication des émotions le plus évident. Nous l'utilisons au quotidien aussi bien pour transmettre nos émotions que pour identifier celles d'autrui. Le froncement des sourcils ou la présence d'un sourire sont des indicateurs facilement reconnaissables. Les travaux d'Ekman [Ekman 71] ont montré l'existence d'un sous-ensemble d'expressions universellement reconnu.

Les expressions faciales semblent être la source de données la plus facile et pratique à utiliser. En effet, une simple caméra peut être employée pour ce traitement. Cependant, cette modalité est à la fois la plus expressive, mais aussi la plus contrôlable. L'utilisateur peut alors facilement tromper le système. De plus, le contexte influe largement sur ce mode d'expressions, la vie en société empêche d'exprimer ses émotions sans retenue. Á contrario, cet inconvénient devient une qualité dans le cadre expérimental. Il est beaucoup plus facile de mimer une émotion au travers du visage, que d'elliciter une véritable émotion afin de capter son expression physiologique, par exemple.

Outre ces problématiques de falsification des expressions faciales, cette modalité reste largement exploitée dans le domaine de l'*informatique affective*. Á tort considérée comme une source de données facile à manipuler, la détection des émotions au travers du visage soulève de nombreuses difficultées. Il existe une grande variété d'expressions faciales pour une même émotion. La modification de l'intensité

émotionnelle, ainsi que les circonstances de l'émotion peuvent produire des expressions variées. Une description standardisée des expressions est proposée au travers de l'outil Facial Action Coding System (FACS) [Ekman 78].

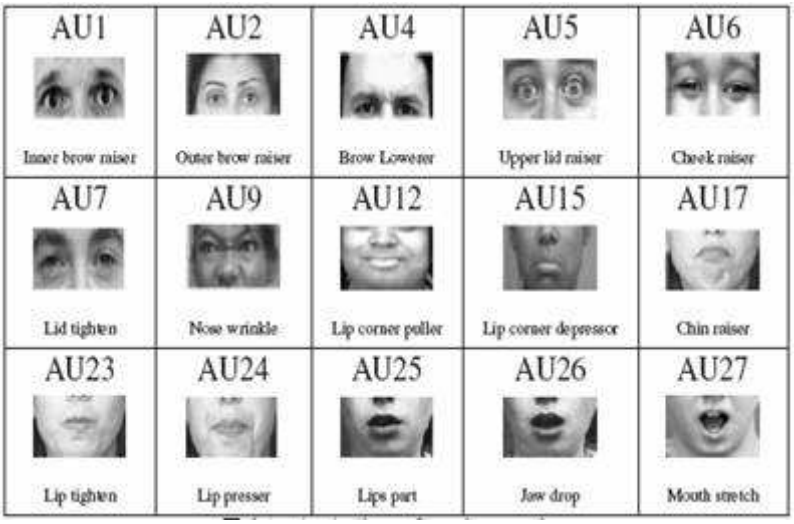


FIGURE 1.14 – Extrait de la base de données des unités d'actions de Kanade [Kanade 00]

Ce dernier permet, au travers de 46 unités d'expressions, de décrire l'ensemble des expressions faciales. Une unité d'action représente une modification locale d'un groupe musculaire du visage. Par exemple, la première unité d'action correspond à l'élévation de la partie interne des sourcils (au niveau du nez). L'observation des différentes unités d'actions permet d'identifier une émotion. Des applications de FACS pour l'analyse émotionnelle existent au travers de Emotional Facial Action Coding System (EMFACS) [Friesen 83] et Facial Action Coding System Affect Interpretation Dictionary (FACSAID) [9]. Il est alors possible de décrire une émotion en fonction des unités d'actions. La problématique informatique se limite alors au bon suivi des changements d'unités d'actions [Hamm 11]. Cependant, même si cette approche tente de faciliter l'interprétation des expressions faciales, le suivi de ces dernières n'est pas simple. Le visage est un vecteur d'émotion, mais il ne se limite pas à cette unique tâche. En effet, les mouvements de la tête, les changements de luminosité ou la parole rendent cette démarche ardue. Une voie autre que l'observation d'un ensemble de changements locaux est également envisagée. Les travaux de Dahmane et Meunier [Dahmane 11] proposent une analyse de la déformation globale du visage, afin d'extraire l'émotion par la classification. D'autres travaux [Sénéchal 13, Hoque 12] en revanche se concentrent sur certaines parties du visage, en particulier la bouche, afin de différencier le vrai sourire du ricannement par exemple.

[9] http://www.face-and-emotion.com/dataface/facsaid/description.jsp

Expressions vocales et textuelles

L'utilisation du son de la voix ou du texte est la seconde voie empruntée pour l'analyse de la modalité psychologique. La particularité ici est que ces deux signaux peuvent être produits au même moment. L'étude de ce signal se divise d'une part en fonction de caractéristiques phonétiques pour extraire les sons et prosodiques pour extraire la mélodie. Les aspects phonétiques sont souvent employés pour l'extraction de mots et les aspects prosodiques pour l'analyse des émotions. Cependant cette délimitation n'est pas aussi franche puisque certains travaux [Nicholson 00] utilisent ces deux caractéristiques pour la détection des émotions. Néanmoins, les outils d'analyses du texte sont très différents de ceux employés pour la mélodie.

L'interprétation émotionnelle d'une phrase s'appuie essentiellement sur des thésaurus [Bradley 99] ou ontologies [Mathieu 05]. En associant une émotion à un terme il est alors possible d'identifier l'émotion globale d'une phrase en fonction du sens donné au terme, par la syntaxe de la phrase. Ceci est compliqué puisque spécifique à la langue utilisée et au niveau de langage employé. En fonction de la syntaxe d'une phrase, un terme correspondant à une émotion négative peut décrire une émotion positive.

En revanche, l'analyse du son de la voie repose en grande partie sur l'extraction de caractéristiques prosodiques et phonétiques. Parmi celles-ci, nous pouvons citer la puissance vocale, la hauteur du son ou encore le débit vocal. Tout comme pour les expressions faciales et EMFACS des propositions existent pour décrire une émotion en fonction des variations de ces caractéristiques. Les travaux de Banse et Scherer [Banse 96] utilisent 18 paramètres acoustiques pour différencier 12 émotions. Une implémentation informatique de ces règles [Jones 08] a permis de mettre en évidence l'utilisabilité de celles-ci pour la détection des émotions. Néanmoins, il existe d'autres approches pour l'analyse de la voie, comme la classification [Nicholson 00], qui est un outil largement exploité en *informatique affective*.

5.4 Les réponses comportementales

Cette dernière modalité correspond à l'expression des émotions au travers du comportement que l'on adopte face à une situation. Le plus souvent représentée par la posture du corps, cette modalité est faiblement employée en raison de l'espace requis pour son expression. Néanmoins, les postures adoptées lors de l'expérience émotionnelle sont fortements liées à la notion de tendance à l'action [Frijda 87]. Par exemple, dans le cadre d'une émotion dominante (colère) une position d'avancement est prise par le sujet. De plus, il est facile d'observer cette modalité grâce à une caméra. Plus récemment, de nouveaux capteurs à bas coût ont été développés facilitant

cette observation, tels que la kinect de Microsoft. Il existe cependant des travaux la concernant [Gunes 07]. Ces derniers utilisent uniquement la partie supérieure du corps pour obtenir l'information émotionnelle.

5.5 Positionnement

Comme nous l'avons montré dans cette section, il existe une grande variété de capteurs émotionnels, chacun fournissant un ou plusieurs aspects de l'émotion ressentie. De plus, il n'existe pas un ensemble de règles permettant d'identifier une émotion en fonction d'un ou plusieurs capteurs.

En se basant sur les théories émotionnelles précédentes, notre cerveau utilise l'ensemble de ces retours afin d'identifier le type d'émotion éprouvé. Il semble donc raisonnable de penser que l'ensemble de ces expressions a un impact sur la détermination du ressenti. Cependant, il est inenvisageable d'appliquer l'ensemble de ces capteurs sur un individu dans un contexte réel. Le choix de l'utilisation d'un capteur plutôt qu'un autre est fortement lié à l'environnement de l'expérimentation ou de l'application. C'est pourquoi nous proposons de laisser ce choix aux concepteurs afin d'adapter notre système à ses propres considérations. Nous spécifierons dans la suite du document certaines limitations quant à l'utilisation de capteurs.

6 Observation et validation des émotions

L'émotion est à la fois un objet éphémère, subjectif et impalpable. Ces trois caractèristiques rendent l'étude de ce phénomène difficile. Dans un cadre expérimental de détection, l'objectif est d'une part observer un objet connu via un système automatisé pour obtenir la nature de l'objet. La comparaison entre la nature réelle et celle détectée permet alors d'identifier si le détecteur est fonctionnel. Par exemple, l'analyse d'images satellitaires permet d'identifier la nature des terrains. La détection consiste alors à appliquer une détection sur une image et d'identifier le terrain photographié. Il est ensuite possible de se rendre à cet emplacement pour confirmer ou non la présence d'un terrain de ce type. Dans le cadre de l'émotion, cette procédure n'est pas aussi évidente. Il faut alors dissocier deux aspects. D'une part, nous devons elliciter une émotion chez une personne puis nous devons nous assurer que l'émotion ellicitée est bien celle ressentie par le sujet. Une fois ces deux étapes accomplies, il est alors possible de comparer les résultats d'un détecteur d'émotion avec ceux récoltés manuellement. Cependant cette démarche nécessite un protocole expérimental complexe, afin de nous assurer de la bonne concordance entre l'émotion souhaitée

et l'émotion ressentie. En psychologie, des travaux [Coan 07] se sont penchés sur la standardisation de méthodologies permettant de répondre à ces problématiques.

6.1 Ellicitation des émotions

Cette première étape a pour but de faire ressentir à un individu une émotion précise. Une première méthode simple est de demander au sujet de mimer une émotion. Cette démarche comporte certaines restrictions. Il est nécessaire d'une part de bien connaître l'expression émotionnelle afin de mimer cette dernière de manière naturelle (FACS), et d'autre part, toutes les réponses corporelles ne sont pas simulables. En effet, la modification du rythme cardiaque ou la sudation de la peau ne sont pas des réactions facilement observables dans le cadre du jeu d'un acteur. Cependant, il existe certaines procédures permettant de faire ressentir à un sujet acteur ou non une émotion. La présentation d'une succession d'images, sons ou vidéos avec une charge émotionnelle proche, peut plonger le sujet dans un certain état émotionnel. Il existe pour cela des bases de données d'images (International Affective Picture System (IAPS) [Lang 05]), de sons (International Affective Digitized Sounds (IADS) [Bradley 07]) ou de vidéos [Gross 95] permettant de faire ce travail. Il reste cependant difficile de faire la différence entre l'ellicitation d'une *émotion pénétrante* d'une *émotion épisodique* [Cowie 05].

6.2 Identification de l'émotion

Une fois l'émotion ressentie, il est nécessaire d'identifier si celle-ci correspond à l'émotion souhaitée initialement. Le questionnement est la solution la plus couramment utilisée. Cependant, de par les confusions existantes entre les différents types d'émotion, un questionnement simple sur le ressenti du sujet n'est pas envisageable. Un des questionnaires le plus souvent utilisé est Self-Assessment Manikin (SAM) [Bradley 94]. Celui-ci utilise des pictogrammes répartis sur une échelle d'intensité, afin de caractériser l'émotion suivant trois axes : la valence, l'éveil et la dominance. Ce type de questionnaire peut également être complété par l'observation du sujet. Une personne entrainée à l'identification des émotions peut alors pondérer les résultats obtenus par son observation.

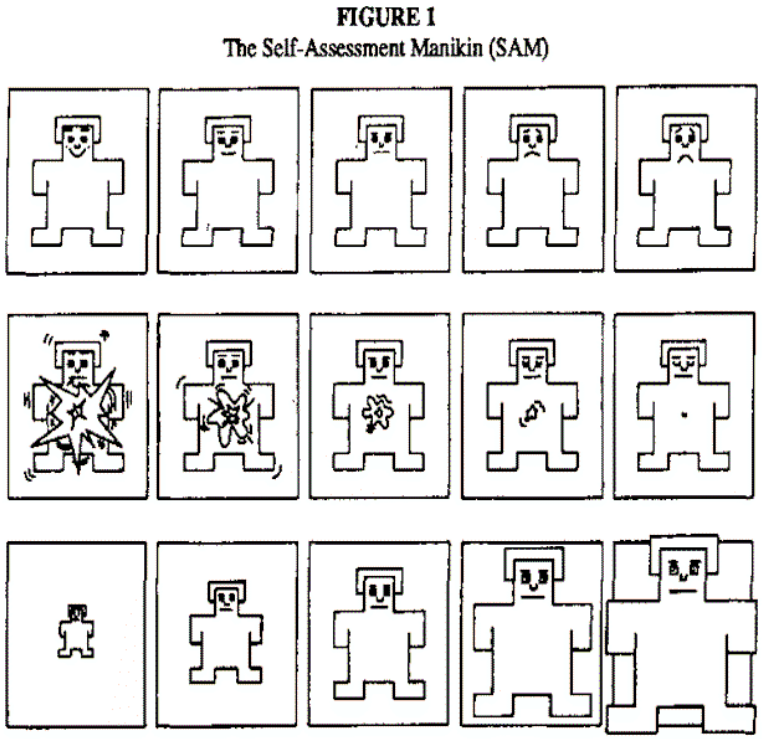


FIGURE 1.15 – Représentation des trois échelles du questionnaire SAM

7 Les détecteurs émotionnels

Le domaine de l'*informatique affective* a permis le développement de nombreux détecteurs des émotions. Nous pouvons distinguer deux grandes catégories, les systèmes à une modalité et les systèmes multimodaux. Le premier type de modèle vise à exploiter les données d'un capteur de manière optimale. La difficulté ici étant qu'un seul capteur peut fournir des données erronées ou des tentatives de dissimulation de l'émotion provoquant des erreurs de détection. Le second type de système, en revanche propose de combiner des sources variées d'informations afin d'affiner les détections. Il devient donc difficile de masquer l'émotion si le nombre de modalités est suffisant, mais cela peut engendrer une dégénéréscence de la capture. Une autre difficulté se situe au niveau de la complexité et de la rapidité d'exécution, qui augmente avec la quantité de modalités observées.

Dans cette partie, nous présenterons donc des détecteurs de ces deux catégories afin de montrer les différences et les tendances de chacun des domaines. Cependant, nous ne proposerons pas de comparaison de performances de ces systèmes. Dans le domaine de l'*informatique affective*, il n'existe pas encore de base de données standardisée pour comparer les systèmes, même si des propositions commencent à

émerger. De plus, les émotions détectées ne sont pas forcément comparables suivant le type d'émotions [Cowie 05] ou la variété considérée.

7.1 Présentation d'une partie des détecteurs existants dans le domaine

Le tableau 1.3, présente une comparaison d'un éventail représentatif des détecteurs émotionnels. Nous proposons un résumé des principaux critères de caractérisation des détecteurs, à savoir les types de signaux utilisés, la méthodologie employée pour l'interprétation de ces derniers, ainsi que les émotions observées.

Travaux	Type	Signaux	Méthodologie	Emotions
		Systèmes multimodaux		
[Haag 04]	Multimodal	EMG, GSR, température de la peau, BVP, ECG, rythem respiratoire	Réseau de neuronnes	Eveil / valence
[Metri 12]	Multimodal	Expressions faciales et gestuelles	classification bayésienne	neutre, surprise, peur, joie, dégoût, tristesse
[Picard 01]	Multimodal	EMG, BVP, GSR, rythme respiratoire	SFFS+FP et K-nearest neighbors (KNN)	neutre, colère, haine, chagrin, amour platonic, amour, joie et révérence
[Takahashi 04]	Multimodal	rythme cardiaque, EEG et GSR	SVM	Joie, colère, tristesse, peur et relaxation
[Chanel 05]	Multimodal	EEG, rythme respiratoire, température de la peau, GSR et pression sanguine	FDA	Eveil / valence

Travaux	Type	Signaux	Méthodologie	Emotions
[Gunes 07]	Multimodal	Visage et corps	Classifieur Bayesien	Anxiété, colère, dégoût, peur, joie et incertitude
Systèmes unimodaux				
[Li 09]	Unimodal	EEG	SVM et CSP	Joie et tristesse
[Anttonen 05]	Unimodal	Rythme cardiaque	Observation du rythme cardiaque en fonction de la valence du stimulus	Positive, négative et neutre
[Bos 06]	Unimodal	EEG	Classifieur FDA	Eveil / valence
[Jones 08]	Unimodal	Voix	Statistique	Ennui, tristesse, colère, joie et surprise
[Nicholson 00]	Unimodal	Voix	Réseaux de neuronnes	Joie, peur, tristesse, dégoût, colère, surprise, neutre et malicieux
[Mathieu 05]	Unimodal	Texte	Ontologie	Irritation, peur, tristesse, amour, fascination, étonnement
[Palomba 97]	Unimodal	rythme cardiaque	statistique	Plaisant, déplaisant et neutre

TABLE 1.3 – Tableau comparatif d'un panel représentatif des détecteurs d'émotions

Au travers de ce tableau, nous avons voulu représenter les principales voies empruntées pour détecter les émotions. Nous pouvons observer qu'il existe une grande

diversité des signaux observés même si la tendance est à l'observation des expressions physiologiques. En particulier, les travaux les plus récents utilisent de plus en plus l'EEG en complément d'autres capteurs physiologiques. Cependant, toutes les modalités d'expressions des émotions ont été exploitées dans ces travaux. Les performances avancées dans ces différents articles bien que difficilement comparables, ne montrent pas de prévalence d'un certain type de capteurs. Nous pouvons également déduire de ce tableau que les systèmes unimodaux sont souvent développés dans le but de prouver l'utilité du capteur employé ou de déduire certains corportements de la modalité observée. Par exemple, les travaux effectués sur le rythme cardiaque [Anttonen 05] montrent la différence de décélaration du rythme cardiaque en fonction de la valence d'un stimulus. En revanche, les systèmes multimodaux sont voués à la détection et à l'évaluation des méthodologies employées.

Concernant les méthodologies employées, nous pouvons dire qu'il existe un fort angouement pour la classification. Une telle approche se justifie par l'aspect non spécifique des réponses émotionnelles et la variété des signaux à observer. L'étude individuelle des capteurs n'ayant pas vu apparaître l'émergence de schémas ou de règles permettant l'identification des émotions, une approche par apprentissage reste une bonne solution pour effectuer la détection. Nous pouvons également remarquer que la classification est souvent précédée par une phase d'extraction de caractéristiques et de réduction de ces dernières par l'utilisation d'outils tels que la projection de Fisher. Il existe néanmoins des approches dédiées à certains capteurs. Par exemple, l'utilisation d'une ontologie est plus adaptée dans le cadre d'une détection des émotions par le texte [Mathieu 05].

7.2 Positionnement

A partir de l'observation des détecteurs existants, nous avons convenu que la détermination d'un jeu de capteurs fixes est difficilement envisageable. En effet, les propositions effectuées par la communauté ne montrent pas un ensemble de capteurs plus performants que les autres. De plus, le choix d'un capteur est très dépendant du contexte applicatif visé. Nous avons donc opté pour concevoir un système qui puisse s'abstraire du nombre et du type de capteurs utilisés. Nous avons cependant voulu nous rapprocher de ce qui existe en nous positionnant sur une détection par régression, qui est une classification dans l'espace continu. L'extraction de caractéristiques est pour nous une considération liée au capteur, c'est pourquoi nous proposons de dissocier cette étape de l'interprétation des signaux, dans le but de concevoir un système indépendant des capteurs.

8 Conclusion

Dans ce chapitre, nous avons commencé par fournir un ensemble de connaissances issues du domaine psychologique afin de cerner la notion d'émotion. Pour ce faire, nous avons commencé par une analyse des différentes définitions proposées par le domaine afin de sélectionner les définitions de Kleinginna et Cowie largement utilisées dans le domaine. Nous avons également montré les caractéristiques émotionnelles permettant de distinguer les émotions de notions connexes engendrant de nombreuses ambiguités.

Nous avons ensuite continué par présenter les principales théories émotionnelles par ordre chronologique afin de montrer l'évolution de la compréhension du processus émotionnel. La description de ces principaux courants de pensée nous a amené à identifier les théories auxquelles nous adhérions, à savoir les théories cognitives telles que la théorie des deux facteurs de Schachter et Singer ou la théorie CPM de Scherer. Ensuite, pour nous rapprocher du domaine de l' *informatique affective*, nous avons montré les modèles émotionnels utilisés par les psychologues, ainsi que les tentatives proposées en informatique à travers les ontologies ou des modèles plus théoriques comme celui de Scherer.

Nous avons terminé ce chapitre par la présentation des principaux capteurs utilisés dans le domaine, ainsi que les détecteurs dans lesquels ils sont utilisés, en distinguant les approches unimodales des approches multimodales.

Chapitre 2

Un modèle en trois couches pour la détection des émotions

L'objectif que nous visons est de proposer un modèle pour la détection des émotions modulaire et flexible pour faciliter l'intégration dans un système nécessitant une reconnaissance en temps réel. Nous présenterons donc notre modélisation divisée en modules interopérables et interchangeables. Nous n'avons pas la prétention d'aboutir à un détecteur des émotions parfait, c'est pourquoi nous souhaitons que notre système puisse contenir des modules interchangeables permettant ainsi d'améliorer la qualité de détection.

L'architecture globale de notre système repose sur la théorie des deux facteurs car c'est la première à définir la dualité expression et contexte dans le processus émotionnel. Cette théorie propose qu'après la perception d'un stimulus, la personne produit une expression corporelle reflexe, ensuite, un processus cognitif est déclenché permettant d'évaluer l'émotion ressentie. La difficulté ici est de pouvoir exploiter à la fois les expressions corporelles mais aussi le processus cognitif pour obtenir une interprétation émotionnelle la plus fidèle. La figure 2.1 reprend le schéma de la théorie des deux facteurs complété par un observateur externe qui peut être une autre personne mais aussi un système informatique. Un observateur humain a typiquement accès aux expressions émotionnelles (B) pour déterminer l'émotion d'une personne. Nous proposons ici d'exploiter deux autres informations (A et C) afin d'approcher l'émotion réellement ressentie. Le chapitre 3 présente l'exploitation des expressions corporelles (B) alors que le chapitre 4 présente l'exploitation du stimulus et du processus cognitif (A et B).

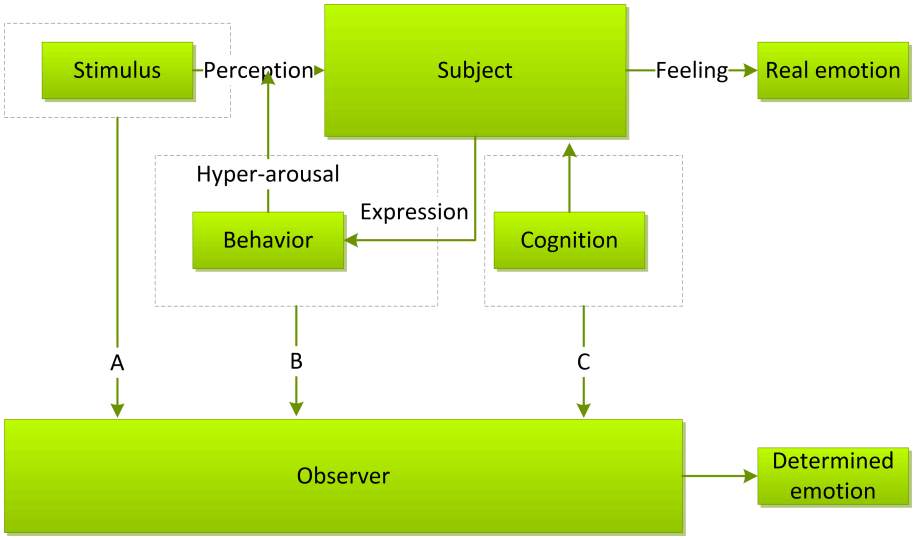


FIGURE 2.1 – Présentation des différentes sources d'informations émotionnelles à partir de la théorie des deux facteurs.

Nos recherches sur l'état de l'art nous ont poussés à diviser notre modèle en trois modules distincts. Chaque module ayant une tâche indépendante et étant centré sur une spécialité précise, sera interopérable afin de fournir une information émotionnelle à des degrés d'abstraction de plus en plus faible. Basé sur la théorie des deux facteurs et sur notre hypothèse selon laquelle nous pensons comme un modèle dimensionnel et nous nous exprimons comme un modèle catégoriel, notre modèle se divise en couches successives reprenant ce que nous pensons être le fonctionnement humain [Fig. 2.2].

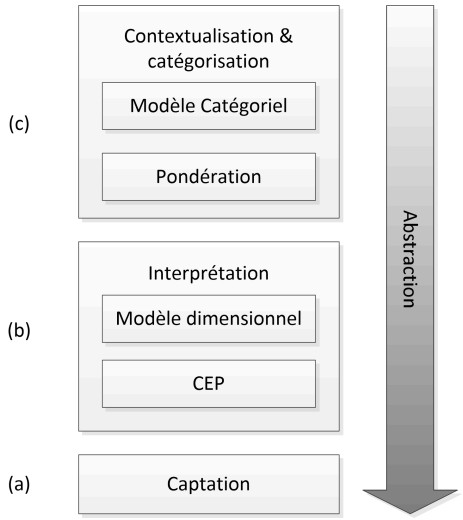


FIGURE 2.2 – Vue d'ensemble de notre modèle en trois couches

Un premier module (a) aura pour objectif de traiter les données issues des cap-

teurs. Cette partie de traitement du signal ne permet pas l'émergence d'une information sur l'émotion ressentie mais des indices sur l'expression de celle-ci. La partie centrale (b) du système correspond à ce que nous appelons CEP ; ce sont ces cartes que nous utiliserons pour la reconnaissance des émotions ainsi que l'apprentissage. L'information issue de cette partie est une annotation dans un modèle dimensionnel. Le dernier module (c), basé sur une ontologie, permet, d'une part, de définir des intervalles émotionnels de notre espace dimensionnel, afin d'associer un terme émotionnel à l'annotation brute issue des CEPs (catégorisation). D'autre part, ce module nous permet également d'utiliser le contexte de l'expérience émotionnelle pour pondérer l'annotation brute. Ceci est réalisable grâce aux capacités d'inférences de l'ontologie. Ce choix d'implémentation est directement lié à notre approche cognitiviste de l'émotion. Cette dernière définie que la connaissance et le contexte du sujet sur le stimulus auquel il est soumis, est une phase importante de l'expérience émotionnelle.

Dans ce chapitre, nous détaillerons l'architecture globale, ainsi que les interactions entre ces différents modules, permettant d'aboutir à l'émotion.

1 De la théorie des deux facteurs au modèle informatique

La réalisation d'un modèle informatique des émotions est sujette à de nombreuses variantes. Le nombre de théories et de modèles existants dans le domaine de la psychologie empêche la définition d'un standard admis par tous. Nous avons donc dû effectuer des choix quant aux connaissances qui devaient être la base théorique de notre modélisation. Nous avons pour cela sélectionné des définitions générales du phénomène fixant ainsi certaines caractéristiques sur l'objet que nous voulons détecter, une théorie décrivant le processus d'apparition de l'émotion et enfin un modèle de description de l'émotion.

Dans cette section, nous allons décrire de quelle façon nous avons créé notre modèle en fonction des choix précédents. Comme indiqué dans le chapitre précédent, le point de départ de nos travaux est la théorie des deux facteurs de Schachter et Singer. Cette dernière propose un découpage du processus émotionnel en deux phases interconnectées. D'une part, comme dans les théories classiques, l'expression des émotions est une réponse issue du cerveau reptilien, permettant ainsi le déclenchement réflexe d'une réponse face à un stimulus. Cette première phase difficilement contrôlable est traduite comme un état d'hyper-éveil amorçant la phase de cognition dont le déroulement est plus long de par sa complexité. Cette seconde étape

utilise les informations issues des réponses émotionnelles couplées au contexte afin d'identifier l'émotion ressentie.

L'utilisation du contexte comme information pour la détermination de l'émotion nous a semblé pertinent du fait de la non spécificité des réponses émotionnelles. Cette information difficilement captable est sous exploitée dans les travaux de détection des émotions. De plus, cette théorie est à l'origine des théories modernes sur les émotions et semble admise par de nombreux psychologues.

Nous avons donc opté pour l'utilisation de cette théorie afin de calquer notre modèle de détection des émotions sur le comportement humain. Nous avons souhaité réaliser une modélisation couvrant l'ensemble du processus de détection depuis l'expression émotionnelle jusqu'à l'expression verbale de l'émotion. Nous pensons en effet, que la phase d'expérience émotionnelle est distincte de la phase de verbalisation de par l'utilisation d'un vocabulaire émotionnel. L'utilisation d'un tel outil implique une discrimination faible des émotions. Nous émettons donc l'hypothèse selon laquelle l'être humain ressent une variété infinie d'émotions mais les exprime via un nombre de termes limité, éliminant ainsi un grand nombre de nuances.

Nous proposons ainsi une détection en deux phases, une première permettant d'extraire une émotion dans un espace continu grâce à l'analyse de l'expression émotionnelle et une seconde permettant d'y associer un terme du vocabulaire émotionnel pondéré par le contexte.

1.1 Architecture générale

La détection des émotions est un domaine fortement interdisciplinaire, c'est pourquoi, nous avons divisé notre réalisation en 3 modules interconnectés et interchangeables. L'intérêt y est double car la division de la détection permet d'obtenir des informations émotionnelles à plusieurs niveaux d'abstraction, mais également de pouvoir modifier un ou plusieurs modules afin de l'adapter au contexte de l'application réalisée. Nous identifions donc trois couches distinctes représentées sur la figure 2.3 qui sont :

- le module de traitement du signal spécifique à chacun des capteurs
- le module de détection et annotation bas niveau appelé CEP
- le module de catégorisation et contextualisation des annotations brutes

Cette division est basée sur les connaissances issues de la psychologie, mais également, définie en fonction des domaines de connaissances impliqués dans ces modules.

Celui de plus bas niveau se concentre sur le traitement du signal, le module de reconnaissance est basé sur la classification et l'analyse de données, et enfin, le dernier, repose sur l'ingénierie des connaissances.

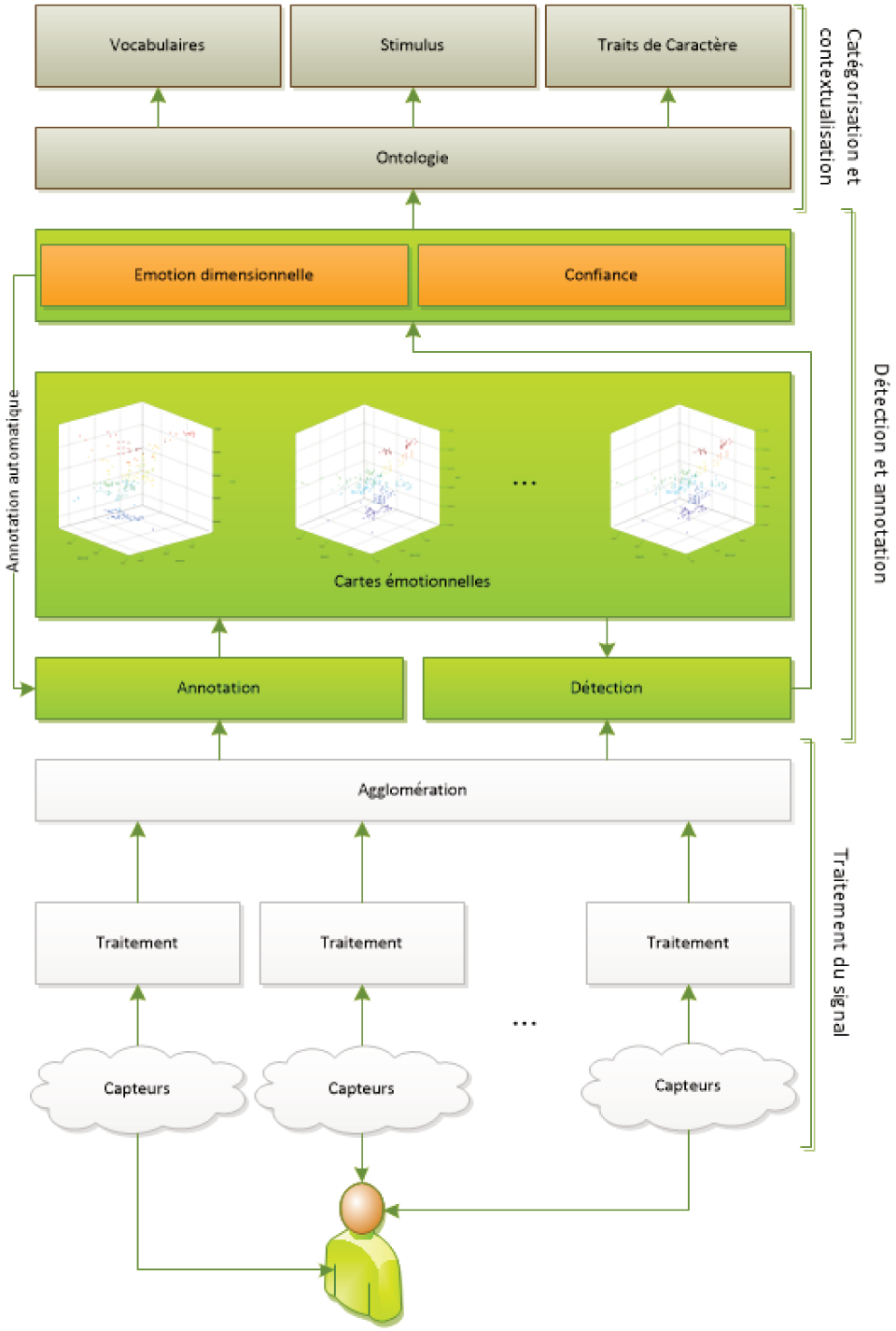


FIGURE 2.3 – Architecture globale du système de détection des émotions

Module de traitement du signal

Le module de plus bas niveau que nous définissons correspond à la phase de traitement des données captées. L'identification des capteurs et le traitement permettant d'extraire des informations exploitables, est un travail important pour détecter les émotions. Ce travail particulièrement complexe et spécifique à chacun des capteurs utilisés demande des connaissances importantes dans le domaine du traitement de signal. La création d'un tel module est nécessaire lors de l'ajout d'un nouveau capteur et comprend l'ensemble du traitement depuis le pilote logiciel permettant de communiquer avec le capteur, jusqu'à l'exposition des données sous forme de valeurs numériques représentant la caractéristique visée. L'exemple que nous avons sélectionné pour nos travaux est celui de la caméra et des expressions faciales. Dans celui-ci, nous avons donc developpé le traitement permettant d'extraire de chaque image huits points caractéristiques des expressions, à savoir, différentes zones représentant la position et la forme de la bouche, ainsi que des sourcils. Nous normalisons ensuite ces informations dans un intervalle $[0\,;1]$ décrivant le déplacement de ces points.

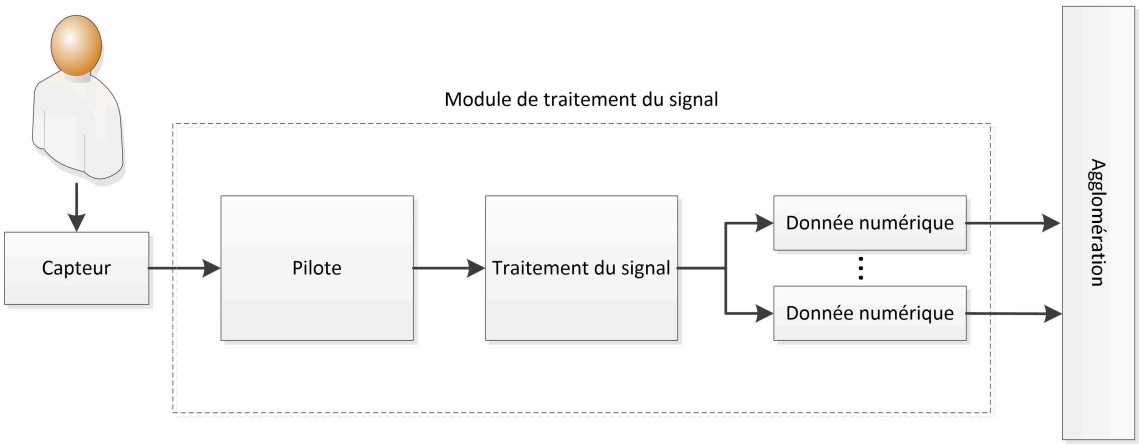


FIGURE 2.4 – Détail du module de traitement du signal

Module de détection et d'annotation

L'objectif pour ce second module est de faire le lien entre les données captées, exposées par le module précédent et une annotation dans un espace continu représentant une émotion. Dans cette première étape, seules les expressions émotionnelles sont utilisées pour identifier l'émotion ressentie. La réalisation de cette phase s'inscrit dans une démarche de régression qui est une transposition de la classification dans l'espace continu. Nous utilisons à la fois un algorithme de détection inspiré de l'algorithme de classification K-nearest neighbors (KNN) [Beyer 99] et une base d'apprentissage constuée des expressions annotées émotionnellement. Dans notre

système, la base d'apprentissage est représentée par les CEPs et l'algorithme de détection s'éxécute grâce à ces dernières. Pour cette première étape de la détection, nous souhaitons décrire les émotions dans un espace continu afin de nous conformer à notre hypothèse initiale sur la cohabitation des modèles dimensionnels et catégoriels. C'est pourqoi nous avons modifié l'algorithme KNN pour faire de la régression. Cette dernière étant une forme de classification dans le domaine continu, ainsi nous ne souhaitons pas prédire une classe précise pour un état émotionnel, mais prédire une valeur numérique pour celui-ci.

Nous distinguons deux temps d'éxécution pour ce module. Le premier temps correspond à l'apprentissage, nous permettant ainsi de générer les CEPs nécéssaires à la détection. Cette phase doit être réalisée en amont de la détection, afin d'insérer des données émotionnelles représentatives. Le second temps d'éxécution est la détection en temps réel de l'émotion, basée sur les mêmes capteurs que la phase précédente. Il est également possible d'utiliser un sous-ensemble de capteurs en effectuant une projection de nos cartes sur les capteurs utilisés lors de la détection. Cette possibilité permet d'améliorer la robustesse de notre système, en prévenant des pannes matérielles par exemple. La détection sera impactée en réduisant potentiellement la fiabilité de la détection. Nous utilisons également les émotions détectées pour affiner nos CEPs grâce à la notion de confiance de la détection, que nous détaillerons dans le chapitre 3. De plus, l'algorithme de détection utilisé s'appuie sur l'émotion précédemment détectée afin d'identifier la suivante. Ceci nous permet de stabiliser la détection car il n'est pas possible de changer d'émotion instantannément. Le résultat final de ce module est l'exposition d'une représentation d'une émotion dans un espace émotionnel dimensionnel basé uniquement sur l'expression de l'émotion.

Pour notre prototype, nous utilisons deux cartes dans le but de décrire les émotions dans l'espace éveil/valence. La construction de celles-ci se base sur une vidéo où une personne exprime différentes émotions. Chaque image de cette vidéo permet de capturer une expression que l'on annote manuellement, afin de l'ajouter dans nos cartes. A partir de l'ensemble des émotions exprimées (ici les six émotions de Ekman), nous pouvons exécuter notre algorithme de détection. Une fois les données captées agglomérées, nous calculons les composantes de l'émotion (éveil ou valence) sur chacune des cartes à partir de l'expression faciale et de l'émotion précédente. Ce calcul nous fournit d'une part une nouvelle annotation émotionnelle dimensionnelle, mais également une information sur la confiance que l'on peut porter sur cette détection. C'est cette dernière information qui nous permettra d'ajouter ou non ce nouvel état au sein de notre base d'apprentissage, en fonction d'un seuil de confiance défini par l'utilisateur.

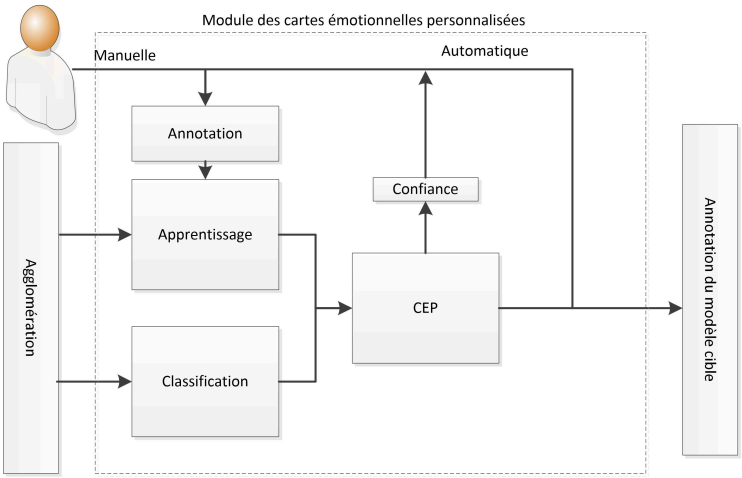


FIGURE 2.5 – Détail du module de CEP

Module de catégorisation et de contextualisation

La dernière étape de notre processus d'extraction des émotions est composée d'un double objectif. La catégorisation et la contextualisation sont deux étapes indépendantes visant des objectifs distincts, mais réalisées par le même module. La contextualisation doit être réalisée en amont de la catégorisation pour obtenir le terme le plus adapté à l'émotion ressentie.

L'étape de contextualisation est réalisée à partir des annotations brutes fournies par les CEPs. Le but ici, est de prendre en compte le contexte de l'experience émotionnelle pour pondérer l'annotation finale proposée par le système. Cet objectif découle directement des théories cognitivistes sur les émotions définissant le contexte comme élément de catégorisation des émotions. Dans notre cas, nous utilisons des informations sur les traits de caractères stockés dans notre base de connaissances. La comparaison entre le stimulus proposé à l'utilisateur et les traits de caractères pouvant correspondre, détermine la pondération à appliquer sur l'annotation initiale. Dans un deuxième temps, la catégorisation utilise cette nouvelle annotation pondérée afin d'identifier le terme de notre vocabulaire émotionnel le plus adapté à cette dernière. Par exemple, une personne arachnophobe ressentira une émotion nettement plus intense face à une araignée qu'une personne ne l'étant pas. C'est cette connaissance tangible et durable que nous souhaitons modéliser afin de fournir une information aussi proche que possible du ressenti réel de l'utilisateur.

L'ensemble des informations utilisées dans ce traitement est dissocié de notre modélisation basée sur une ontologie. Les traits de caractères et leur impact sur l'émotion ressentie sont définis par l'utilisateur, pouvant ainsi modifier l'intensité de son trait de caractère. D'un autre côté, la traduction des annotations continues en

annotation discrète doit être établie par des spécialistes du domaine en fonction du modèle émotionnel choisi pour la génération des CEPs.

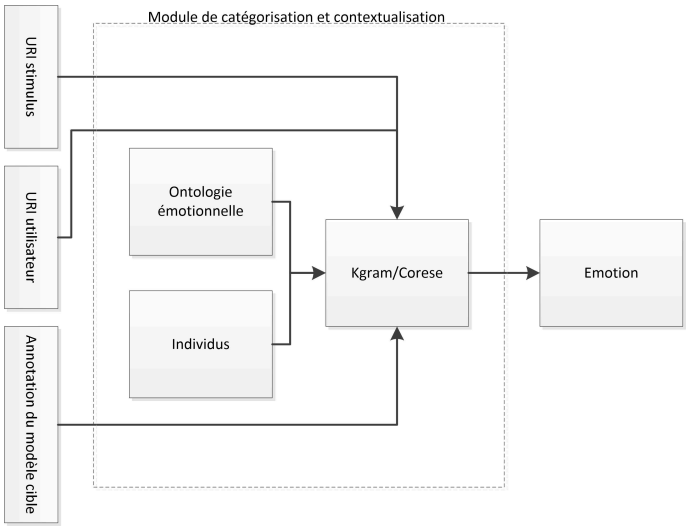


FIGURE 2.6 – Détail du module de catégorisation et contextualisation

2 Présentation du prototype

Dans cette section, nous présenterons le prototype d'application que nous avons réalisé à des fins expérimentales. L'objectif final étant de masquer le fonctionnement interne de notre détection pour proposer un outil clé en main aux développeurs souhaitant intégrer la notion d'émotion dans leur application.

Le prototype que nous présentons est divisé en sous-applications indépendantes suivant le découpage de notre modèle. Cette démarche a pour ambition d'éprouver les capacités de modularité de notre système.

Nous commencerons tout d'abord par le module de plus bas niveau utilisant une caméra comme capteur. La partie de traitement du signal étant une solution ad'hoc pour notre système, nous détaillerons son fonctionnement dans cette partie. L'implémentation que nous en proposons repose sur des outils sensibles au changement de luminosité et de position du corps ce qui rend son utilisation difficile dans un cadre réel. Nous souhaitons néanmoins le présenter pour fournir une bonne compréhension des données que nous traiterons par la suite et des contraintes que nous avons précisés pour la création de modules similaires.

Nous présenterons ensuite les deux principaux modules de notre détecteur sans spécifier les détails techniques de leurs mises en oeuvre, puis nous parlerons d'un dernier module d'exploitation d'une caractéristique de nos cartes émotionnelles visant à synthétiser les expressions faciales.

2.1 Module de capture des expressions faciales

Comme défini précédemment, l'acquisition des données est dissociée de l'interprétation de celles-ci. Le module que nous présentons ici est une implémentation utilisant la caméra comme source d'information. L'utilisation d'un tel capteur se justifie pour deux raisons. D'une part, nous favorisons l'utilisation de capteurs à faible coût et facilement accessibles, ce qui est le cas des caméras. D'autre part, l'expression faciale est un vecteur émotionnel facilement contrôlable dans le cadre d'un expérimentation. Nous utilisons cet unique capteur dont nous extrayons plusieurs informations afin de caractériser l'expression de l'utilisateur.

La figure 2.7 présente l'interface de contrôle du module. Nous pouvons observer sur l'image de la caméra la détection des régions d'intérêts. Nous utilisons la librairie OpenCV [1] pour ses fonctionnalités de reconnaissance des visages, ainsi que les outils de traitement d'images qu'elle fournit.

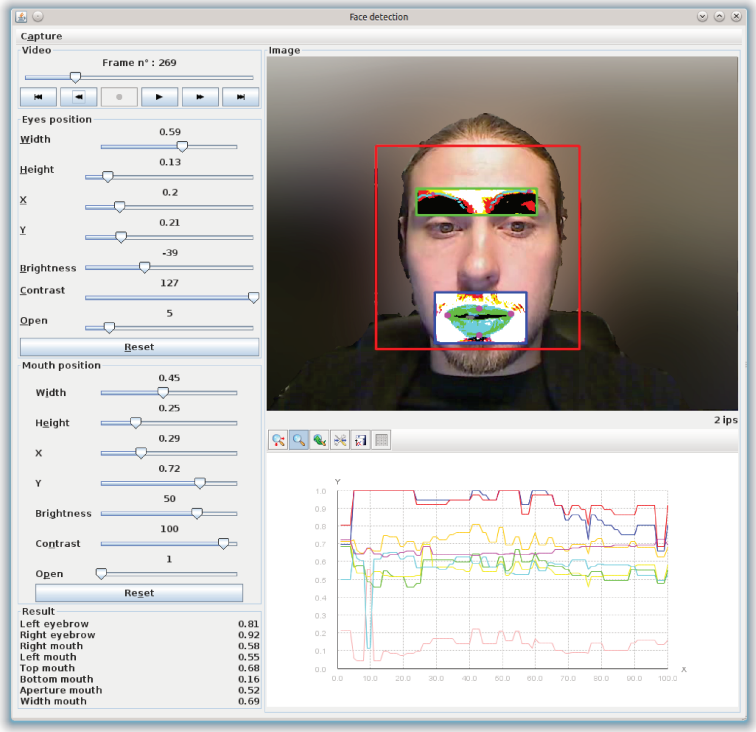


FIGURE 2.7 – Interface graphique du module de capture des expressions faciales

Le carré rouge encadrant le visage est déterminé par l'utilisation des descripteurs de Haar intégrés à la librairie. Cet outil permet de facilement repérer un visage sur une image et grâce au temps d'exécution relativement faible, nous pouvons l'utiliser

[1] http://opencv.org/

sur une séquence vidéo pour faire de la détection en temps réel. A partir de la zone du visage, nous configurons les régions de la bouche et des sourcils afin d'exécuter nos algorithmes de détection de points d'intérêts pour les expressions faciales. La position de ces deux régions est configurable dans la partie de gauche et est exprimée en pourcentage du visage (zone rouge). Cette démarche à pour but de donner un peu de flexibilité lors de la détection en cas de déplacement du visage.

Une fois ces deux régions définies, nous appliquons un traitement permettant de repérer les contours de la bouche ou des sourcils. Nous utilisons la différence de coloration de ces différentes parties et les variations de contraste et de luminosité afin de détacher ces éléments du reste du visage. Sur l'image nous pouvons voir les sourcils ressortir en noir et la bouche en vert. Pour stabiliser cette détection, nous utilisons également l'algorithme de morphisme qu'est l'ouverture, afin d'éliminer les pixels isolés. La figure 2.8 présente l'application de notre processus d'extraction des contours sur la bouche. Nous pouvons voir que sur cette image seuillée pour conserver uniquement les pixels verts, les pixels blancs représentent la bouche. A partir de cette image, nous utilisons les caractéristiques de forme de la bouche pour en extraire les points caractéristiques.

FIGURE 2.8 – Extraction des contours de la bouche

À partir de la camera, nous identifions six points caractéristiques pour la bouche dont nous souhaitons suivre l'évolution :

- Les commissures de la bouche : extrait en recherchant les points horizontalement les plus extérieurs de la zone d'intérêt.
- Les bords supérieurs et inférieurs : extrait en recherchant les points verticalement les plus extérieurs.
- La distance entre les commissures de la bouche ce qui représente la largeur de la bouche.
- La distance entre les bords supérieurs et inférieurs ce qui représente l'ouverture de la bouche.

Pour les sourcils, nous avons sélectionné une information par sourcil, à savoir le déplacement du point le plus haut de ces derniers.

Ce module permet également de visualiser les informations détectées de manière graphique ou numérique. Ces informations représentent la position des points caractéristiques en pourcentage de la région d'intérêt dans laquelle ils ont été extrait.

Enfin, ce module permet d'enregistrer en temps réel les images de la caméra en appliquant les traitements sur les images et de rejouer ces séquences à posteriori.

2.2 Module de génération des CEPs

Ce second module a pour but la gestion et la visualisation de l'apprentissage et de la classification des données captées. La figure 2.9 présente l'interface de ce module.

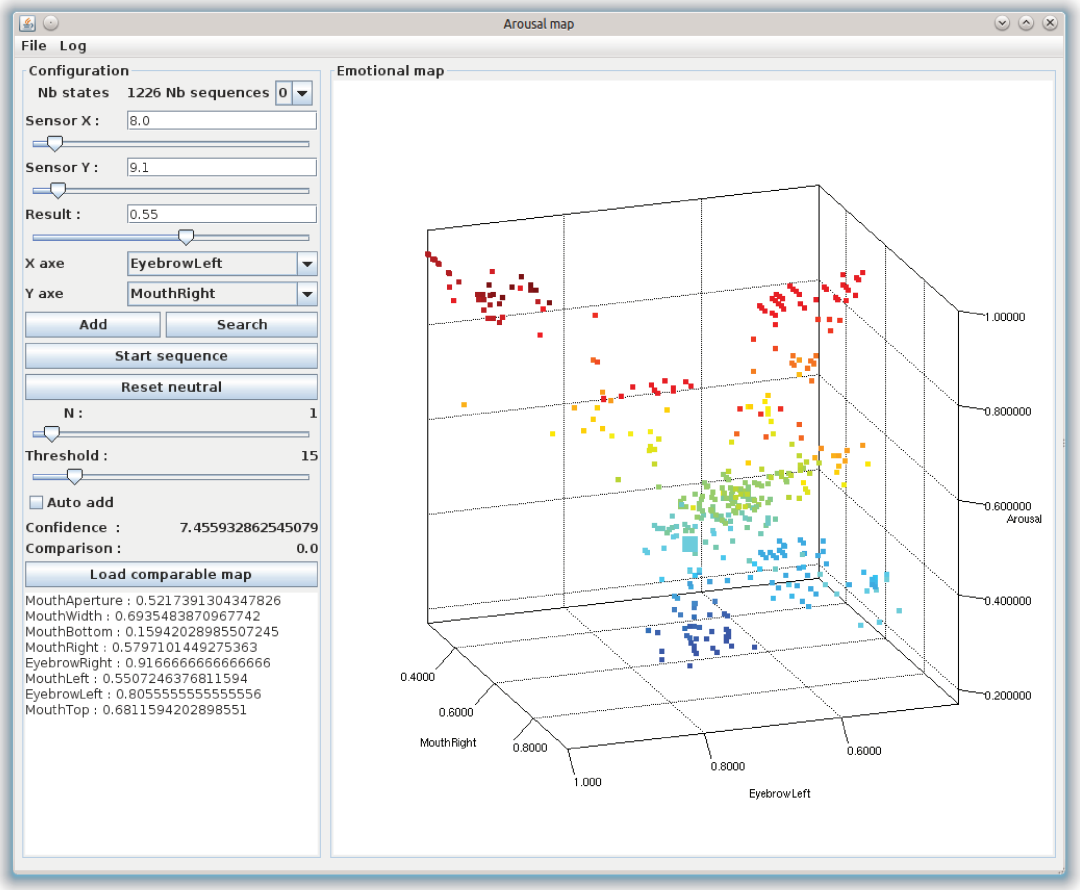


FIGURE 2.9 – Interface graphique du module de génération et d'exploitation des CEPs. Ici, nous pouvons observer une projection sur la commisure droite de la bouche et le sourcil gauche de la carte d'éveil. Le dégradé de couleur représente l'intensité de l'éveil avec en bleu un état calme et en rouge un état éveillé.

Sa première utilisation consiste à créer une CEP. Nous chargerons donc un module distinct par CEP utilisée dans le système global. Comme nous le détaillerons dans les chapitres suivants, la création des cartes consiste en l'annotation manuelle des données captées lors d'une expérience émotionnelle. Afin de regrouper les annotations par type d'expressions, nous définissons une notion de séquence contenant des annotations émotionnelles pour une expression particulière. Une carte pouvant contenir plusieurs séquences, nous pouvons via cette interface en créer une nouvelle et sélectionner parmi celles existantes la séquence à laquelle nous souhaitons affecter notre nouvelle annotation.

L'utilisation de ces cartes est réalisée automatiquement lors de la modification des données captées. Dès qu'une nouvelle information est reçue, l'algorithme de classification est exécuté et affiche au sein même de notre visualisation un point distinctif. Cela permet lors de l'analyse en temps réel, d'observer le déplacement de notre état émotionnel sur les cartes. En cas d'absence d'un ou plusieurs capteurs, il est possible d'effectuer la détection sur ces cartes en les projetant sur les capteurs utilisés. Ceci permet de rendre robuste notre système aux défaillances matériels. L'impact d'un tel incident sur la détection est variable suivant les capteurs utilisés, mais tendra à réduire la fiabilité du système. Après remplacement du capteur défectueux, le système peut alors exploiter à nouveau les informations de ce dernier et ainsi améliorer la détection. Cependant, si on souhaite ajouter un nouveau capteur il est nécessaire de refaire la phase de création des cartes, afin de lier le comportement de ce nouveau capteur au jeu préalablement utilisé. Nous pouvons également faire varier des paramètres d'optimisation de notre algorithme de détection afin d'observer l'impact de ces derniers sur les performances de détection du système. L'analyse des émotions sous l'angle du changement donne une place prépondérante à l'état initial de l'utilisateur. C'est pourquoi nous pouvons réinitialiser notre système à l'état neutre via cette interface.

Les deux premiers cas d'utilisation que nous avons cités sont les points clés de ce module, néanmoins pour une phase d'expérimentation, nous avons porté un soin important à la visualisation de nos cartes. N'ayant pas défini de restriction quant au nombre de capteurs utilisables par nos cartes, il devient difficile de représenter celles-ci lorsque le nombre de capteurs est important. C'est pourquoi, nous avons opté pour une projection des cartes sur deux capteurs évitant ainsi la surcharge d'informations lors de la visualisation. Nous avons également laissé la possibilité de modifier les capteurs utilisés pour la projection, afin d'observer au mieux le comportement de notre système.

Cependant, l'utilisation de ces cartes est lié à la qualité de la phase d'apprentissage. L'ajout d'expressions au sein nos CEPs peut être long et fastidieux. C'est pourquoi, nous avons élaboré une fonctionnalité permettant de réduire cette dernière.

Nous proposons donc, à partir d'une carte minimale, de l'améliorer dynamiquement pendant la phase de détection. L'utilisation de l'information de confiance nous permet de définir si une nouvelle expression est suffisamment proche de celles contenues, afin de pouvoir l'ajouter. Nous utilisons pour cela un seuil d'acceptabilité basé sur la confiance afin de prendre cette décision. L'utilisation de cet apprentissage permet d'accélérer la création des cartes, mais la définition du seuil est capital pour la qualité des cartes résultantes. En effet, un seuil trop faible ajoutera un grand nombre d'expressions et tendra à faire dégénérer nos CEPs, alors qu'un seuil trop important ne permettra pas l'ajout d'expressions rendant ainsi inutile cette fonctionnalité.

Parmi les problématiques initiales, la question de la présence d'hysteresis en est une directement liée à la classification. Nous avons donc rajouté la possibilité d'enregistrer les annotations effectuées par le système afin d'observer cette discontinuité pendant cette phase.

Une autre interrogation était de pouvoir comparer des cartes entre elles, ceci afin d'observer des points communs ou divergeants entre les utilisateurs du système. Nous pouvons donc comparer les cartes visuellement en superposant deux CEPs et en calculant la distance entre l'ensemble des points des deux cartes portant sur la même composante émotionnelle.

2.3 Module de catégorisation et de contextualisation

La phase de catégorisation et de contextualisation est déportée dans un module distinct. Les données utilisées proviennent de la couche inférieure de classification des données captées. A partir de l'annotation brute exprimée dans un modèle dimensionnel des émotions, nous utilisons le contexte de l'expérience émotionnelle par l'intermédiaire des traits de caractères (phobies et philies) de l'individu, afin de pondérer la catégorisation, qui permet d'associer un terme du vocabulaire émotionnel à l'annotation initiale.

L'interface de ce module [Fig. 2.10] est relativement simple. Nous permettons uniquement la sélection des informations permettant la contextualisation ainsi que la description de l'état émotionnel dans le modèle ciblé. Afin de présenter les capacités de contextualisation de l'ontologie sous-jacentes, nous pouvons définir l'utilisateur face au système ainsi que le stimulus auquel il est exposé. Ces deux informations permettent alors d'identifier les données contenues dans notre base de connaissances sur l'individu et d'inférer sur les liens sémantiques entre ses traits de caractères et le stimulus présenté.

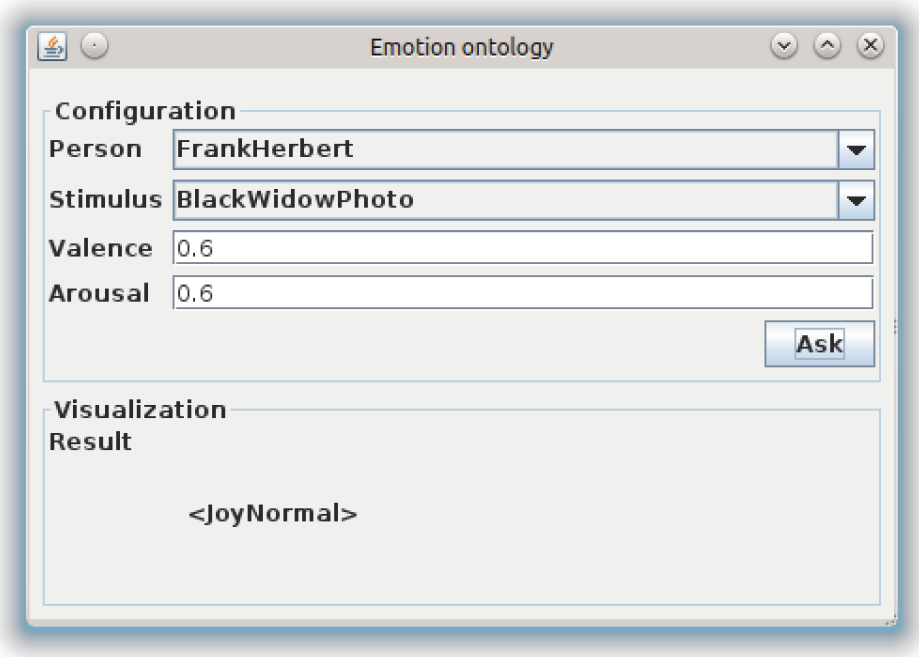


FIGURE 2.10 – Interface graphique du module de catégorisation et de contextualisation des émotions

Sur l'exemple précédent, nous avons fourni une annotation brute représentant la peur, cependant l'émotion en sortie du système indique de la surprise. Cela est obtenu en définissant le trait de caractère "arachnophile" pour l'utilisateur observé. La présentation d'un stimulus représentant une araignée sera pondérée par l'augmentation de la valence de l'émotion. Et c'est la première fois à notre connaissance que le contexte est utilisé afin de pondérer la détection des émotions au travers d'une ontologie.

2.4 Module de synthétisation des expressions faciales

Le dernier module que nous présentons est un module annexe exploitant certaines caractéristiques de nos cartes permettant d'inverser son utilisation. Dans notre exemple, l'utilisation des expressions faciales pour la détection des émotions permet d'annoter un nombre important d'expressions faciales. Nous nous sommes alors posés la question de comment exploiter ces informations afin non pas de détecter les émotions, ce que nous faisons déjà, mais de synthétiser les expressions à partir de nos cartes. Nous avons donc pu mettre en place un algorithme d'exploitation inverse de nos CEPs. Ainsi, nous recherchons l'ensemble des états contenus dans

nos cartes, compatibles avec une annotation émotionnelle donnée. Une fois cet ou ces états récupérés, nous obtenons la valeur des capteurs permettant de reproduire l'expression faciale sur un avatar virtuel.

Sur la figure 2.11, nous avons construit un avatar représentant un "smiley" en courbe de bézier permettant ainsi de configurer son expression via huit points caractéristiques. Les points utilisés ici correspondent aux capteurs exploités lors de l'apprentissage émotionnel.

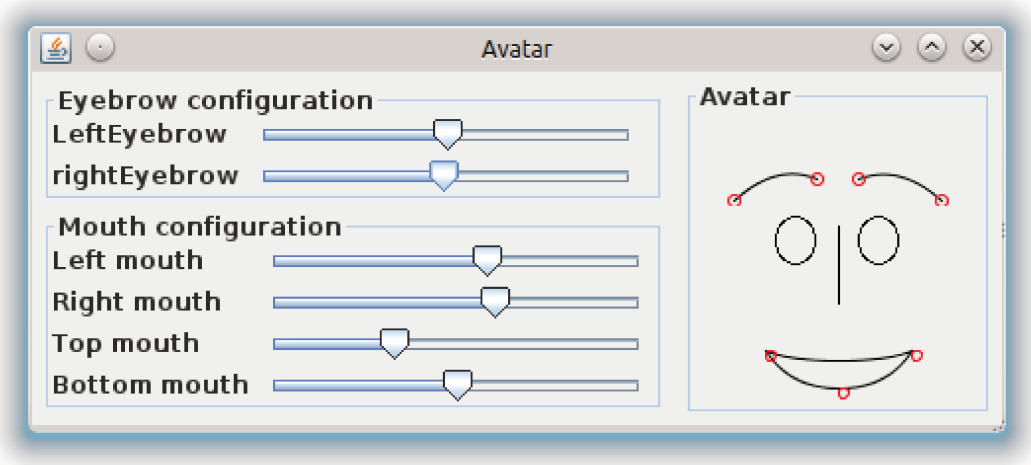


FIGURE 2.11 – Interface graphique du module de synthétisation des expressions faciales

Nous avons donc pu mettre en place dans notre processus un retour virtuel de l'expression émotionnelle détectée. L'annotation fournie en sortie de nos cartes est alors réinjectée dans notre avatar, lui permettant ainsi de rechercher l'expression faciale la plus ressemblante à celle exprimée par l'utilisateur. Il est également possible d'utiliser une annotation catégorielle afin de retrouver l'ensemble des expressions d'une émotion, mais cela rend le choix de l'expression plus difficile. Le fonctionnement de ce module et plus particulièrement de l'algorithme utilisé, est détaillé dans la section 4 du chapitre 3. Le fonctionnement optimal de ce module nécessite l'utilisation des expressions faciales, afin de générer l'avatar. Il est également possible d'imaginer l'utilisation des postures pour une synthétisation plus complète d'avatar. Ce mode de fonctionnement permet de personnaliser les expressions émotionnelles des avatars avec les propres expressions de l'utilisateur.

3 Conclusion

Dans ce chapitre, nous avons présenté l'architecture générale de notre système ainsi qu'une description des modules que nous avons implémenté dans notre prototype. Le détecteur se décompose en 3 modules distincts permettant à chacun de traiter une partie du problème de la détection des émotions. Nous ciblons la tâche du plus bas niveau au plus haut, de traitement du signal de chaque capteur, la génération des CEPs avec l'algorithme de classification et la contextualisation, couplée à la catégorisation des annotations. Les modules sont interconnectés permettant d'enrichir l'annotation dans chacune des étapes de la détection. Pour cela, le module de plus bas niveau permet l'extraction d'informations caractéristiques des émotions. Ces informations sont alors transmises au module de plus haut niveau, visant à classifier ces données grâce aux CEPs. Ce dernier module fournit une annotation brute qui sera alors contextualisée et catégorisée par le module de plus haut niveau. Cette dernière étape permet de prendre en compte des informations durables sur le comportement de l'utilisateur et sur le stimulus déclencheur de l'émotion.

La seconde partie de ce chapitre est dédiée à la présentation de notre prototype. Nous avons sélectionné la caméra comme premier capteur afin de développer un détecteur basé sur les expressions faciales. Ces données sont transmises au module CEP, qui est une implémentation d'un modèle psychologique, basé sur les systèmes complexes. Une variante de l'algorithme de classification KNN, modifié pour effectuer de la régression, est utilisée pour exploiter ces cartes, afin d'identifier l'émotion ressentie. Enfin, une ontologie permet la classification et la contextualisation des émotions, c'est-à-dire une utilisation du contexte émotionnel pour interpréter les expressions ainsi qu'une traduction des annotations dimensionnelles en annotations catégorielles. Un autre module est décrit dans ce chapitre montrant la génération des expressions faciales à partir d'une annotation brute.

Nous avons donc mis en avant dans ce chapitre la structuration en module de notre système afin de répondre aux problèmatiques de flexibilité, induites par une grande variété de contextes applicatifs. En effet, l'*informatique affective* tend à utiliser une grande variété de capteurs, mais les outils disponibles sont souvent spécialisés dans l'utilisation d'un ou plusieurs capteurs définis au préalable. Nous visons ici la possibilité de développer des modules indépendants pour ajouter de nouveaux capteurs en fonction de l'application réalisée. De plus, la structure de nos modules permet d'obtenir des informations à des niveaux d'interprétation variés. Une première interprétation permet d'obtenir une annotation dimensionnelle basée sur les expressions émotionnelles uniquement. Une seconde annotation est disponible après pondération des annotations dimensionnelles en prenant en compte le contexte. Cette dernière annotation est disponible soit dans un format dimensionnel ou catégoriel

après traduction. Nous décrivons ici, notre vision de la structuration d'un détecteur émotionnel capable de répondre à la plus grande variété de contextes applicatifs, en identifiant certaines contraites, comme l'utilisation de capteurs fournissant une donnée numérique continue, ou l'utilisation des modèles émotionnels dimensionnels pour la classification des émotions.

Chapitre 3

Les variations émotionnelles sous l'angle des sytèmes complexes

Dans le chapitre précédent, nous avons présenté l'architecture globale de notre système. Ici, nous allons détailler le fonctionnement de la couche centrale de notre système à savoir le module d'interprétation des expressions émotionnelles.

En nous basant sur les théories du processus émotionnel, nous pouvons dire que l'environnement (mais également les stimulus internes comme les souvenirs) est le facteur de changement émotionnel. Un objet de notre environnement peut devenir le stimulus déclencheur de l'émotion que nous allons ressentir. Cependant, il est difficile de déduire d'un stimulus, l'émotion ressentie. La solution couramment utilisée est d'observer une partie de l'expression émotionnelle par l'intermédiaire de capteurs. En d'autres termes, l'environnement déclenche un changement émotionnel par l'intermédiaire d'un stimulus, ce qui provoque une expression observable chez l'individu. Cette indirection nous permet d'obtenir des données tangibles pour la détection émotionnelle.

Á partir de cette source d'informations, deux voies principales d'interprétation existent : la recherche de patrons ou la classification. Dans le cadre de ce document, nous avons opté pour une approche régressive afin de nous conformer à la vision continue des émotions. De plus, le phénomène que nous souhaitons observer possède certaines caractéristiques difficilement modélisables. Nous basons notre réflexion à partir du modèle de Scherer, lui même utilisant les travaux de Thom sur les systèmes complexes pour la modélisation des émotions. Le modèle proposé n'ayant, à notre connaissance, pas d'implémentation connue, nous souhaitons par l'intermédiaire de nos travaux apporter une étude expérimentale de ce dernier. En effet, l'expérience émotionnelle est stable, c'est-à-dire, que le changement d'émotion n'est pas proportionnel à l'intensité du stimulus. Par exemple, l'augmentation progressive de la difficulté d'un jeu va entrainer une réaction d'énervement lorsque le joueur ne pourra

plus avancer dans la tâche qui lui est confiée. Ce moment correspond à une rupture qui délimite l'émotion initiale (joie ou émotion neutre), de l'émotion finale (colère). Si l'on considère maintenant la situation inverse, la diminution progressive de la difficulté ne va pas permettre le retour au calme au même niveau que dans la situation précédente, mais à un niveau bien inférieure de difficulté.

Ce phénomène est modélisable mathématiquement par la théorie des catastrophes sous le nom d'hysteresis. Une modélisation prenant en compte ce comportement et s'appuyant sur cette théorie est proposée par Klaus R. Scherer [Scherer 00, Scherer 09].

Une autre difficulté dans l'exploitation de l'expression émotionnelle est la spécificité de celle-ci. En effet, le lien entre l'expression et l'émotion ressentie est spécifique à chaque individu. C'est pourquoi nous souhaitons proposer un modèle personnalisable des émotions.

C'est à partir de ces différentes observations, que nous proposons un modèle d'interprétation des expressions émotionnelles que nous appelons CEP. Cette modélisation repose à la fois sur les travaux de Klaus R. Scherer et son modèle dynamique, mais aussi sur une modification de l'algorithme de classification KNN adapté à la régression. La définition de ce modèle se fera indépendamment des capteurs et des modèles émotionnels utilisés pour la détection des émotions, afin de rendre notre outil adaptable au maximum de contextes applicatifs.

Dans ce chapitre, nous nous concentrerons sur la présentation des CEPs. Dans un premier temps, une définition permettant de structurer le concept sera proposée. Ensuite, nous décrirons les différentes fonctions applicables sur nos cartes, en particulier, les méthodes d'apprentissage et de détection des émotions, mais également la notion de confiance que nous pouvons en extraire. Cette dernière permettra d'aborder la problématique de l'apprentissage automatique et les questions d'enrichissement des cartes. Enfin, nous présenterons les capacités des CEPs pour la synthétisation des émotions par des avatars virtuels.

1 Les Cartes Émotionnelles Personnalisées

1.1 Utilisation et objectifs

Ce que nous appelons CEP est utilisée comme lien entre les données d'expressions captées et l'annotation émotionnelle dimensionnelle. Ces cartes tiendront donc lieu de base d'apprentissage pour l'exécution de notre algorithme de régression. L'objectif principal est de pouvoir annoter une expression émotionnelle à partir des données d'apprentissage. Cependant, nous visons d'autres objectifs comme l'adaptabilité et

une détection en temps réel. Ce que nous définissons comme l'adaptabilité est le fait que notre modélisation puisse d'une part être utilisée avec une grande variété de capteurs, mais aussi de décrire l'émotion dans la plus part des modèles émotionnels.

Ce que nous proposons pour cette modélisation est de nous appuyer sur des bases théoriques solides afin d'observer et de comparer le principe d'émotion. La nature même de l'émotion pose problème de part sa complexité, son aspect éphémère et subjectif. Nous souhaitons pour cela une base théorique nous offrant des outils d'analyses afin de poursuivre l'exploitation de ces cartes en dehors du cadre de ce document.

Nous avons opté pour la modélisation complexe de Scherer. Cette proposition a l'avantage de fournir une vision continue des émotions en se basant sur les systèmes complexes et plus particulièrement sur la théorie des catastrophes. Une telle approche permet de modéliser les changements émotionnels par des discontinuités proposées dans la théorie des catastrophes. Dans ses articles, Scherer prend l'exemple de l'hystérésis pour décrire le besoin de discontinuité concernant la modélisation du changement émotionnel. Dans la section résultat, nous essayerons d'observer ce phénomène lors d'une détection, afin de vérifier par l'expérimentation cet exemple.

1.2 Modèle dynamique

Le modèle dynamique que nous souhaitons utiliser n'a, à notre connaissance, jamais été implémenté. Nous souhaitons à travers les CEPs proposer une implémentation informatique de ce modèle afin d'évaluer ses performances dans un contexte expérimental. L'utilisation de la théorie des catastrophes permet d'étudier et de modéliser certains aspects du comportement d'un sytème complexe. Une définition d'un système complexe peut être : un grand nombre d'informations en interconnexion facilement compréhensibles individuellement mais dont le comportement global ne l'est pas. Dans le contexte de l'émotion, le système complexe correspond à l'expérience émotionnelle et les informations en interaction à l'environnement ayant déclenché cette dernière. Comme nous l'expliquions en introduction, l'observation de l'environnement est une tâche très complexe de par sa variabilité. Nous privilégierons dans nos travaux les expressions émotionnelles. Nous pouvons alors modifier notre enoncé précédent en indiquant que nos informations en interaction sont les expressions émotionnelles.

En prenant un exemple de capteur physiologique comme le rythme cardiaque, il est facile de comprendre qu'une augmentation de ce dernier induit une augmentation de l'éveil et inversement. Cependant, l'expression émotionnelle est une notion complexe activant un certain nombre de réponses corporelles. Selon les théories

émotionnelles, les réponses émotionnelles sont non-spécifiques. De même, certaines réponses physiologiques peuvent être produites par d'autres facteurs que l'émotion. Par exemple, lors d'une activité physique, l'augmentation du rythme cardiaque n'est pas un signe d'éveil ou encore la dilation des pupilles peut être liée à une diminution de l'intensité lumineuse environnante. Il est alors nécessaire d'en observer une grande quantité afin d'identifier l'émotion ressentie. C'est pour cela que nous proposons d'intégrer nos travaux dans la vision des émotions par les systèmes complexes.

Néanmoins, dans le cadre d'une détection en temps réel, basée sur des capteurs d'expressions émotionnelles, il est difficile d'utiliser directement ce modèle. C'est pourquoi nous proposons de modifier le modèle de Scherer en fonction de nos contraintes applicatives. Les modifications que nous allons proposer dans la suite de ce document respectent le formalisme défini par la théorie des catastrophes, en particulier concernant la définition de notre espace multi-dimensionnel. La définition de cet espace repose sur deux éléments, l'espace de contrôle et la dimension représentant le comportement de notre système. Ce premier élément caractérise l'ensemble des variables impactant le comportement de notre système. Ces variables définissent un espace multi-dimensionnel. Dans l'exemple proposé par Klaus R. Scherer, l'espace de contrôle correspond à la puissance et l'obstruction de celle-ci. Dans le contexte de la détection, nous ne pouvons pas observer ces deux facteurs. Nous proposons alors de remplacer les deux facteurs par les expressions émotionnelles. Le terme même d'espace de contrôle tend à définir qu'il est possible de modifier les variables du système de manière contrôlée. Dans notre situation, les expressions ne sont pas contrôlables de part leur nature réflexe. Cependant, le contrôle de l'environnement de l'utilisateur est possible en particulier dans un environnement virtuel. La possibilité de modifier le stimulus initial impacte directement l'expression émotionnelle. Nous utiliserons donc cette indirection pour justifier le choix de notre espace de contrôle. Même s'il est plus aisé d'observer des réponses émotionnelles, celles-ci sont nombreuses et ne peuvent être capturées par un unique capteur. Nous proposons donc une définition de notre espace de contrôle indépendamment des capteurs utilisés lors de l'implémentation. Cependant, nous fixons certaines contraintes sur les capteurs compatibles avec notre modélisation. Ceux-ci doivent fournir une information numérique continue. Par exemple, un ECG est compatible avec notre définition alors qu'un texte ne l'est pas. De même certains capteurs peuvent sembler incompatibles comme la caméra, mais en considérant le déplacement de points caractéristiques sur les différentes images, il est possible d'utiliser ce capteur. Dans la suite du document, quand nous parlerons de capteurs, nous ferons référence à l'ensemble des capteurs répondant aux contraintes fixées. D'une manière concrète, chaque dimension de notre espace de contrôle sera définie par un des capteurs utilisés dans l'implémentation de notre modèle. Par exemple, une implémentation utilisant un ECG et un Galvanic

Skin Response (GSR) définira un espace de contrôle à deux dimensions, chacune représentant un capteur différent.

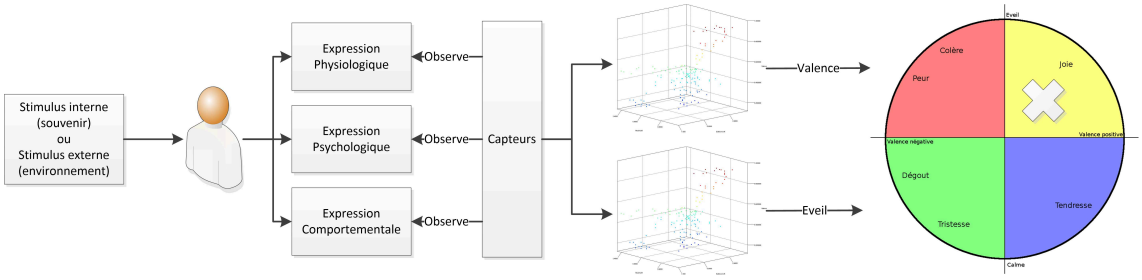


FIGURE 3.1 – Fonctionnement global des CEPs

Le second élément que nous devons décrire est la dimension résultat ou comportement de notre système. C'est l'évolution de cette dimension que nous devons observer pour identifier le comportement de notre sytème après modification de nos variables de contrôle. Cette dernière doit être unique afin d'isoler l'impact d'une modification de l'espace de contrôle sur le comportement du système. Dans l'exemple proposé par Klaus R. Scherer, la dimension "comportement" correspond aux émotions. Compte tenu de notre modélisation continue, nous devons décrire les émotions par l'intermédiaire de modèles émotionnels dimensionnels. Ces derniers étant les seuls capables de décrire l'ensemble des émotions et leurs intensités dans un espace continu, la contrainte précédente imposée par la théorie des catastrophes ne permet pas d'utiliser les modèles comportant plus d'une dimension comme dimension "résultat". Nous proposons de résoudre cette problématique en créant une CEP par dimension du modèle émotionnel sélectionné. En reprenant l'espace de contrôle de l'exemple précédent, et sachant que l'on souhaite exprimer les émotions dans le modèle eveil/valence, nous devons créer une carte pour la dimension de valence et une seconde pour la dimension d'éveil, chacune partageant le même espace de contrôle. Cette décomposition de nos espaces complexes en fonction des dimensions du modèle cible, nous permet de rendre compatible notre système avec l'ensemble des modèles dimensionnels.

Le prototype que nous avons proposé dans le chapitre 2, présente une implémentation de nos CEPs à partir d'une caméra et du modèle de Russell. Cette implémentation permet de construire des cartes à partir de données fournies par la caméra. Comme expliqué précédemment, nous utilisons ce capteur pour récupérer le déplacement de huits points caractéristiques de l'expression faciale. Ces huits éléments représentent notre espace de contrôle pour nos CEPs. Ce dernier est donc de dimension huit. Pour fournir une annotation dans notre modèle cible, nous déclarons les dimensions d'éveil et de valence afin que le système génère deux cartes, une

pour la valence et une pour l'éveil. Nous utilisons ces cartes de dimension neuf afin de stocker notre base d'apprentissage, représentée par des expressions corporelles, préalablement annotées émotionnellement, dans le but d'appliquer notre algorithme de régression.

1.3 Apprentissage et annotation des états émotionnels

Comme nous l'évoquions précédemment, les CEPs sont utilisées comme base d'apprentissage pour notre algorithme de régression. Ceci, implique une phase de population de nos CEPs après les avoir définies. Cette étape d'apprentissage va permettre d'intégrer des expressions corporelles annotées émotionnellement au sein de notre espace complexe. Ces états sont définis par l'ensemble des valeurs captées à un instant donné et associées à une annotation émotionnelle déterminée par le modèle utilisé pour la construction des CEPs. De manière générale, la construction des CEPs pendant la phase d'apprentissage se déroule de la manière suivante :

1. on présente au sujet un ensemble de stimulus
2. ses réactions sont indéxées par $I_a \in [1, 2, 3...]$ correspondant à une séquence regroupant les mesures des capteurs
3. on associe les annotations émotionnelles aux mesures des capteurs

De manière plus formelle,

$$I_a \mapsto \mathbb{R}^M \text{ pour } M \text{ capteurs} \tag{ii}$$

Cependant, nous recherchons l'annotation émotionnelle à partir des données captées

$$I_a \mapsto \mathbb{R}^N \text{ pour } N \text{ dimensions émotionnelles} \tag{iii}$$

Nous définissons alors les CEPs comme une application composée :

$$CEP : I_a \to \underbrace{\mathbb{R}^M \to \mathbb{R}^N}$$

Une fois cet apprentissage effectué nous considérons une CEP comme un ensemble de points CEP $\subset \mathbb{R}^{M+N}$.

Pour notre approche concernant la détection émotionnelle, cette phase d'apprentissage doit permettre de capturer l'apparition de l'émotion jusqu'à sa disparition, ceci afin d'observer le changement émotionnel provoqué par le stimulus. Nous proposons de regrouper au sein de nos cartes la succession d'expressions proposées

par un utilisateur lors du ressenti d'une émotion. Nous définissons alors la notion de séquence comme l'ensemble des états successifs ayant permis l'expression d'une émotion. Par exemple, pour apprendre au système à détecter la joie, nous créons une nouvelle séquence émotionnelle dans nos CEPs contenant les expressions de l'utilisateur lors du ressenti de la joie. Les cartes peuvent contenir un grand nombre de séquences et potentiellement plusieurs pour une même émotion. La figure 3.2 montre quelques-unes des séquences contenues dans une carte expérimentale.

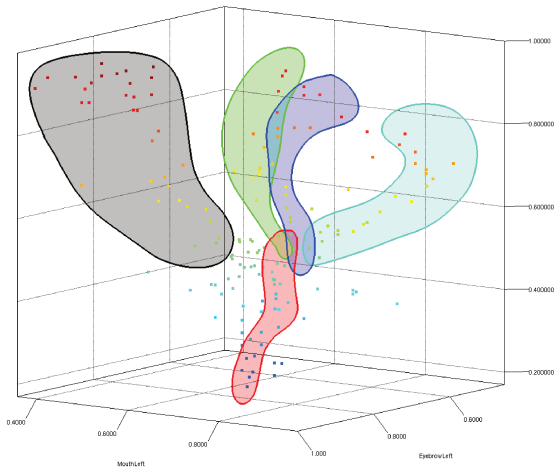


FIGURE 3.2 – Représentation des séquences contenues dans une CEP expérimentale

Cette méthode de construction permet de conserver le lien temporel lors du ressenti émotionnel. Cette démarche se justifie dans le cadre d'une détection continue où nous pouvons observer le déclenchement d'une expression émotionnelle. Par exemple, l'expression faciale de la joie peut être l'apparition d'un sourire. Cependant, sur un laps de temps suffisamment court, il est possible de voir apparaitre le déplacement des commissures de la bouche vers le haut. Cette succession d'expressions va permettre à notre système de sélectionner la séquence d'expressions la plus proche de celles détectées.

Dans le cadre de notre prototype basé sur les expressions faciales, nous avons mimé un ensemble d'émotions en partant de l'émotion neutre pour atteindre le ressenti le plus intense, puis revenir au calme. Nous avons alors utilisé l'expression neutre comme délimiteur de nos séquences.

Il est possible grâce à ce découpage, de construire une séquence à partir d'une expression inconnue par le système et d'utiliser les annotations proposées par celui-ci lors de la détection, comme nouveaux états. Cette démarche a pour objectif l'observation de l'émergeance de nouvelles émotions dans le cadre de notre modèle dynamique. Ce type d'observation sera effectuée dans le cadre de la poursuite de nos travaux afin de monter une expérimentation validant ou invalidant ce genre d'utili-

sation par notre système. Nous tenons à préciser que ce type d'utilisation a motivé notre choix de conception.

2 Reconnaissance des émotions

La recherche de l'annotation émotionnelle à partir des CEPs se base sur l'hypothèse de Scherer [Scherer 09] selon laquelle un individu ressentant une certaine émotion tend à conserver son état le plus longtemps possible. L'évolution continue d'un stimulus va permettre le changement émotionnel lorsque le sujet ne peut plus accepter l'intensité de la stimulation et opte donc pour une nouvelle réaction émotionnelle. De plus, nous visons une approche de la détection sous l'angle du changement, ceci ayant pour intérêt d'observer le déclenchement de l'émotion et donc de stabiliser la détection dans une direction ayant été initiée par les changements précédents.

Pour cela, nous utilisons comme base l'algorithme de classification K-nearest neighbors (KNN) [Fig. 3.3]. Afin de l'employer comme outil de recherche de l'état émotionnel, nous en proposons une version modifiée répondant aux critères définis ci-dessus. Premièrement, le choix de cet algorithme est motivé par cette approche de l'émotion par le changement, ainsi que par nos objectifs de temps réel. Cet algorithme, bien que considéré comme peu performant dans le domaine de la fouille de données, possède une simplicité nous permettant d'atteindre une vitesse d'exécution de l'ordre du temps réel. De plus, la disposition des états émotionnels de nos cartes forme des chemins facilement exploitables par cet algorithme.

Cependant, une classification directe des données captées est soumise à de grandes perturbations de part la non spécificité des réponses émotionnelles. Par exemple, une classification basée sur le rythme cardiaque ne peut pas fournir une information stable du fait d'une augmentation similaire dans le cas de la colère, peur ou joie. Les modifications physiologiques peuvent être causées par un évènement extérieur à l'émotion, en particulier les activités physiques pour le rythme cardiaque ou une modification d'intensité lumineuse pour la dilation pupillaire. C'est pourquoi nous intégrons dans le vecteur à classifier une nouvelle dimension correspondant à l'annotation émotionnelle précédente. Cette information tend à stabiliser la classification en augmentant l'intérêt d'une annotation proche de la précédente. L'initialisation est effectuée avec une annotation neutre, mais peut être modifiée par l'utilisateur en fonction de l'émotion ressentie lors du démarrage du système.

La deuxième modification réside dans la stratégie de classification utilisée. Une utilisation classique du KNN détermine la classe de l'objet à classifier, comme la classe majoritaire des k voisins les plus proches. Dans notre cas, nous n'avons pas à proprement parlé de classe, mais d'une annotation continue représentant l'émotion.

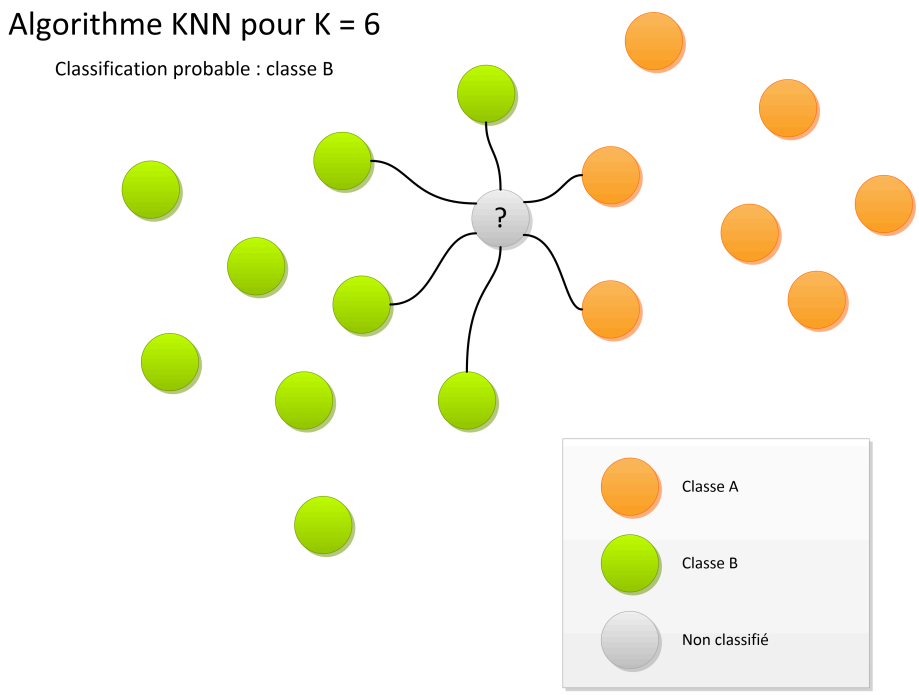


FIGURE 3.3 – Schématisation de l'algorithme KNN. Les classes A et B pouvant représenter deux émotions distinctes.

C'est pourquoi, nous utilisons le barycentre pondéré par l'inverse de la distance entre le point recherché et ses voisins pour définir l'annotation de notre nouvelle émotion. Cela permet de réduire l'impact de points trop éloignés de l'annotation recherchée, surtout dans le cas d'utilisation d'un nombre de voisins importants.

La figure 3.4 présente le fonctionnement de notre algorithme. À partir d'un nouvel ensemble de données captées, nous devons construire le nouvel état $D_t \in \mathbb{R}^{M+N}$ où t correspond à un indice temporel, nous cherchons à retrouver l'émotion correspondant à cette expression à partir de nos CEPs. Conceptuellement,

1. D_t correspond alors à la combinaison entre les N dimensions émotionnelles préalablement calculées pour D_{t-1} associées aux M nouvelles données captées, nous permettant de construire un état temporaire où $D_t \in \mathbb{R}^{M+N}$;

2. notre algorithme détermine la nouvelle émotion pour D_t comme le barycentre de l'ensemble des K plus proches voisins de nos CEPs.

La valeur du parmètre K, le nombre des plus proches voisins à considérer pour le calcul de la nouvelle émotion, est un paramètre de notre système qui reste à définir. Nous proposons une détermination de celui-ci dans le chapitre 5 section 1.2

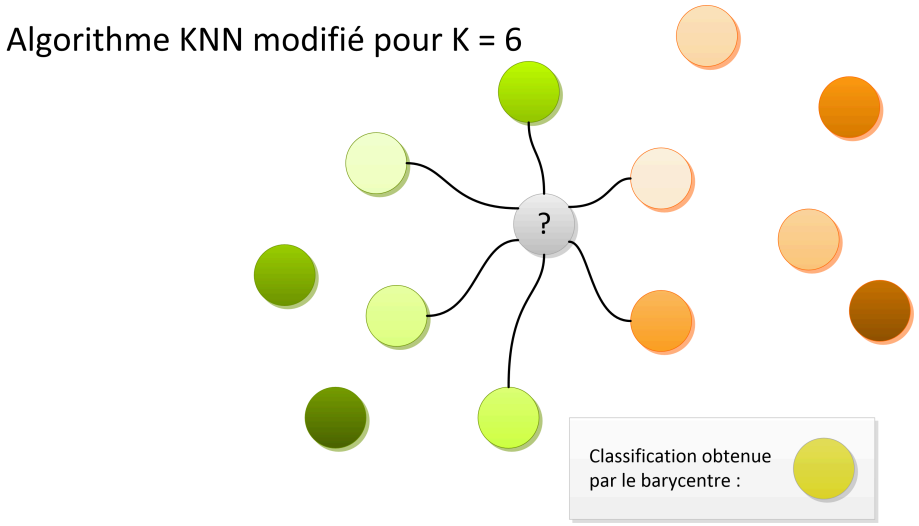


FIGURE 3.4 – Schématisation de l'algorithme KNN modifié

Comme nous l'indiquions dans les sections précédentes, nos annotations sont divisées au sein de nos cartes en fonction de la séquence utilisée pour leur génération. L'application de cet algorithme est reproduite sur chacune des séquences afin d'obtenir un ensemble d'annotations potentielles. Le choix de l'annotation s'effectue en fonction de la confiance attribué par notre algorithme à chaque exécution, afin de déterminer laquelle est la plus fiable. Nous utilisons l'inverse de la distance moyenne afin de définir la confiance de la classification. La confiance est décrite par la formule suivante :

Soit K, le nombre de plus proches voisins, $\mathcal{E}$, un état parmi les plus proches voisins, $\mathcal{S}$, l'état recherché et M, le nombre de capteurs, alors :

$$confiance = \frac{1}{\sum_{i=1}^{K}\left(\sqrt{\sum_{j=1}^{M+1}\left(\mathcal{E}_{i_j} - \mathcal{S}_j\right)^2}\right)/K} \tag{3.1}$$

Cette information permet d'identifier si la classification proposée a été déterminée grâce à un grand nombre de points similaires et donc si elle est potentiellement fiable. La confiance est utilisée par la suite pour automatiser l'apprentissage, afin d'affiner les cartes préalablement annotées.

Le pseudo-code fournit en annexe A ainsi que le schéma suivant en [Fig. 3.5] montrent le fonctionnement global de notre algorithme de recherche. Il se décompose en trois grandes étapes, une première permettant de parcourir les séquences de notre CEP et d'appliquer le traitement sur chacune d'elles. La deuxième étape correspond au fonctionnement du KNN, en cherchant les K plus proches voisins du vecteur comprenant les nouvelles données et l'annotation précédente, à partir du calcul de la distance euclidienne. Enfin, la dernière étape vérifie pour chaque séquence

si la confiance moyenne de l'annotation proposée sur la séquence courante est plus fiable que les précédentes, permettant ainsi de calculer le barycentre représentant l'annotation finale. Cette dernière étape permet donc de faire la sélection du barycentre le plus pertinent pour la détection de l'émotion en fonction de la confiance attribuée par le système.

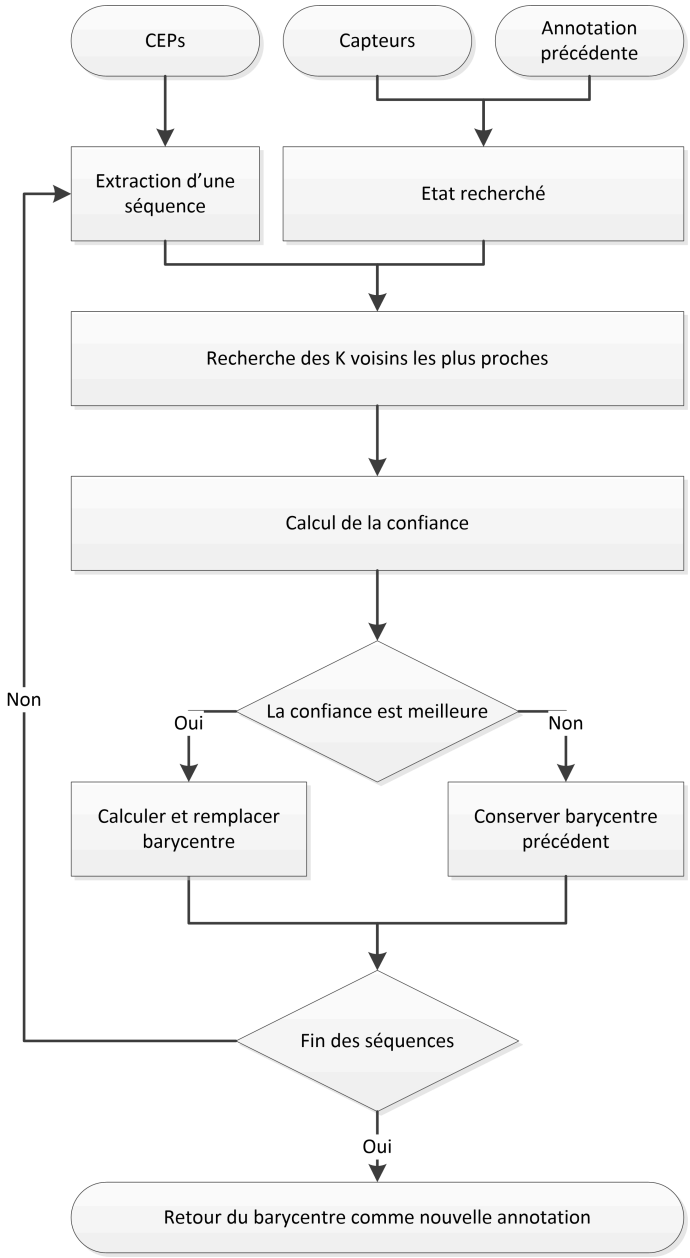


FIGURE 3.5 – Schématisation du processus de détection des émotions dans les CEPs

3 Apprentissage automatique

Comme nous l'indiquions précédemment, la notion de confiance est une information calculable grâce à la structure de nos cartes et à l'algorithme de classification utilisé. Cette donnée fournit une information sur la pertinence de la détection et peut être utilisée pour avertir d'un défaut du système lors de l'annotation. Cette première utilisation est intéressante dans des contextes ayant besoin d'un haut niveau de fiabilité. Cependant, une autre utilisation de la confiance peut être effectuée afin d'obtenir une amélioration du système. En effet, la phase d'apprentissage est longue et fastidieuse surtout pour des séquences vidéos contenant de nombreuses images. C'est pourquoi, nous utilisons la confiance pour affiner nos cartes et ajouter de nouveaux états au sein des CEPs.

Lors de la détection d'une émotion, le système produit deux informations, l'annotation et la confiance sur chacune des cartes. Cette annotation couplée aux données captées représente un vecteur émotionnel comparable à ceux existant au sein de nos cartes. Nous proposons de réinjecter ce vecteur comme un nouvel état émotionnel dans la séquence ayant permis son annotation. La confiance est ici utilisée pour déterminer si ce vecteur est potentiellement viable pour être réinjecté. Pour ce faire, nous définissons un seuil de confiance minimum pour indiquer au système d'ajouter ou non cet état dans les cartes. Un seuil trop faible aura pour conséquence de faire diverger ces dernières et donc d'empécher la détection des émotions. Au contraire, un seuil trop important ne permettra pas l'ajout de nouveaux états. Le choix de ce seuil dépend de plusieurs facteurs, d'une part la structure même de nos cartes mais également du nombre de voisins utilisé par l'algorithme de classification. Par exemple, pour un K égal à 1 une confiance infinie peut être obtenue si la valeur exacte des données captées a déjà été intégrée au système. En revanche, pour des valeurs de K plus importantes, l'algorithme permet de générer de nouvelles annotations correspondant au barycentre des plus proches voisins, il devient donc extrêmement rare d'obtenir un niveau comparable. La détermination de ce seuil doit être effectuée après la définition du paramètre K, afin de le considérer comme une constante dans la formule de la confiance. Il n'est cependant pas possible de déterminer un seuil optimum adaptable aux différentes situations de test. Le choix devra être effectué en fonction des capteurs utilisés, de l'intervalle des données issues des differents capteurs et de la valeur du paramètre K, le tout pondéré par la fiabilité que l'on souhaite obtenir pour nos cartes.

L'utilisation de cette fonctionnalité ne permet pas la génération de cartes automatiquement, il est nécessaire de fournir une carte minimale contenant l'éventail complet des annotations que nous souhaitons obtenir. L'apprentissage automatique se limite donc a l'affinage de nos cartes, afin d'augmenter la probabilité de bonne

classification.

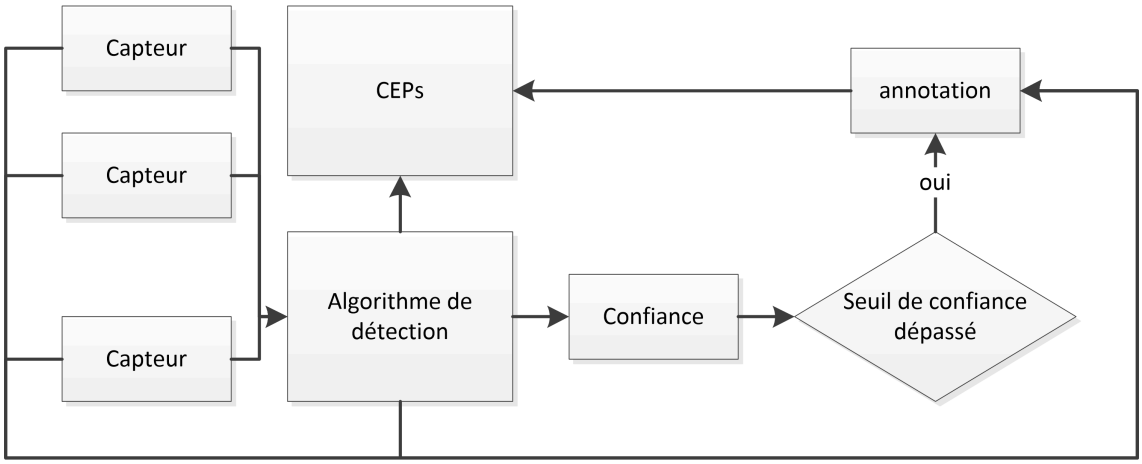


FIGURE 3.6 – Schématisation du processus d'apprentissage automatique

4 Algorithme de génération d'expressions émotionnelles

L'utilisation des émotions peut également être vue comme un outil de communication. En effet, les échanges émotionnels sont utilisés, dans le cadre d'une discussion, pour améliorer la compréhension de certains messages. Pour reproduire ce système de fonctionnement entre un humain et une machine, il est nécessaire dans un premier temps, de reconnaître l'émotion exprimée par l'humain, c'est cette partie que nous visons pour notre réalisation. Dans un second temps, le système informatique doit pouvoir exprimer des émotions. Cette seconde partie n'était pas un objectif au début de nos travaux. Cependant la structure de nos CEPs nous permet d'envisager ce type d'utilisation.

Initialement, nous avons développé ces cartes afin de relier les données captées avec une annotation émotionnelle. Dans le cadre de notre algorithme de détection, nous fournissons les données captées afin de récupérer une annotation émotionnelle. Ici, nous interrogeons nos cartes avec une annotation émotionnelle afin de récupérer les données captées. Le fonctionnement de notre algorithme se base sur l'existence d'états annotés émotionnellement au sein de nos cartes. Nous proposons de récupérer, sur chacune de nos cartes, les états dont l'annotation émotionnelle est la plus proche de celle recherchée. De part la construction de nos cartes, il est possible d'obtenir plus d'un état correspondant aux critères définis. C'est pourquoi, à partir des états candidats, nous définissons une stratégie de sélection de l'expression émotionnelle adaptée. Cette phase est complexe car elle ne possède pas de données tangibles

permettant d'effectuer cette sélection. Dans l'implémentation que nous proposons, nous avons opté pour une stratégie de moyennage des états, capteur par capteur. Nous détaillerons dans la section 4, les performances d'une telle approche.

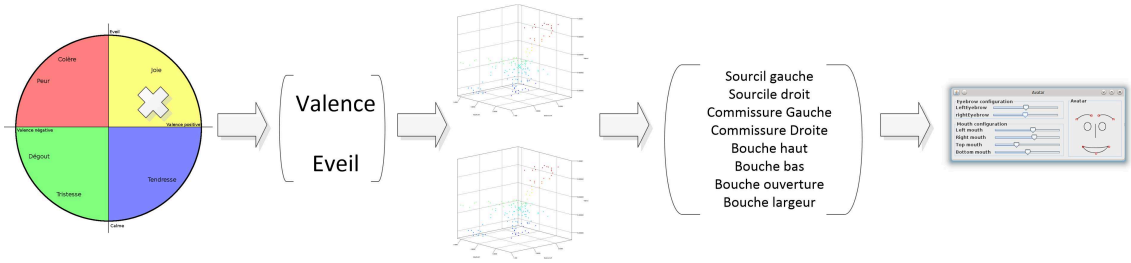


FIGURE 3.7 – Principe général de fonctionnement de la synthétisation des émotions

Une telle approche n'est valable que dans le cas où la détection est basée sur un ensemble de capteurs utilisés pour la synthétisation des émotions. Pour notre prototype, nous utilisons les expressions faciales comme sources de données. Nous pouvons donc appliquer cette fonctionnalité, afin de récupérer les déplacements des points caractéristiques et donc, de reproduire l'expression faciale de l'utilisateur sur notre avatar. Dans le cas d'une détection basée uniquement sur des capteurs physiologiques, il semble plus difficile d'utiliser ces informations pour la génération d'avatar. Cependant, il est possible d'utiliser ce type d'information, afin d'augmenter le réalisme de l'avatar en réutilisant les expressions de l'utilisateur.

5 Conclusion

Dans ce chapitre, nous avons détaillé le fonctionnement de notre module intermédiaire de détection des émotions. L'objectif ici était d'identifier l'émotion ressentie en nous basant uniquement sur des données d'expressions émotionnelles. Comme nous l'évoquions précédemment, le déclencheur de l'émotion provient de l'environnement de l'utilisateur aussi bien externe (situation du monde réel) qu'interne (souvenir, mémoire). La difficulté d'observation de ces informations est contournée en nous penchant sur les expressions émotionnelles. Celles-ci sont le reflet des modifications corporelles induites par le stimulus déclencheur de l'émotion. Nous nous sommes donc servis de cette indirection, afin d'utiliser les expressions comme source de données pour notre détection.

Dans le domaine, de nombreux travaux utilisent à la fois ces informations et des algorithmes de classification pour réaliser ce traitement. Dans notre cas, nous avons souhaité appréhender ce problème sous un angle nouveau. Les travaux de Scherer sur la modélisation des émotions par les systèmes complexes et les théories

de Thom, nous ont servi de base pour l'élaboration de cette partie. L'objectif d'une telle approche est double. D'une part, nous souhaitions décrire les émotions comme un phénomène continu, afin de nous conformer à l'hypothèse que nous pouvons ressentir une infinité d'émotions mais que notre vocabulaire nous limite dans la communication de celles-ci. D'autre part, la majorité des systèmes existant dans le domaine utilise un jeu de capteurs, déterminé à l'avance, et les émotions sont décrites par un modèle spécifique. Ici, nous avons opté pour une modélisation plus fexible permettant aux utilisateurs de notre système de définir à la fois les capteurs à utiliser, mais aussi le modèle de description à employer. Pour cela nous avons fixé deux contraintes, à savoir l'utilisation de capteurs numériques continus et uniquement des modèles dimensionnels des émotions. L'utilisation de ces limitations, couplée à une modélisation basée sur les sytèmes complexes, nous a permis de donner une réponse à cette problématique au travers de nos CEPs.

De plus, à notre connaissance aucune implémentation du modèle de Scherer n'a été proposée. Bien que nos cartes ne soient pas une implémentation directe de ce modèle, nous avons pu proposer une solution exploitant ces hypothèses afin de tester dans des conditions expérimentales, le fonctionnement de cette approche. L'évaluation de notre modèle est disponible dans le chapitre 5.

Au travers de cette partie, nous avons pu montrer les modifications apportées au modèle de Scherer, afin de l'adapter à une détection en temps réel, basée sur des données expressives. Nous avons également présenté les outils permettant de créer ces cartes, en effectuant un apprentissage manuel, complété par une tentative d'apprentissage automatique, permettant d'affiner ces dernières, ainsi que la structuration en séquence des émotions au sein de nos CEPs. Nous avons ensuite présenté notre algorithme de détection à proprement parler. Celui-ci basé sur l'algorithme KNN et modifié pour l'appliquer dans le cadre de la régression, exploite nos cartes, afin de prédire l'annotation émotionnelle correspondant à un jeu de données captées à un instant donné. Outre le cadre de la régression employé pour la détection, cet algorithme utilise l'annotation émotionnelle précédemment détectée, afin de stabiliser la prédiction, en donnant un poids supplémentaire à une détection compatible avec la précédente. Enfin, nous avons proposé une utilisation inversée de nos cartes, afin de synthétiser les expressions faciales sur un avatar, en réutilisant les expressions de l'utilisateur.

Chapitre 4

Catégorisation et contextualisation par une ontologie

Dans le chapitre précédent, nous avons décrit le fonctionnement des CEPs. Les outils d'apprentissage et de détection développés pour ce module permettent d'obtenir une annotation brute, dépendante d'un modèle dimensionnel en sortie des cartes. L'objectif ici, est de fournir un enrichissement et une simplification de cette information grâce à des connaissances sur l'utilisateur stables dans le temps. Les annotations dimensionnelles ont l'avantage de pouvoir décrire les émotions dans un espace continu, permettant ainsi de décrire l'ensemble des nuances de l'émotion. Cependant, la compréhension d'une telle annotation par un individu non expert peut être difficile. De plus, en se référant à la théorie des deux facteurs, une reconnaissance basée sur l'expression des émotions ne permet pas d'appréhender l'émotion dans sa globalité. Cette théorie ainsi que les théories plus modernes proposent de prendre en compte l'expérience du sujet face au stimulus qu'il rencontre. Par exemple, l'arachnophobie, est un trait de caractère donnant une information importante sur la réaction d'une personne face au stimulus "araignée". Face à une mygale, certaines réactions telles que la fuite ou l'augmentation du rythme cardiaque seront communes de par notre instinct de survie. Cependant, une personne arachnophobe bien que présentant des signes communs, ressentira une émotion exacerbée. C'est cette information que nous souhaitons traiter afin de pondérer l'émotion brute détectée.

L'utilisation d'une ontologie comme socle pour ce module se justifie par le besoin de raisonnement. La question qui se pose ici est le moyen de faire le lien entre un trait de caractère et un objet de la vie courante. La capacité d'inférence permise par les ontologies et l'interopérabilité avec de vastes sources de données telles que dbpedia [1], permet de décrire n'importe quel type de trait de caractère répondant à la définition suivante : un trait de caractère est une modification du ressenti émotionnel

[1] http://dbpedia.org

(inhibition/galvanisation) en fonction d'un objet concret et dûe à l'expérience du sujet.

Dans ce chapitre, nous décrirons la structure et le fonctionnement de notre ontologie EmOCA [2] des émotions. Nous commencerons par définir ce que nous considérons être le contexte émotionnel dans le cadre de notre travail. Nous continuerons sur les raisons des choix technologiques pour arriver sur la modélisation de notre base de connaissances. Enfin, nous terminerons ce chapitre en décrivant l'exploitation de cette base avec d'une part, l'implémentation de celle-ci, suivi des outils permettant son interrogation.

1 Contexte émotionnel

Nous définissons la catégorisation comme la capacité du système à associer un terme de notre vocabulaire à une annotation émotionnelle. De même, nous définissons la contextualisation comme la capacité de notre système à utiliser le contexte de l'expérience émotionnelle pour pondérer la catégorisation. La notion de contexte est le point de départ pour la réalisation de cette phase de détection émotionnelle. Nous proposons une définition du contexte comme étant l'ensemble des informations ayant déclenché l'émotion ou permettant d'expliquer son apparition. La diversité et la complexité de ces données ne permettent pas une prise en compte globale du processus. A défaut de tenir compte des stimuli déclencheurs et des connaissances de l'individu sur ces derniers, nous proposons de considérer une information permettant d'expliquer certaines réactions disproportionnées par rapport à un stimulus donné. Cette information correspond aux différentes craintes ou attirances d'un individu face à un objet de la vie courante.

Par exemple, les phobies sont des traits de caractère impactant de manière négative l'expérience émotionnelle. L'arachnophobie en est un exemple qui caractérise les personnes ayant une forte répulsion pour les araignées ou plus généralement les arachnides. Á l'inverse, nous pouvons aussi considérer les philies qui impactent de manière positive le ressenti d'une émotion. La cynophilie caractérise les personnes ayant une forte attirance pour les chiens ou canidae.

Cependant, les phobies et les philies ne s'expriment pas de manière identique sur l'ensemble de la population. L'impact de ces dernières est sujet à variation ce qui empêche une caractérisation unique de ce comportement. De plus, ces traits de caractère sont activés par des stimuli précis. Dans le cas de l'arachnophobie, la présence d'une espèce d'araignée telle que la veuve noire ou d'une variété voisine comme une mygale ou un scorpion doit activer cette phobie. Bien que nous ayons

[2] http://ns.inria.fr/emoca

restreint le contexte à un sous-ensemble de l'environnement du ressenti émotionnel, la prise en compte de cette information comporte un coût de traitement non négligeable pour une implémentation classique. C'est pourquoi le choix technologique est ici crucial afin de limiter le coût en terme de modélisation mais également de capacité de raisonnement sur le contexte.

2 Choix technologique

Afin de modéliser les phobies et philies des utilisateurs, les approches classiques ne permettent pas une réalisation aisée et peu coûteuse en terme de ressource et de traitement. Nous proposons pour notre réalisation, l'utilisation de bases de connaissances afin de pouvoir connecter notre système à celles existantes, mais aussi dans le but de profiter des capacités de raisonnement sur les données de ces dernières.

Depuis l'apparition du web-sémantique en 2006, visant à structurer les données sur le web, de nombreuses bases de connaissances ont été développées reposant sur des standards issus du W3C. Ces standards développés pour le web favorisent l'interconnexion de ces sources d'informations. Le second intérêt des bases de connaissances est que le fondement théorique de ces dernières repose sur la logique de descriptions. Ceci permet d'introduire la notion d'inférence et de raisonnement afin de déduire des connaissances d'un ensemble structuré de données. La combinaison de ces deux caractéristiques permet de répondre aux problématiques que nous avons identifiées pour la modélisation du contexte émotionnel. D'une part, la réutilisation d'informations issues d'autres bases de connaissances nous permet de décharger la modélisation des stimuli et d'autre part, le raisonnement nous permet de retrouver les liens hierarchiques entre deux stimuli. Cet outil permet de fournir à l'utilisateur une façon de décrire ses traits de caractères de manière structurée, afin d'être utilisé pour améliorer la détection de ses émotions.

En reprenant l'exemple de la cynophilie, l'utilisation d'une base de connaissances nous permet dans un premier temps de définir le sujet activateur de ce trait de caractère. Pour notre exemple, nous utilisons la famille des "canis", récupérée de fr.dbpedia.org. Nous pouvons donc définir la "cynophilie" comme un trait de caractère stimulé par l'ensemble des animaux appartenant au genre "canis". A partir de cette définition, nous pouvons interroger notre base de connaissances afin d'identifier si ce trait de caractère est éveillé par le stimulus présenté à l'utilisateur. En montrant une photo représentant un "loup" ou un "dalmatien", nous pouvons grâce aux capacités d'inférence obtenir l'information suivante : le trait de caractère "cynophilie" est activé dans les deux cas. La figure 4.1 reprend cet exemple afin de montrer la logique recherchée pour l'identification de l'activation de ces traits de caractères.

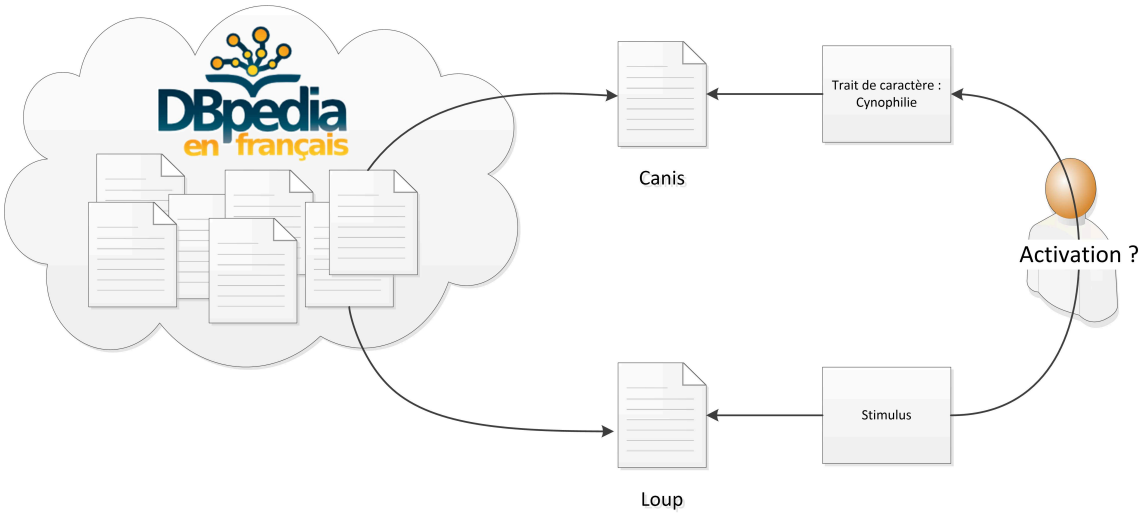


FIGURE 4.1 – Logique d'activation d'un trait de caractère à partir d'une base de connaissances

A partir de cette information, nous pouvons alors définir l'impact de ces traits de caractère sur le ressenti émotionnel et donc pondérer la catégorisation qui en découle.

La contextualisation que nous décrivons dans cette section est une première étape dans la prise en compte du contexte émotionnel. La définition des philies et des phobies est un exemple qui nous permettra de montrer l'intérêt du contexte dans la détection des émotions. La considération de ces traits de caractère motive également notre choix technologique vers une base de connaissances qui n'est pas sans impact sur le coût d'une telle implémentation.

3 Modélisation de EmOCA

3.1 Modélisation de la contextualisation

La modélisation de notre base de connaissances repose sur la création d'une ontologie permettant de structurer les connaissances sur les traits de caractère. L'objectif ici est de pouvoir décrire les philies et phobies d'un individu mais également l'impact de ces dernières sur le ressenti émotionnel. Une première étape consiste à contextualiser les annotations émotionnelles issues des couches inférieures de notre système, alors que la seconde, permet la catégorisation des émotions par la définition d'un vocabulaire contrôlé.

La figure 4.2 représente l'architecture générale de notre ontologie.

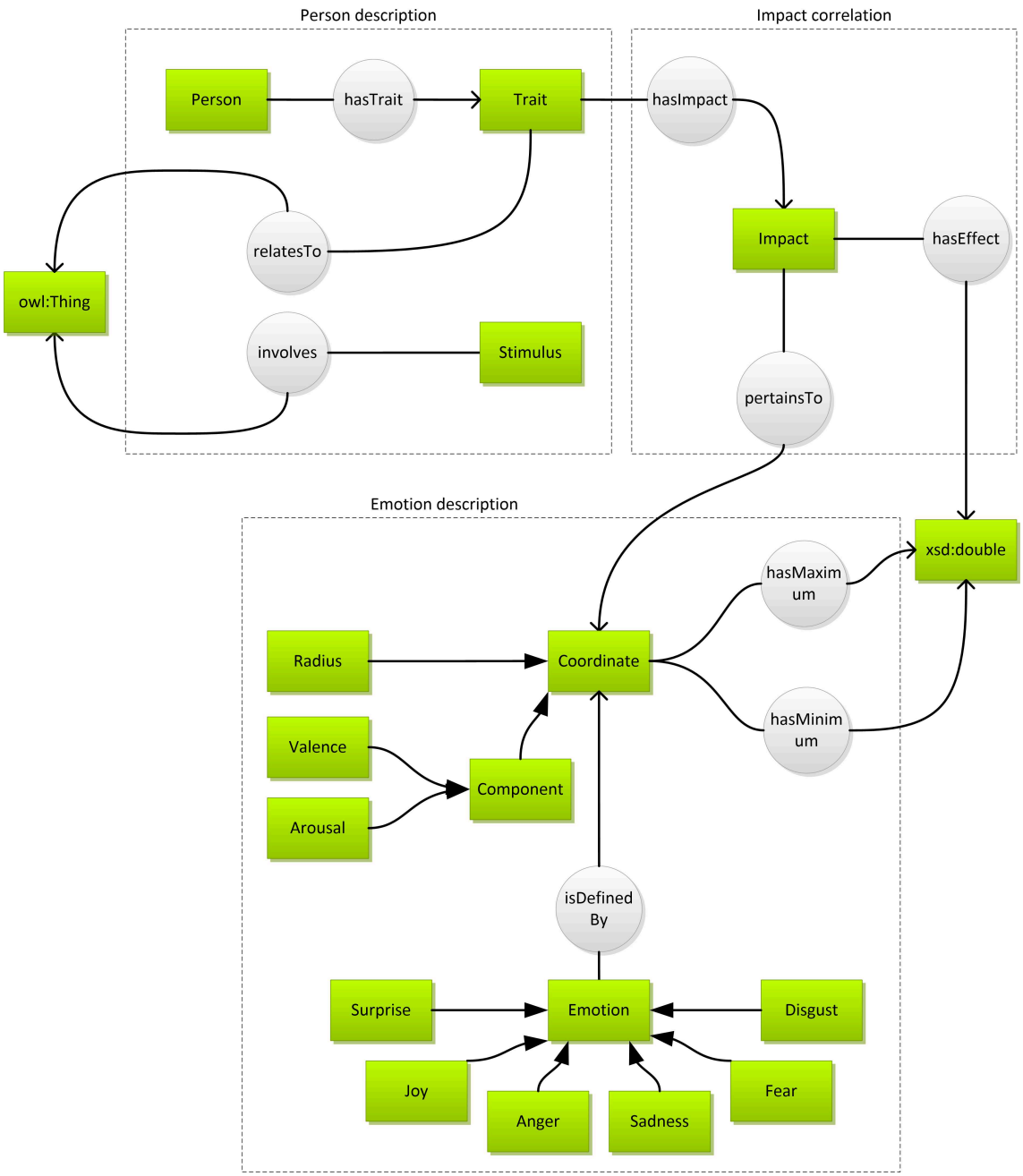


FIGURE 4.2 – Architecture globale de EmOCA

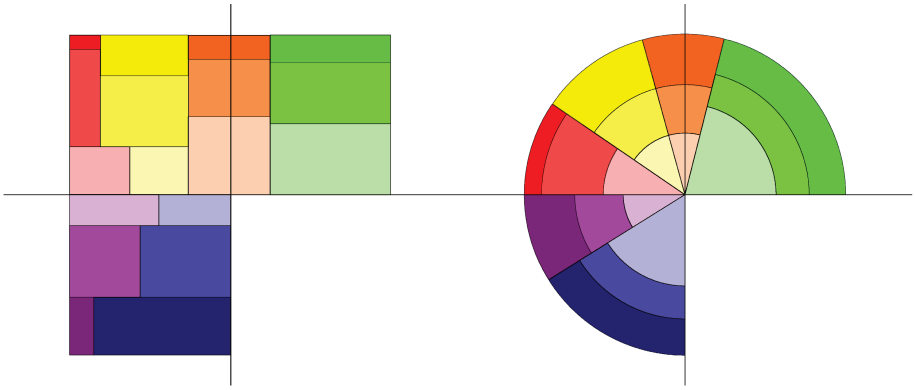
Nous avons découpé ce schéma en trois parties afin d'identifier les différents éléments nécessaires à la modélisation de notre contexte. Nous avons en premier, la représentation d'un concept "person" décrivant les individus utilisant notre système. Le concept même de personne n'est pas défini afin de respecter la logique de réutilisation et d'interconnexion des ontologies. Nous pouvons définir ce concept comme étant identique au concept "person" de l'ontologie Friend Of A Friend (FOAF) et ainsi profiter d'une description très précise des individus et des relations existantes entre eux. Ce que nous souhaitons ajouter à la définition d'une personne correspond

aux traits de caractère qu'il possède, ainsi que les stimuli auxquels il est confronté. A l'image de l'exemple précédent 4.1, les deux concepts "trait" et "stimulus" sont en relation avec un concept particulier "owl :thing" qui est le parent de l'ensemble des concepts. Cette définition permet donc la description d'un trait de caractère ou d'un stimulus en rapport avec des concepts issus de n'importe quelle base de connaissances.

3.2 Modélisation de la catégorisation

A l'autre extrémité de ce schéma, nous avons la description des émotions composée d'un vocabulaire contrôlé et de sa définition dans les modèles dimensionnels. Le vocabulaire utilisé correspond aux termes définis par Ekman [Ekman 71] dans sa première liste des émotions universelles. Nous avons opté pour ce modèle catégoriel du fait de sa simplicité et de la réutilisation de ces termes dans de nombreux modèles catégoriels. De plus, le nombre de termes décrits n'est pas limité, il est possible d'augmenter ou de réduire ce vocabulaire afin de correspondre aux besoins applicatifs. Il est également possible de décrire des intensités émotionnelles lors de la définition des individus peuplant notre base de connaissances.

La seconde partie de la modélisation des émotions correspond à la traduction des émotions dans les modèles dimensionnels. Cette partie nous permet d'interpréter les annotations dimensionnelles issues de notre module de classification en annotations catégorielles. Ceci est réalisé par la définition d'intervalles dans l'espace dimensionnel correspondant à un terme de notre vocabulaire. Nous proposons de définir ces intervalles grâces aux coordonnées cartésiennes en fournissant une valeur minimale et maximale sur chaque dimension des modèles, ou en coordonnées sphériques en définissant sur chaque dimension un intervalle angulaire et un intervalle correspondant au rayon. L'utilisation de ces deux types de coordonnées permet une définition plus flexible des intervalles. Le schéma 4.3 présente deux configurations possibles utilisant les deux types de coordonnées.



(a) Coordonnées cartésiennes (b) Coordonnées sphériques

FIGURE 4.3 – Utilisation des coordonnées cartésiennes et sphériques sur le modèle de Russell

L'utilisation des coordonnées sphériques, bien que plus difficiles à manipuler, trouve son intérêt lorsque l'on associe le rayon à l'intensité émotionnelle. Il devient alors beaucoup plus simple de définir des intensités émotionnelles en proposant des intervalles de rayon caractérisant chacune d'entre elles, alors que les intervalles angulaires restent les mêmes pour l'émotion définie.

Enfin, la dernière partie de notre ontologie définit l'impact d'un trait de caractère sur la catégorisation de l'émotion. Cette phase est réalisée par la définition d'une valeur numérique modifiant l'intervalle de description des émotions. Cette information permet de définir si une émotion se voit inhibée ou galvanisée par l'activation d'un trait de caractère.

4 Exploitation de la base de connaissances

4.1 Implémentation

Avant de pouvoir exploiter notre base de connaissances, nous devons en proposer une implémentation. Comme nous l'indiquions dans la section 3.1, un des points forts des bases de connaissances est leur capacité à s'interconnecter facilement entre elles. Ceci est possible grâce aux standards issus du W3C que nous souhaitons exploiter. Parmi ces recommandations, nous avons opté pour le langage RDFs de description d'ontologie couplé au langage RDF de description des triplets peuplant notre base. Ce choix est motivé par le niveau d'expressivité requis par notre système, ainsi que la compatibilité de ce langage avec des langages plus expressifs tels que OWL également standardisé par le W3C. Ce qui permet d'envisager une évolution vers ce type de langage si le besoin d'expressivité augmente sans pour autant devoir redéfinir ce que

nous proposons. Nous pensons aussi aux impératifs de temps réel que nous nous sommes fixés ce qui limite le raisonnement envisagé, afin de conserver un temps d'exécution acceptable. L'exploitation de cette base de connaissances s'effectuera grâce au langage de requête SPARQL Protocol And RDF Query (SPARQL) et au moteur d'inférence COnceptual REsource Search Engine (Corese).

La réalisation de notre base de connaissances s'est divisée en deux phases, une première visant à implémenter l'ontologie, et une seconde, permettant de peupler cette dernière avec des triplets, décrivant les instances exploitées par notre système. En nous basant sur la modélisation décrite dans les sections 3.1 et 3.2, nous avons pu décrire chacun des concepts de notre ontologie avec le langage RDFs. La représentation d'un concept est décrite de la manière suivante :

```
1 <rdfs:Class rdf:about="&emoca;Joy">
2   <rdfs:subClassOf rdf:resource="&emoca;Emotion" />
3 </rdfs:Class>
```

Alors que les propriétés reliant nos concepts sont décrites par cette syntaxe :

```
1 <rdf:Property rdf:about="&emoca;hasEffect">
2   <rdfs:domain rdf:resource="&emoca;Impact" />
3   <rdfs:range rdf:resource="&xsd;double" />
4 </rdf:Property>
```

Concernant la phase de population de notre base, nous avons dû définir un ensemble d'individus minimum permettant de mettre en exergue les capacités de raisonnement de notre outil de contextualisation. En respectant la structuration de notre ontologie, les triplets que nous fournissons permettent de décrire les utilisateurs de notre système, ainsi que les traits de caractère et les stimuli auxquels il sont soumis, mais également les émotions et leurs descriptions dans les modèles dimensionnels. Voici un exemple de la description d'une personne :

```
1   <emoca:Impact rdf:about="&base;ModeratelySuppressValence">
2     <emoca:pertainsTo rdf:resource="&emoca;Valence" />
3     <emoca:hasEffect>0.2</emoca:hasEffect>
4   </emoca:Impact>
5
6   <emoca:Impact rdf:about="&base;LightlyStimulateArousal">
7     <emoca:pertainsTo rdf:resource="&emoca;Arousal" />
8     <emoca:hasEffect>−0.2</emoca:hasEffect>
9   </emoca:Impact>
10
11  <emotion:Trait rdf:about="&base;Arachnophobia">
```

```
12      <emoca:relatesTo rdf:resource="http://umbel.org/umbel/rc/Arachnid" />
13      <emoca:hasImpact rdf:resource="&base;ModeratelySuppressValence" />
14      <emoca:hasImpact rdf:resource="&base;LightlyStimulateArousal" />
15    </emoca:Trait>
16
17    <emoca:Person rdf:about="&base;JohnDoe">
18      <emoca:hasTrait rdf:resource="&trait;Arachnophobia" />
19    </emoca:Person>
20
21    <emoca:Stimulus rdf:about="&base;ScorpionPhoto">
22      <emoca:involves rdf:resource="http://dbpedia.org/resource/Scorpion" />
23    </emoca:Stimulus>
```

Cet exemple montre une définition de l'utilisateur "John Doe" possédant un trait de caractère qui est l'arachnophobie. Cette phobie est décrite comme étant activée par les "arachnides" et impacte la valence négativement alors que son impact sur l'éveil est positif. Nous définissons également un stimulus permettant d'activer cette caractéristique par la présentation d'une photo de scorpion.

De même, nous pouvons décrire les émotions grâce à cette syntaxe :

```
1     <emoca:Arousal rdf:about="&base;ArousalFear">
2       <emoca:hasMinimum>0.55</emoca:hasMinimum>
3       <emoca:hasMaximum>1</emoca:hasMaximum>
4     </emoca:Arousal>
5     <emoca:Valence rdf:about="&base;ValenceFear">
6       <emoca:hasMinimum>0</emoca:hasMinimum>
7       <emoca:hasMaximum>0.45</emoca:hasMaximum>
8     </emoca:Valence>
9
10    <emoca:Fear rdf:about="&base;FearNormal">
11      <emoca:isDefinedBy rdf:resource="&base;ValenceFear" />
12      <emoca:isDefinedBy rdf:resource="&base;ArousalFear" />
13    </emoca:Fear>
```

Cet exemple montre la description de la peur dans notre système qui est caractérisée par un intervalle sur l'axe de l'éveil et de la valence compris entre 0.45 et 0. Nous pouvons également définir des intensités pour cette émotion en proposant un nouveau triplet dont l'intervalle est une partie de celui de la peur.

4.2 Interrogation de la base de connaissances

L'implémentation de l'ontologie nous a permis de structurer les connaissances du contexte de l'expérience émotionnelle ainsi que la standardisation d'un vocabulaire. L'objectif ici est de raisonner sur ces informations, afin d'identifier si le stimulus déclencheur est en relation avec un trait de caractère de l'utilisateur, et d'appliquer l'impact de celui-ci sur l'annotation brute, afin de sélectionner le terme de notre vocabulaire compatible avec notre pondération. Pour réaliser ce raisonnement, nous utilisons le moteur d'inférence Corese et le langage de requête SPARQL. La requête que nous allons détailler dans cette section est exécutée à chaque fois qu'une nouvelle annotation brute est calculée par notre module de classification.

Le fonctionnement global de cette requête est décrit par la figure 4.4. Nous identifions trois parties de raisonnement exécutées en séquence avec en premier, la détermination du trait de caractère, activé par le stimulus présenté à l'utilisateur, suivi par la pondération de l'annotation brute en fonction de l'impact du trait de caractère,et pour terminer, la sélection d'un terme de notre vocabulaire.

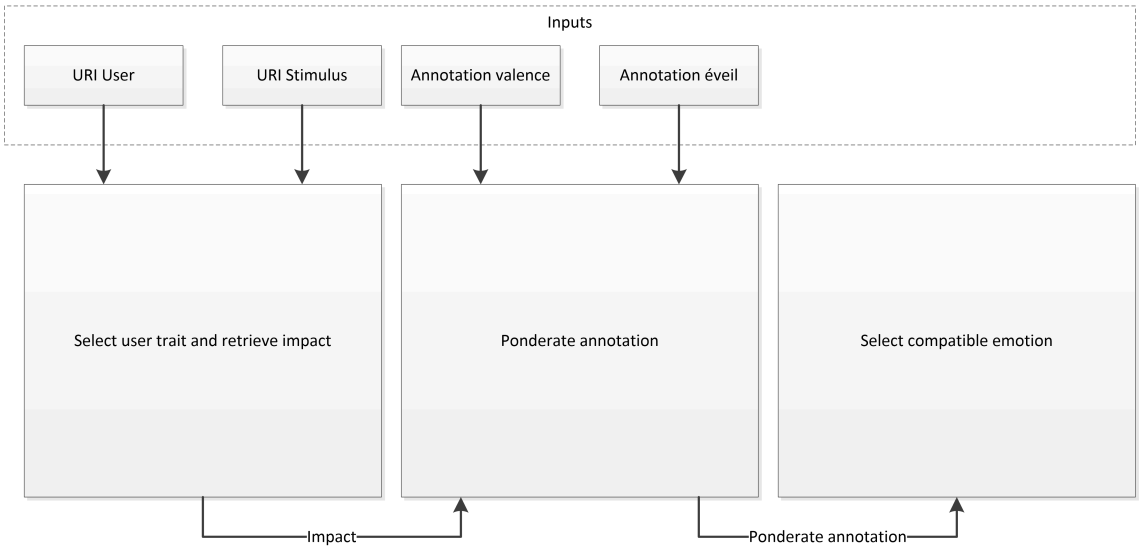


FIGURE 4.4 – Schéma global de l'exécution de la requête SPARQL

Pour décrire le fonctionnement de notre requête, nous allons prendre l'exemple de l'utilisateur "JohnDoe" auquel nous allons présenter une image de "dalmatien". Les phases précédentes de notre détection ont permis de trouver une annotation correspondant à l'émotion dans le modèle de Russell à savoir une valeur de 0.6 pour la valence et 0.9 pour l'éveil. Nous montrerons le déroulement du raisonnement appliqué dans notre base de connaissance en trois étapes afin d'obtenir un terme de notre vocabulaire correspondant à l'émotion contextualisée.

La première partie de notre raisonnement permet d'identifier les liens sémantiques entre les traits de caractère et le stimulus présenté. Dans la figure 4.5, nous prenons l'exemple de l'utilisateur "JohnDoe" auquel nous présentons une image de "dalmatien".

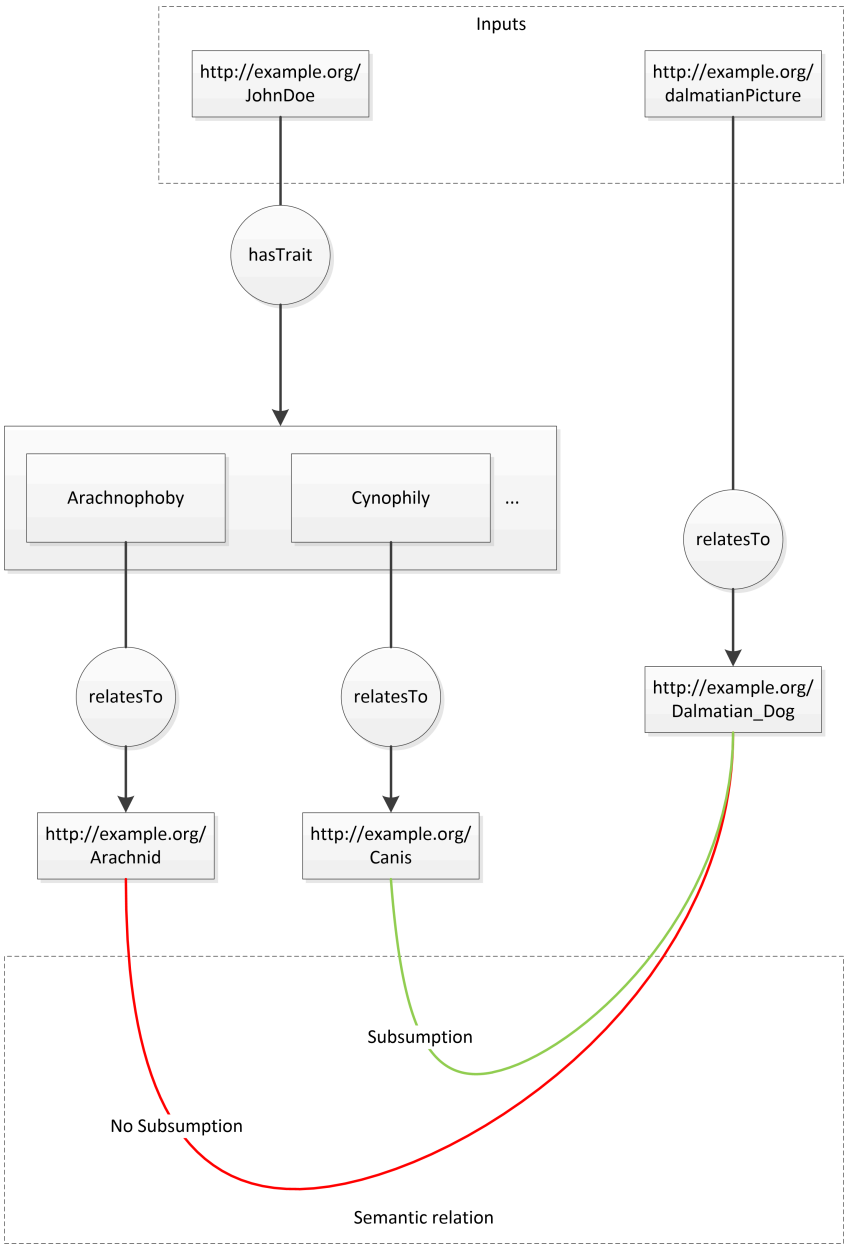


FIGURE 4.5 – Détermination du trait de caractère activé

Nous utilisons les Uniform Resource Identifier (URI) des triplets pour caractériser à la fois la personne et le stimulus. A partir de ces deux informations, nous interrogeons d'une part, la base pour obtenir les traits de caractère de l'utilisateur, ici "JohnDoe", et d'autre part, le sujet du stimulus. Dans cet exemple, nous pouvons savoir que l'utilisateur est à la fois arachnophobe mais également cynophile.

Afin de pouvoir raisonner sur ces caractérisques, nous recherchons les concepts en relation avec ces phobies et philies. Ici, nous déterminons que l'arachnophobie est activée en présence d'arachnides et que la cynophilie est activée en présence d'animaux de la famille canis. Le stimulus quant à lui a pour sujet un chien de la race des dalmatiens. Le raisonnement commence à partir de ces informations afin de retrouver les liens sémantiques entre les arachnides ou les canis, et le dalmatien. Dans le premier cas, aucun lien hiérarchique n'est identifiable car le dalmatien n'est pas une sous-espèce d'arachnide alors que le dalmatien appartient bien à la famille des canis. Cette information nous permet de sélectionner le trait de caractère activé par ce stimulus.

Une fois le trait de caractère identifié, nous pouvons déterminer l'impact de ce dernier sur l'annotation émotionnelle brute. Cet impact correspond à la pondération que nous appliquons à l'annotation dimensionnelle. La figure 4.6 présente ce processus. La définition d'un trait de caractère est composée à la fois de l'objet qui l'active mais également de l'impact sur l'émotion qu'il induit. Ces deux informations sont définies par l'utilisateur du système afin de personnaliser son ressenti émotionnel face à un sujet donné. Dans notre exemple, c'est la cynophilie de notre utilisateur qui est activée.

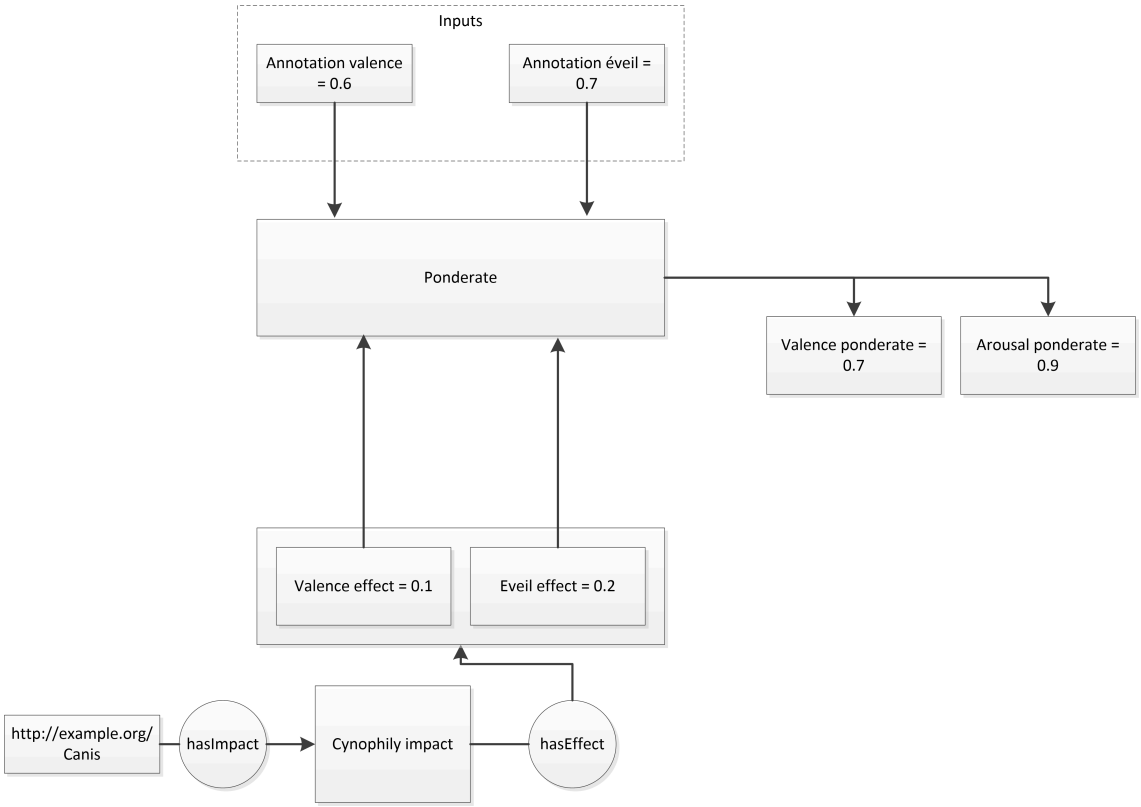


FIGURE 4.6 – Pondération de l'annotation émotionnelle brute

Nous pouvons alors récupérer l'impact de cette philie qui est décrite par l'utilisateur. L'annotation initialement classifiée est de 0.6 pour la valence et 0.7 pour l'éveil, avec un impact pour la cynophilie de +0.1 sur la valence et +0.2 sur l'éveil, nous obtenons une nouvelle annotation brute pondérée par le contexte émotionnel, qui est de 0.7 pour la valence et 0.9 pour l'éveil. A ce stade, un utilisateur non expert ne peut toujours pas identifier l'émotion ressentie.

La dernière étape de notre traitement consiste en la détermination d'un terme de notre vocabulaire compatible avec l'annotation contextualisée. Ceci se fait par la comparaison de notre annotation avec l'ensemble des intervalles émotionnels décrits dans notre base de connaissances. Si la description de l'émotion est compatible sur l'ensemble des axes du modèle dimensionnel, alors l'annotation contextualisée est décrite par ce concept. Ici, deux émotions sont décrites, la peur et la joie, avec des intervalles sur les dimensions de valence et d'éveil. La comparaison montre qu'il y a un problème de compatibilité avec la peur où l'intervalle de valence ne comprend pas l'annotation pondérée, alors que la joie est compatible avec cette annotation. L'annotation contextualisée et catégorisée est représentée par le nom du concept qui est fourni en sortie de notre système.

C'est l'information de plus haut niveau que nous proposons pour notre système. Cette annotation est alors exploitée par l'application utilisant notre détecteur pour effectuer une action en fonction des différentes émotions ressenties par l'utilisateur. Dans le cadre des "serious games", il est possible de modifier l'environnement ou le comportement du système en fonction des émotions décrites dans notre base de connaissances. Cet ensemble d'émotions limite les réponses de notre système et permet ainsi une détection limitée à un ensemble restreint d'émotions. L'annotation brute contextualisée peut également être fournie, afin d'évaluer de manière précise l'intensité de l'émotion ressentie.

Le schéma 4.7, représente la dernière étape de la catégorisation. Ici, le but est de trouver le terme compatible avec l'annotation pondérée précédemment dans la phase de contextualisation. Nous pouvons voir que chaque émotion est définie par un intervalle sur l'axe de l'éveil et un sur l'axe de la valence. Pour cet exemple, l'annotation pondérée est compatible uniquement avec la joie, puisque dans le cas de la peur, l'intervalle caractérisant cette émotion ne comprend pas la valeur de l'annotation.

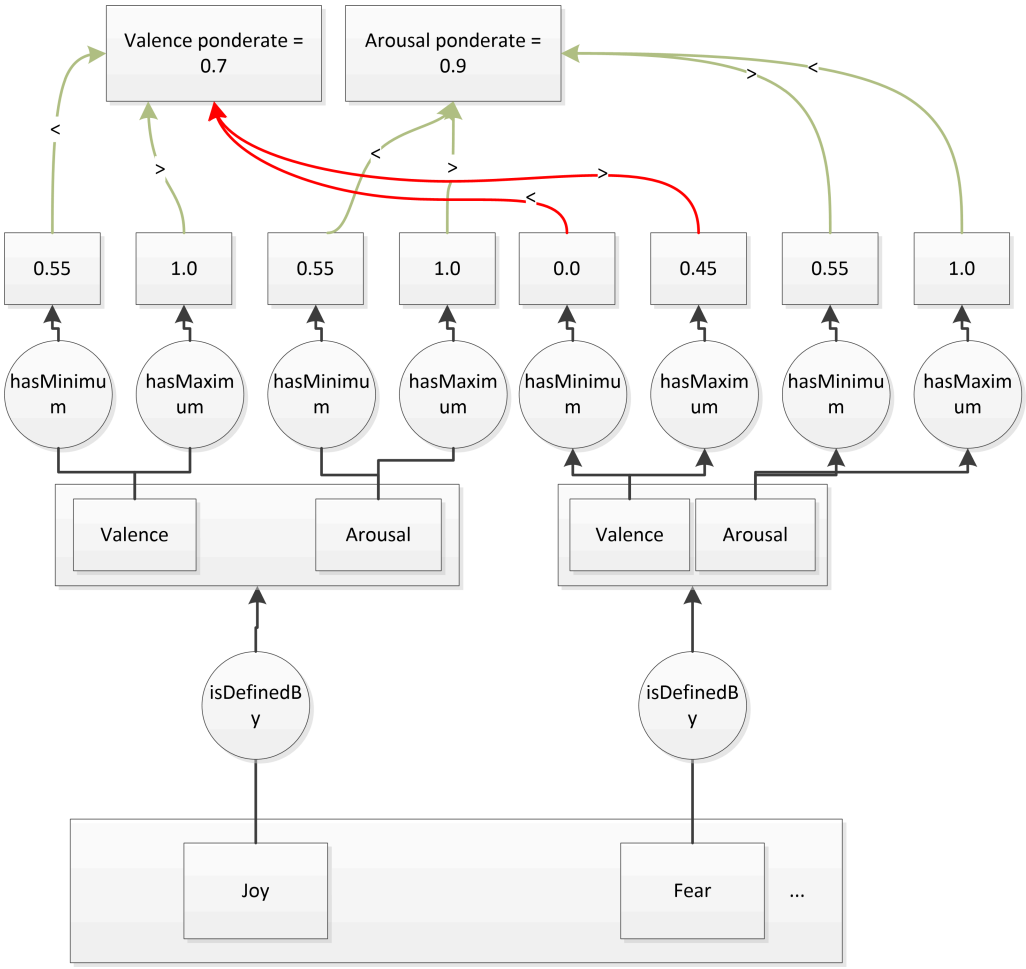


FIGURE 4.7 – Sélection du (ou des) terme(s) émotionnel(s) compatible(s) avec la pondération

5 Conclusion

Dans ce chapitre, nous avons présenté le fonctionnement de la contextualisation et de la catégorisation des annotations brutes. Cette étape finale de notre processus de détection commence après la classification des émotions et se termine par l'exposition d'une annotation brute contextualisée et d'un terme issu de notre vocabulaire standardisé. Nous avons commencé par définir le contexte émotionnel comme étant l'ensemble des informations impactant la compréhension d'une expression émotionnelle, afin d'approcher au mieux le ressenti émotionnel de l'utilisateur. Pour notre réalisation, nous avons montré l'implémentation d'une partie de ce contexte en modélisant la notion de phobie et de philie.

La prise en compte de ce contexte est réalisée par la modélisation d'une base de connaissances. Nous avons justifié ce choix technologique par les capacités de

raisonnement et d'interconnexion issues de la logique de description. De plus, l'engouement pour le web sémantique depuis 2006 a vu apparaître une grande variété de bases pouvant être réutilisées par notre système afin de décharger la définition des stimuli dans ces dernières. Le raisonnement que nous avons montré se concentre sur l'identification de liens sémantiques entre les sujets, les traits de caractère et les stimuli.

Nous avons également présenté l'implémentation de cette base de connaissances réalisée en RDFs pour l'ontologie et RDF pour les triplets peuplant notre base de connaissances. Cette implémentation a permis de déterminer la syntaxe des différents triplets que nous souhaitons utiliser dans notre système, afin de définir les personnes, les traits de caractère et les émotions.

Enfin, nous avons montré comment nous exploitons cette base de connaissances dans notre processus de détection émotionnelle. L'utilisation de notre base est réalisée par l'intermédiaire du moteur d'inférence Corese et du langage de requête SPARQL. Nous avons montré que l'exécution de cette requête est réalisée pour chacune des classifications effectuées par notre module CEP. Nous avons présenté le fonctionnement de notre requête à partir d'un exemple concret afin de mettre en évidence les trois étapes de raisonnement.

Chapitre 5

Résultats et discussion

Dans les chapitres précédents, nous avons explicité les différents modules utilisés au sein de notre prototype de détecteur émotionnel. Dans ce dernier chapitre, notre objectif est de fournir des informations qualitatives et quantitaves sur les résultats obtenus pendant notre expérimentation. De plus, nous présenterons une utilisation de notre prototype ainsi que le protocole de quantification utilisé.

Notre problématique initiale nous a poussé à définir un modèle des émotions à la fois flexible et interopérable pour s'adapter à une grande variété de contextes d'utilisation. Lors de l'élaboration d'un tel système deux questions ont émergé. D'une part, nous souhaitons fournir un système fiable pouvant répondre de la manière la plus précise possible sur l'état émotionnel d'une personne et cela dans un délai assimilable à du temps réel pour les émotions. D'autre part, nous proposons une première implémentation à notre connaissance du modèle complexe de Scherer. C'est pourquoi, nous voulons également définir si cette modélisation des émotions correspond aux hypothèses émises lors de son élaboration.

Pour répondre à notre première interrogation, nous proposons dans ce chapitre une quantification de notre taux de bonne classification. Cela se concentrera sur la mise au point d'un protocole de test basé sur un capteur facilement contrôlable (la caméra) afin d'écarter des problèmes de traitement du signal ou de mauvaise interprétation des signaux. Notre modèle étant divisé en modules, nous proposerons deux quantifications, une pour les CEPs et une pour la base de connaissances. Pour les cartes, nous utiliserons la notion de déviation afin de déterminer si la détection tend vers des valeurs cohérentes avec l'expression faciale exprimée. Pour l'ontologie, nous examinerons la qualité de catégorisation de cet outil.

Concernant notre seconde interrogation, nous proposons de vérifier la présence de discontinuités de type hystérésis lors de la détection. Comme présenté sur la figure 1, ce type de catastrophe est utilisé pour représenter la non spécificité d'une émotion

en fonction d'un stimulus. La présence d'un tel comportement dans la classification des émotions met en évidence l'intérêt de l'observation des émotions sous l'angle des systèmes complexes. De plus, la quantification précédente donnera une indication quant à la viabilité de son utilisation dans le cadre de la détection.

Enfin, nous discuterons autour des résultats obtenus afin de proposer des évolutions futures pour notre système.

1 Evaluation des performances de détection émotionnelle

Les premiers résultats que nous souhaitons valoriser sont les performances de détection de notre système. Nous présenterons dans cette section trois niveaux d'évaluations afin de valider le choix de notre algorithme de détection basé sur l'algorithme KNN [Fig. 3.4] en montrant l'influence de la constante K de notre algorithme, puis nous quantifierons la détection sur des séquences inconnues, pour enfin terminer sur l'intérêt de la contextualisation des émotions.

1.1 Méthodologie

Comme expliqué précédemment, notre modèle possède deux niveaux d'accès pour l'annotation de l'émotion. D'une part, nous avons l'annotation dite "bas-niveau" correspondant à un vecteur de dimension n, répondant au nombre de dimensions du modèle psychologique cible. Pour notre expérimentation, nous utiliserons le modèle de Russell de dimension 2 (valence / éveil). D'autre part, nous avons l'annotation dite de "haut-niveau" qui est une catégorisation de l'annotation précédente dans un modèle psychologique catégoriel ou un thésaurus émotionnel. Dans ce cas précis, nous utilisons les émotions basiques de Ekman complétées par les intensités correspondantes du modèle circomplexe de Plutchik.

Nous divisons donc notre méthodologie en trois étapes. Premièrement, nous définssons la méthode de construction de notre base de séquence vidéo de référence et la méthode d'annotation de cette dernière. Ensuite, nous calculerons le niveau de déviation entre la détection automatique de bas-niveau et l'annotation manuelle effectuée sur cette première séquence, afin d'identifier la meilleure valeur de la constante K de notre algorithme de classification. Enfin, nous montrerons la capacité de catégorisation sur une séquence video inconnue du système.

Construction de la base d'apprentissage

L'utilisation de la modalité faciale pour la validation de nos travaux est motivée par l'aptitude humaine à reconnaitre les émotions chez autrui. Dans un contexte réel, nous identifions les émotions dans deux cas, soit pour définir l'émotion ressentie soit pour identifier les émotions exprimées par un autre individu. Dans le premier cas, nous utilisons l'ensemble des informations pour catégoriser notre émotion, mais cette étape est instinctive et difficilement reproductible face à des données issues de capteurs physiologiques. Dans le second cas en revanche, nous utilisons des informations essentiellement visuelles comme les expressions faciales ou les postures. L'utilisation d'une caméra permet d'exploiter cette capacité afin de reconnaitre les émotions sur une video et donc d'effectuer une annotation manuelle pour peupler notre base d'apprentissage. Notre seconde motivation concerne la facilité de contrôle de cette modalité. L'elicitation des émotions est un sujet qui fait encore débat au sein de la communauté des psychologues. L'utilisation de capteurs physiologiques nous semble trop couteux en terme de temps et de ressources pour obtenir nos premiers résultats. De plus, une mauvaise elication des émotions peut induire des erreurs dans la classification, ce qui rendrait difficile l'identification des points faibles de notre système. C'est pourquoi nous optons pour une approche visant à mimer les expressions émotionnelles afin de garder le contrôle sur l'émotion exprimée.

Capture : Afin d'observer l'évolution des expressions émotionnelles, nous avons dû construire notre propre jeu de données bien que certains travaux tendent à standardiser ce type d'informations [Healey 02]. Pour cette première capture de notre jeu de données, nous avons demandé à un utilisateur de notre système, que nous définirons comme un acteur dans la suite du document, de mimer les expressions faciales des six émotions basiques de Ekman [Fig. 5.1] avec un retour au neutre entre chaque.

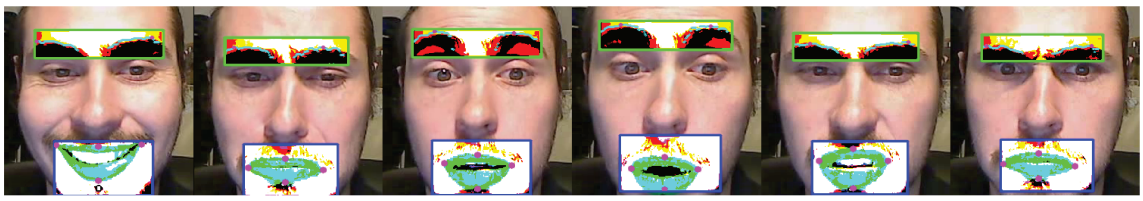


FIGURE 5.1 – Images extraites de la vidéo d'apprentissage représentant les 6 expressions faciales de Ekman

Nous souhaitions également observer la variation d'intensité dans l'expression émotionnelle, c'est pourquoi nous avons demandé de pousser les expressions au maximum de leur intensité [Fig. 5.2]. La séquence a été reproduite à plusieurs reprises,

afin d'accumuler des variantes d'expressions pour chacune des émotions. Cette capture a été effectuée en une seule prise, afin d'écarter des problèmes de luminosité ou de positionnement pouvant fausser les résultats.

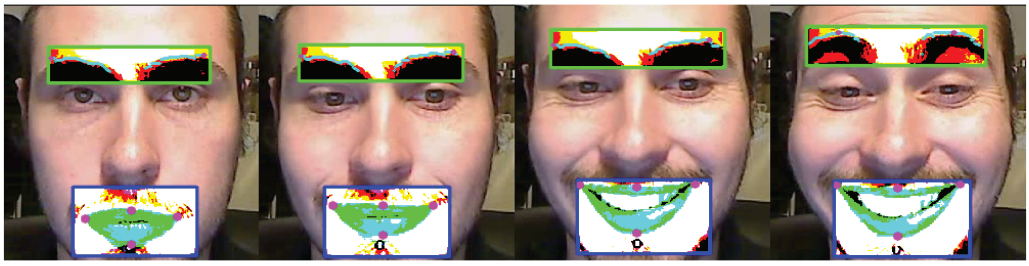


FIGURE 5.2 – Images extraites de la vidéo d'apprentissage représentant les variations d'intensité pour la joie

Extraction des séquences : Une fois la vidéo obtenue, nous avons procédé à un découpage et un nettoyage du fichier afin d'obtenir une suite de séquences contenant chacune une émotion distincte. La phase de nettoyage a consisté à supprimer les séquences non utilisables et réduire les séquences neutres. Pour la découpe, nous avons conservé quelques images neutres avant et après l'expression de l'émotion.

Avec ces séquences, nous avons recréé deux videos. La première contient une séquence de chaque émotion soit 1232 images, nous la désignerons comme vidéo d'apprentissage. La seconde contient 17 séquences soit 5757 images, nous la désignerons comme vidéo de test.

	Nb images	Nb émotions	Nb séquences
Vidéo d'apprentissage	1232	6	6
Vidéo de test	5757	6	17

TABLE 5.1 – Tableau récapitulatif des vidéos utilisées pour l'analyse des performances

Annotation manuelle : Cette dernière tâche correspond à l'annotation manuelle des 1232 images de la vidéo d'apprentissage avec un point dans l'espace de Russell. Nous avons fait correspondre chaque expression faciale à une annotation correspondant au niveau d'intensité et d'émotion exprimé sur le cliché. Ces annotations seront utilisées comme base de référence pour notre analyse. La figure 5.3 représente la répartition des annotations pour la séquence de joie.

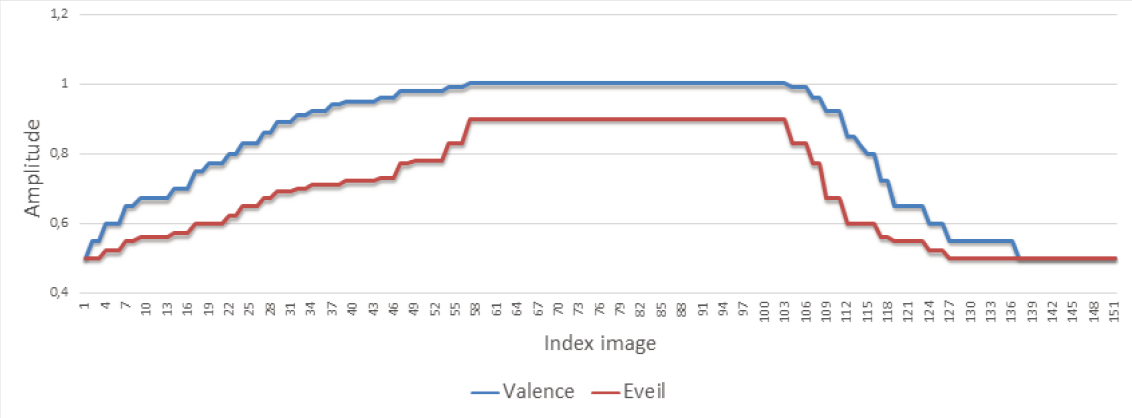
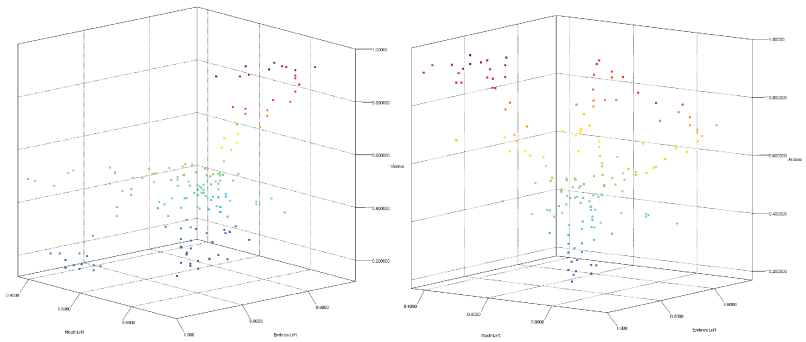


FIGURE 5.3 – Répartition des annotations pour une séquence représentant l'émotion joie

Création des CEPs

A partir de l'annotation manuelle précédente, nous créons nos CEPs. La construction des cartes consiste en la définition d'images caractéristiques auxquelles nous affectons l'annotation manuelle correspondante, chacune de ces images nous donnant 8 points caractéristiques (voir 2.1) de l'expression faciale que nous appelerons capteur.

Pour cette expérimentation, nous avons sélectionné 163 images de la vidéo. Nous avons donc obtenu une carte par composante de l'émotion (valence / éveil) contenant chacune 163 états annotés répartis en 6 séquences. La figure 5.4 présente une projection de nos deux cartes sur les capteurs "commissure des lèvres gauche" et "sourcil gauche". Les cartes complètes étant de dimension 9, nous ne présenterons ici que des projections sur deux capteurs maximum, même si la classification s'effectue sur l'ensemble des dimensions.



(a) CEP de valence (b) CEP d'éveil

FIGURE 5.4 – CEPs de valence et d'éveil pour les expressions faciales

Ces cartes serviront pour la classification des images de nos vidéos. L'annotation obtenue en sortie de notre algorithme de classification est nommée annotation automatique. Elle est composée de deux valeurs, une pour la valence, l'autre pour l'éveil, ainsi que de deux valeurs de confiance correspondant à l'inverse de la distance entre l'état inconnu et les états annotés de nos cartes.

Indicateurs de performances

Pour quantifier les performances de notre système, nous définissons la notion de déviation. Cette information correspond à la différence entre l'annotation manuelle et l'annotation automatique. Nous calculerons la déviation moyenne par composante, par émotion et pour l'ensemble des émotions.

Nous définissons donc la déviation pour une composante par la formule suivante :

$$\Delta = \frac{\sum_i^n (anno_{man_i} - anno_{auto_i})}{n} \tag{5.1}$$

Pour la déviation moyenne d'une émotion, nous utilisons la formule :

$$\mu_{emotion} = \frac{\sum_i^n |\Delta_i|}{n} \tag{5.2}$$

La déviation moyenne globale quant à elle est définie par :

$$\mu = \frac{\sum_i^n |\mu_{emotion_i}|}{n} \tag{5.3}$$

Nous utilisons ce premier indicateur afin d'observer la variation de la détection en fonction de l'émotion recherchée. Le second indicateur permet d'observer la stabilité générale du système lors de changement de certains paramètres comme notre constante de classification ou le nombre d'états initialement annotés. Enfin, notre dernier indicateur permet de fournir une vision globale de la capacité de détections toutes émotions confondues.

Le schéma 5.5 représente la méthodologie complète mise en oeuvre pour la quantification des performances de notre système. La première partie est réalisée avec l'utilisateur alors que le reste du processus est réalisé à posteriori. Nous pouvons donc reproduire cette analyse afin de voir l'évolution de notre système en fonction de modifications internes.

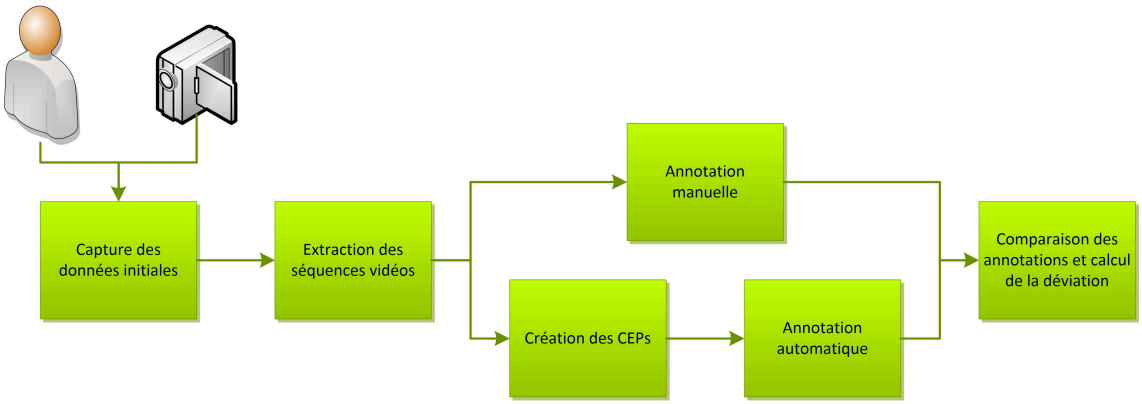


FIGURE 5.5 – Récapitulatif de la méthodologie de test

1.2 Résultats

L'application de la méthodologie précédente est utilisée pour deux raisons. D'une part, nous utiliserons cette procédure dans le but d'identifier la valeur de notre constante K de notre algorithme de classification, et d'autre part, nous observerons la fiabilité de notre système dans les conditions de performances optimales.

Détermination de la constante K

Afin d'identifier la valeur optimale de notre constante, nous avons reproduit l'analyse précédente en faisant varier la valeur de K. Pour ce faire, nous avons repris notre séquence vidéo d'apprentissage dont nous avions extrait 13% des images pour la génération de nos cartes, le but étant ainsi d'observer la qualité de l'annotation automatique sur des expressions préalablement annotées.

Les figures 5.6, 5.7, 5.8 présentent trois histogrammes contenant chacun la déviation par émotion et la déviation moyenne pour des valeurs de K différentes.

Les premiers résultats [Fig. 5.6] obtenus pour une valeur de K égale à 1 montrent un niveau d'annotation très bon pour l'ensemble des émotions sauf pour la peur. Une analyse manuelle des résultats brutes montre que cette mauvaise classification est due à la confusion entre peur et surprise. C'est pourquoi nous avons un taux d'erreur si important sur la valence. L'annotation que nous donnons pour la surprise dans le modèle de Russell correspond au vecteur $\binom{0.5}{1}$ alors que celui de la peur est $\binom{0.1}{1}$. La déviation moyenne quant à elle reste faible à cause des taux de bonne classification des autre émotions. Nous pouvons noter en particulier les valeurs très basses des déviations de la joie et de la colère. Nous pouvons expliquer ces résultats très faibles par l'utilisation d'une valeur de K à 1. En effet, la méthode de calcul de notre

algorithme de recherche se base sur le barycentre des points les plus proches. Dans ce cas précis, l'utilisation d'un seul voisin fournit obligatoirement une annotation existante, engeandrant ainsi des différences d'annotations nulles dans les cas de bonne classification.

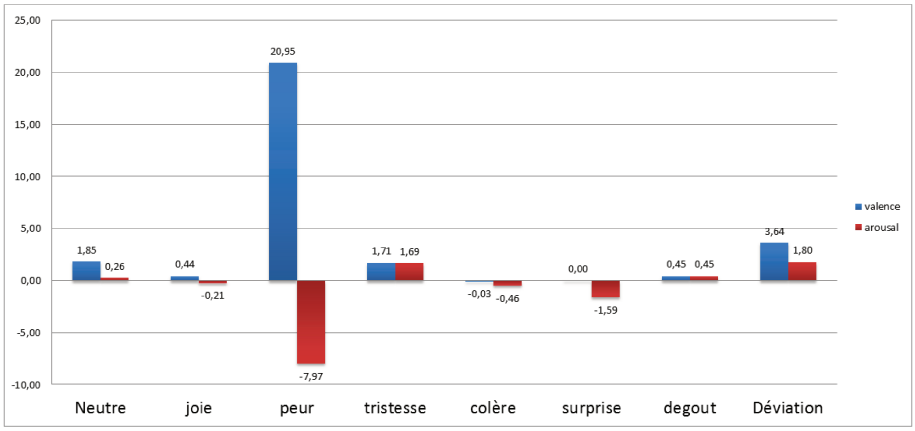


FIGURE 5.6 – Déviation pour $K = 1$

Ce second histogramme [Fig. 5.7] correspond à la meilleure classification que nous avons pu obtenir. L'information la plus importante de ce graphique est la disparition du pic de déviation pour la peur. En utilisant une valeur de K égal à 6 le système est capable de faire la différence entre les émotions de peur et surprise. En revanche, nous pouvons noter que la déviation de certaines émotions a augmenté par exemple pour la joie et la colère. Nous avons des taux qui ont considérablement augmenté. Cependant, la déviation moyenne reste stable du fait du gain considérable effectué sur la peur. Ces modifications sont directement liées à notre classifieur qui pour une valeur de K si importante produit de nouvelles annotations.

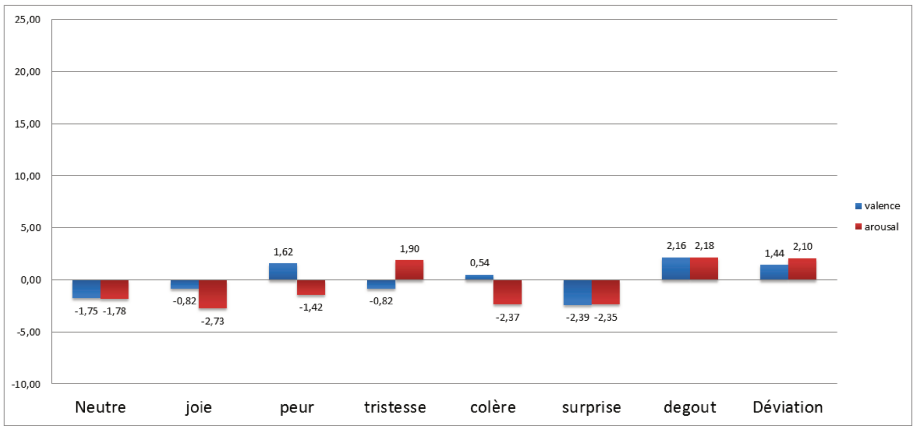


FIGURE 5.7 – Déviation pour $K = 6$

Enfin, ce dernier exemple [Fig. 5.8] pour une valeur de K égal à 15 montre une augmentation de la déviation pour la majorité des émotions. La déviation moyenne est également impactée, par cette modification. Nous observons donc pour cette valeur de K une perte de performance générale par rapport au test précédent sans pour autant obtenir un gain particulier.

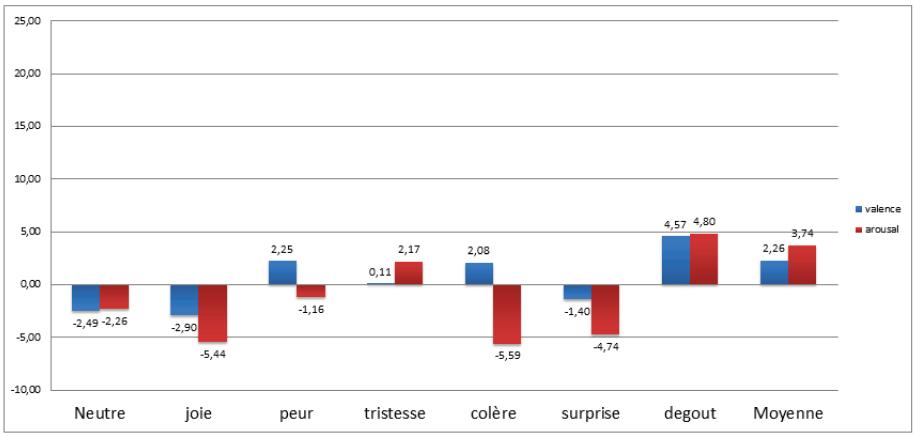


FIGURE 5.8 – Déviation pour $K = 15$

Pour conclure, le graphique 5.9 présente la déviation moyenne par émotion et globale pour différentes valeurs de K.

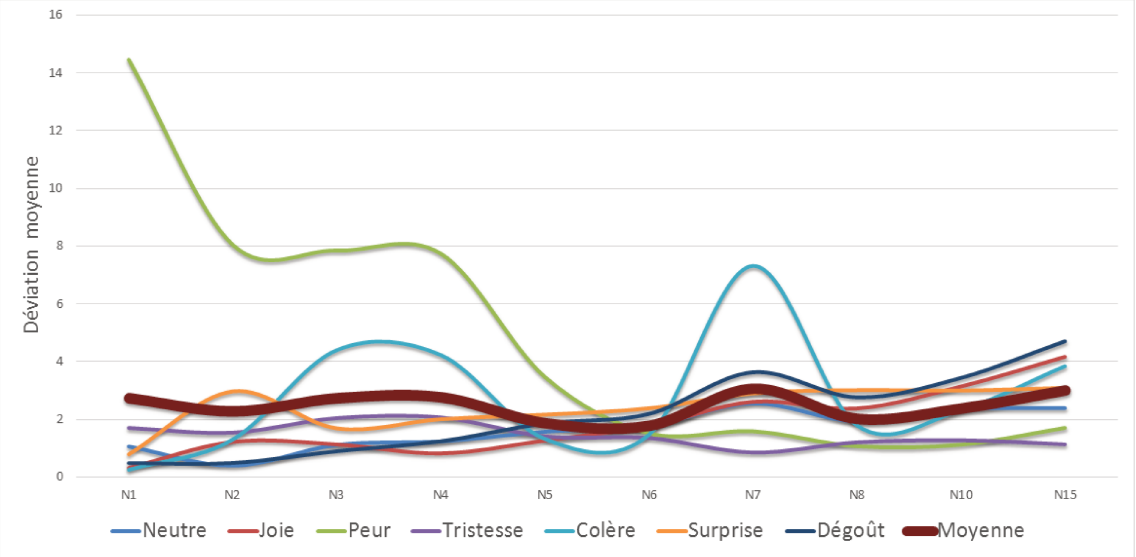


FIGURE 5.9 – Déviation moyenne en fonction de la valeur de K

Nous observons donc sur ce dernier graphique une forte diminution de la déviation de la peur pour des valeurs de K inférieur à 6. Cependant, comme identifié précédemment, l'augmentation de la valeur de K impacte également la déviation des

autres émotions. Le gain obtenu pour la peur se réduit considérablement à partir de K égal 6, nous avons donc opté pour l'utilisation de cette dernière valeur qui est pour nous le meilleur compromis dans le cadre de cette expérimentation.

Annotation automatique sur la vidéo de test

La deuxième étape de notre quantification se résume par l'annotation automatique de la vidéo de test. Celle-ci, n'ayant pas été utilisée pendant la phase de construction des cartes, contient des variantes d'expressions faciales pour les émotions définies. Nous comparons ici l'annotation brute issue du modèle de Russell ainsi que la catégorisation de cette dernière par l'ontologie. Le cadre de l'expérimention ne permet pas l'évaluation de la prise en compte du contexte au sein du raisonnement. L'acteur ayant pour tâche de mimer des émotions, il n'est pas soumis à un stimulus déclencheur.

Pour effectuer ce travail, nous avons dupliqué la vidéo de test. Dans une première version, nous regroupons les séquences par émotion alors que la deuxième version contient des séquences ordonnées aléatoirement. Cette seconde vidéo à pour but de montrer les capacités d'annotation dans un contexte plus proche de la réalité où l'utilisateur ne reproduit pas constamment les mêmes émotions.

La figure 5.10 montre l'annotation automatique d'une partie de la vidéo de test contenant uniquement des expressions de joie. Ici, nous pouvons facilement identifier trois expressions faciales séparées par un retour au neutre.

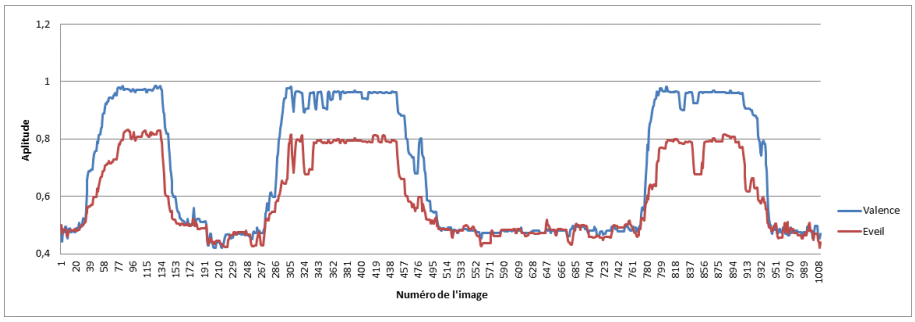


FIGURE 5.10 – Annotation automatique d'une séquence vidéo regroupant trois expressions de la joie

La figure 5.11 quant à elle présente l'annotation automatique de la vidéo de test contenant uniquement l'expression de la colère. Nous pouvons également observer l'apparition de trois séquences exprimant cette émotion. Cependant, la dernière séquence annotée montre une intensité de l'émotion inférieure aux précédentes.

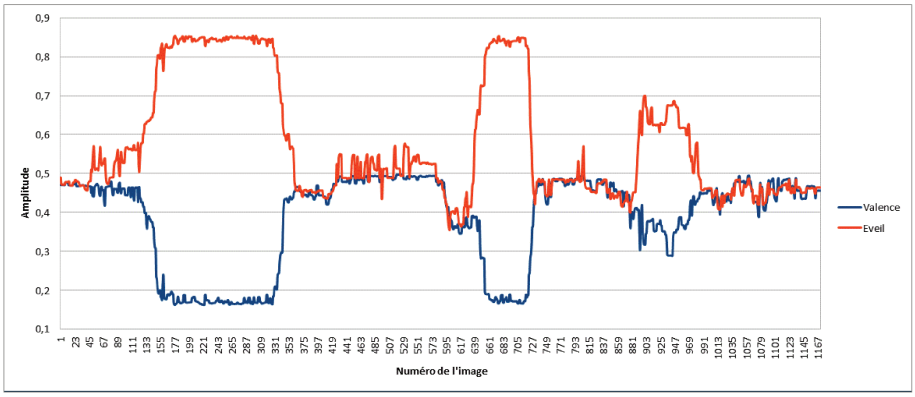


FIGURE 5.11 – Annotation automatique d'une séquence vidéo regroupant trois expressions de la colère

Enfin, la figure 5.12 montre l'annotation automatique de l'ensemble de la vidéo de test triée aléatoirement. Chaque rectangle réprésente l'émotion exprimée sur la vidéo. La première observation que nous pouvons effectuer ici est que l'ensemble des émotions exprimées ont été détectées. Néanmoins, en observant l'annotation automatique, nous pouvons identifier quelques erreurs d'annotations. Les deux dernières expressions du dégoût sont détectées comme une peur ou colère de faible intensité. De plus, l'avant-dernière expression de la surprise contient deux erreurs d'annotations en début et fin d'expression même si la partie centrale identifie bien l'émotion de surprise.

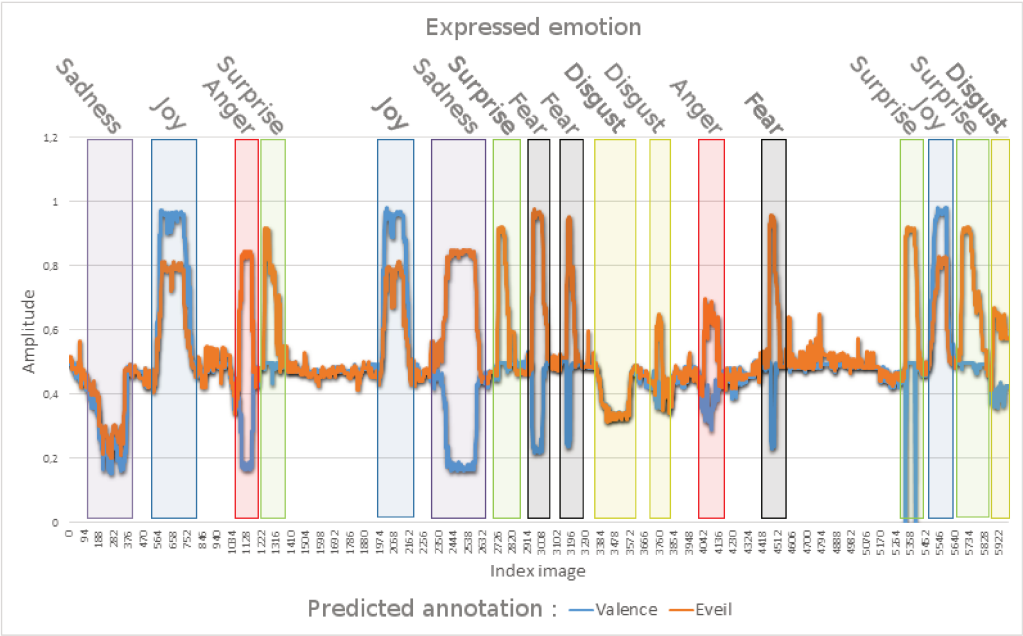


FIGURE 5.12 – Annotation automatique d'une séquence vidéo regroupant trois expressions de la colère

1.3 Discussion

Cette première phase de validation avait pour objectif de comparer l'interpréation automatique des expressions faciales face à une annotation humaine. Nous avons scindé ce travail de manière à paramétrer notre algorithme afin d'obtenir dans un second temps une interprétation fiable de la part de notre algorithme. Le paramétrage de notre système a révélé une grande amélioration pour un valeur de K égale à six. Nous pouvons déduire de ce comportement, que l'utilisation d'un algorithme de régression basé sur KNN est particulièrement adapté dans notre cas. Cet algorithme bien que considéré comme peu performant dans le domaine de la classification nous permet d'interpréter les expressions faciales de manière précise mais aussi en temps réel. A partir de notre vidéo de test, nous avons pu obtenir une vitesse d'éxécution variant de 8 à 12 images par secondes. Ces performances obtenues sur un ordinateur personnel peuvent s'apparenter à du temps réel car la durée d'une émotion varie de quelques secondes, voir minutes. En observant au minimum 8 images par secondes, nous pouvons suivre l'évolution de l'expérience émotionnelle et donc exploiter la structure de nos cartes contenant les phases de changement émotionnel.

A partir de ce constat, nous avons effectué un ensemble de tests sur des séquences vidéos contenant des émotions similaires dans un premier temps puis variées. D'une manière générale, chaque expression émotionnelle contenue dans les vidéos a été identifiée. Cependant, deux interprétations sur 17 sont erronnées. Notre système possède certaines difficultés à différencier le dégoût de la peur ou de la colère. En comparant les images de ces différentes émotions, nous pouvons voir une grande similarité entre le dégoût et la colère. Ceci peut expliquer l'erreur commise par notre système. Deux solutions sont envisagées pour améliorer ce comportement. Nous pensons d'une part, à continuer l'apprentissage de notre système en ajoutant des expressions faciales correspondant au dégoût et d'autre part, il est possible que les caractéristiques faciales observées ne soient pas suffisantes. En ajoutant de nouveaux points caractéristiques ou modélités, nous pensons pouvoir faire progresser la détection sur ce point.

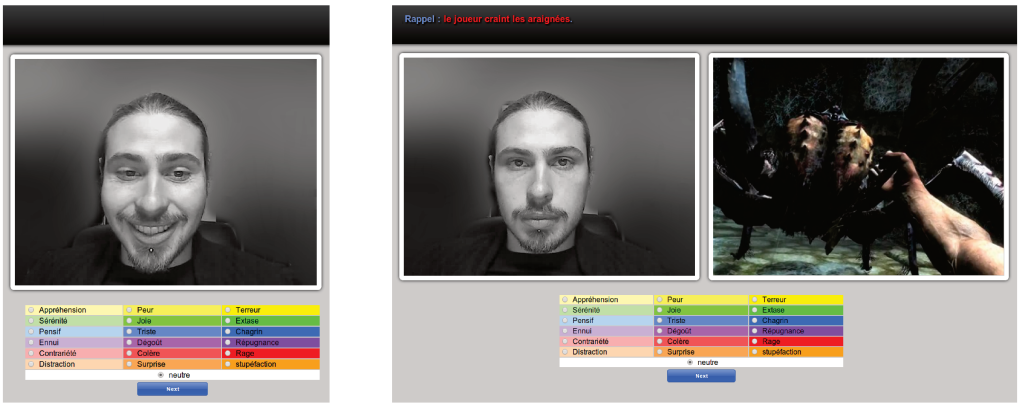
2 Contextualisation des annotations

Afin de quantifier la pertinence des résultats de notre système, nous avons mis en place un questionnaire permettant de comparer d'une part l'annotation en sortie de l'ontologie par rapport à une annotation humaine et d'autre part, d'observer l'influence du contexte dans l'analyse émotionnelle. Nous proposons donc aux personnes qui ont participé au questionnaire de prendre la place de notre système dans deux situations différentes.

La première phase permet de comparer l'annotation sans prise en compte du contexte dans lequel l'émotion apparait. Nous présentons donc à l'utilisateur un ensemble d'images représentant un acteur mimant des émotions. La tâche demandée est de sélectionner parmi les six émotions de Ekman divisées en trois intensités chacunes, celle qui correspond le mieux à l'expression proposée. En plus des images présentant l'acteur, nous insérons les images utilisées par Ekman pour représenter les émotions. L'utilisation de ces images nous permet d'évaluer la capacité de la personne répondant au questionnaire, à identifier les expressions émotionnelles.

La seconde phase permet quant à elle de comparer l'annotation en tenant compte du contexte émotionnel. Pour ce faire, nous présentons le même ensemble d'expressions émotionnelles complété par deux informations, à savoir un trait de caractère de l'acteur et une image correspondant au stimulus déclencheur de l'expression. L'utilisateur doit alors remplir la même tâche que précédemment en indiquant l'émotion qu'il considère la plus pertinente. Les traits de caractères proposés sont décrits sous la forme suivante : "la personne aime les araignées" ou "la personne craint les araignées". Le sujet de l'attirance ou de la répulsion est un élément parmi ours, araignée et chien. Nous pouvons proposer au participant un des 6 traits de caractères. Ce que nous voulons observer ici c'est de quelle manière les utilisateurs vont pondérer leur annotation lorsqu'ils auront connaissance de ces deux informations.

Nous avons donc développé un site internet permettant de récolter les résultats de ce questionnaire. L'accès à ce dernier reste limité afin de contrôler la population observée pour les résultats que nous présentons. Les captures d'écran 5.13a et 5.13b montrent les deux principales phases du questionnaire où l'on retrouve les images présentées, ainsi qu'un tableau permettant de sélectionner les émotions exprimées.



(a) Sans contexte

(b) Avec contexte

FIGURE 5.13 – Présentation des deux phases du questionnaire

2.1 Hypothèse initiale

A partir des théories émotionnelles sélectionnées, nous avons identifié que la notion de contexte est une information importante dans l'expérience émotionnel. Pour Schachter et Singer [Schachter 79], le contexte correspond à la connaissance d'une personne sur une certaine situation. Dans l'expérience qu'ils ont mené sur l'injection d'épinéphrine, ils ont pu démontrer que la connaissance des conséquences d'une telle injection permettait de limiter celles-ci sur le ressenti émotionnel. D'autre part, le monde informatique sous-exploite cette information. Pour d'autres théoriciens comme Frédéric Toates ou P. Philippot, le contexte est associé également à la mémoire et plus généralement à la cognition. Dans notre cas, nous avons proposé une modélisation du contexte sous la forme de philies ou phobies. Nous souhaitons ici montrer quelques pistes en correlation avec nos hypothèses initiales.

La principale hypothèse est que le contexte de l'expérience émotionnelle impacte le ressenti émotionnel. C'est-à-dire qu'un utilisateur arachnophobe en présence d'une araignée ressentira une émotion plus intense qu'un utilisateur non arachnophobe pour une même expression. Dans notre réalisation cela se traduit par une différence de l'annotation émotionnelle en fonction des phobies/philies et du stimulus déclencheur. Nous souhaitons donc observer si les participants à notre expérimentation vont modifier l'annotation d'une même image après avoir pris connaissance des traits de caractère et du stimulus présenté.

La seconde hypothèse est qu'un trait de caractère n'est pas généralisable pour une population donnée. C'est la notion même de personnalisation que nous souhaitons évaluer au travers de ce questionnaire. En effet, nous partons du principe que les traits de caractères sont définis par l'utilisateur de notre système. Cette tâche fastidieuse et complexe à réaliser doit être paramétrée par la personne connaissant ses réactions en fonction de certains stimuli. Dans notre questionnaire, cela correspond à la concordance des annotations avec le contexte des participants. Nous souhaitons observer si un consensus existe dans la définition d'un trait de caractère comme l'arachnophobie par exemple. Pour cela, nous calculerons le pourcentage le plus important d'accord entre les participants au questionnaire.

2.2 Résultat

Déroulement de l'expérimentation

Nous avons demandé à un groupe de 10 personnes constitué de 4 femmes et 6 hommes âgés de 22 à 36 ans ayant déjà utilisé des jeux vidéos, de remplir notre questionnaire. L'ensemble des participants a effectué la même phase initiale de notre test.

Cependant, pour la seconde étape, nous avons divisé le groupe afin de modifier les informations contextuelles. Le premier groupe a eu pour information que la personne représentée sur les photos était en train de regarder des images d'araignées et qu'il en avait peur. Pour le second groupe, nous avons modifié cette information en présentant des images d'ours et en indiquant que la personne aimait les ours. Nous avons présenté 38 expressions faciales dont 6 extraites des images de Ekman pour la première partie du questionnaire. Dans la seconde partie, nous avons réutilisé les 32 expressions faciales précédentes auxquelles nous avons associé soit des images d'araignées, soit des images d'ours, issues de jeux vidéo.

Evaluation des performances de reconnaissance des émotions des participants

Nous avons évalué les performances de reconnaissance des émotions de chacun de nos participants. En utilisant les réponses fournies lors de la présentation des six images de Ekman, nous avons pu établir un niveau de performance, analysé d'une part sans prise en compte de l'intensité émotionnelle, puis avec prise en compte de l'intensité émotionnelle. La figure 5.14 présente les résultats obtenus ainsi qu'une moyenne pour l'ensemble des deux groupes.

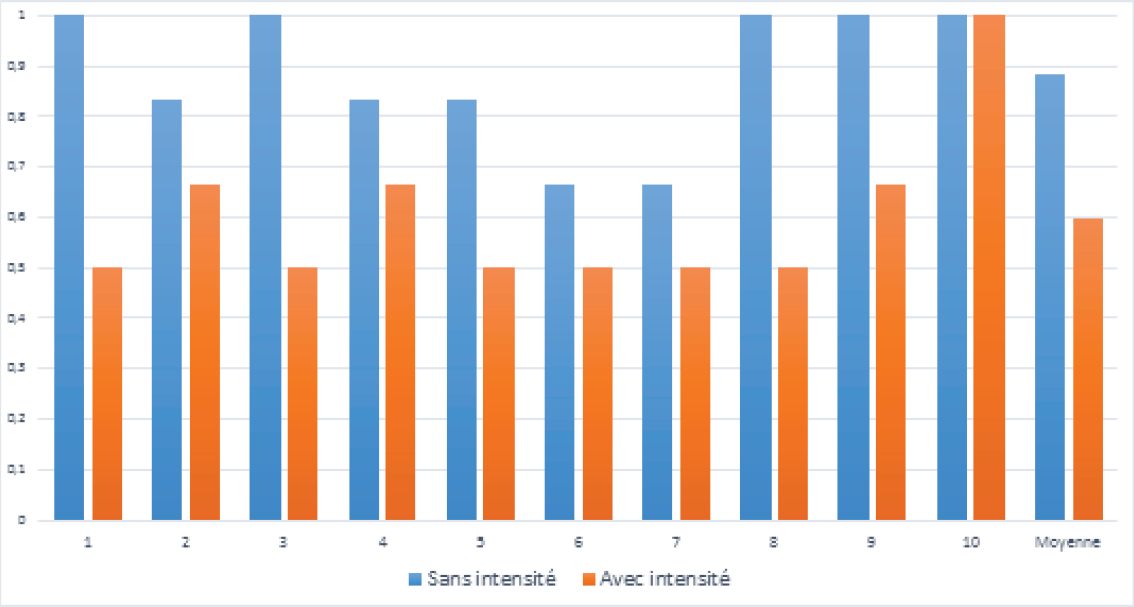


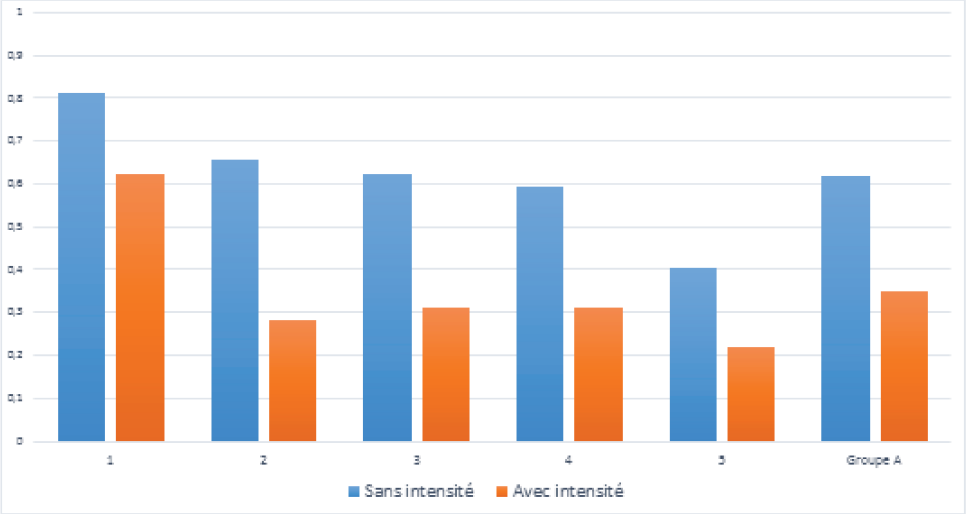
FIGURE 5.14 – Capacité de reconnaissance des émotions de nos participants

L'ensemble des participants a obtenu un score supérieur à 66% de bonnes réponses si l'on ne considère pas l'intensité émotionnelle (par exemple rage et colère seront considérées comme identiques). Si l'on intègre cette intensité les scores

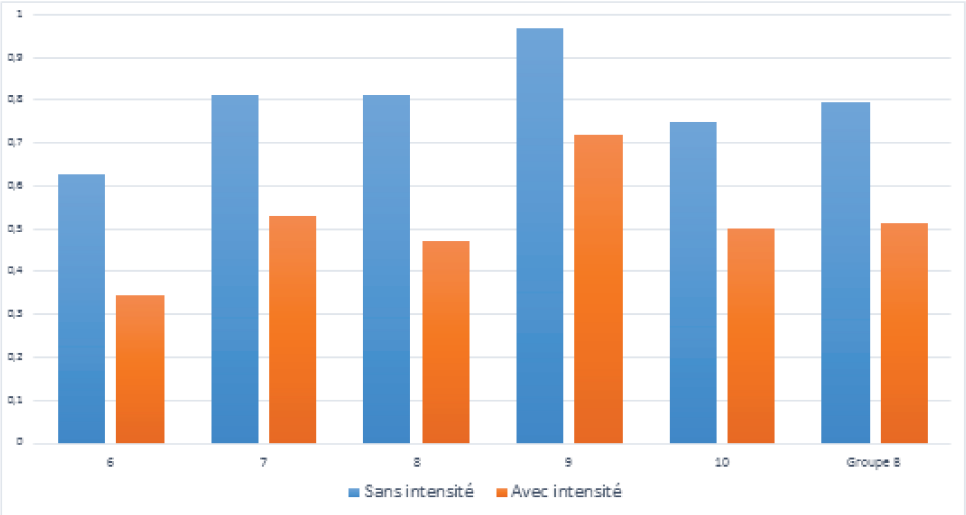
tombent à 50% au minimum. Compte tenu du faible nombre d'images émotionnelles utilisées, un score de 66% est considéré comme suffisamment bon pour ne pas éliminer de participant de notre étude.

Comparaison des réponses en fonction du contexte émotionnel

Ici, nous avons voulu voir si pour une expression faciale, l'émotion reconnue par nos participants variait en fonction de la présence ou non du contexte. Nous avons pour chaque utilisateur comparé les annotations proposées sans connaissance du contexte face aux annotations proposées avec connaissance du contexte. Les figures 5.15 et 5.16 représente les résultats obtenus.



(a) Groupe A (phobie)



(b) Groupe B (philie)

FIGURE 5.15 – Différence d'annotation en fonction du contexte pour les deux groupes

D'une manière générale, nous pouvons observer sur ces premiers graphiques [Fig. 5.15] que les participants ont modifié leurs réponses une fois le contexte émotionnel connue. Sans considération de l'intensité émotionnelle, l'impact du contexte reste faible environ 30% des annotations ont changées. En prenant en compte cette intensité, cette différence augmente pour atteindre 55%. Cette information, tend à montrer que les utilisateurs identifient bien le type d'émotion mais utilisent l'intensité de celle-ci pour adapter le ressenti en fonction du contexte.

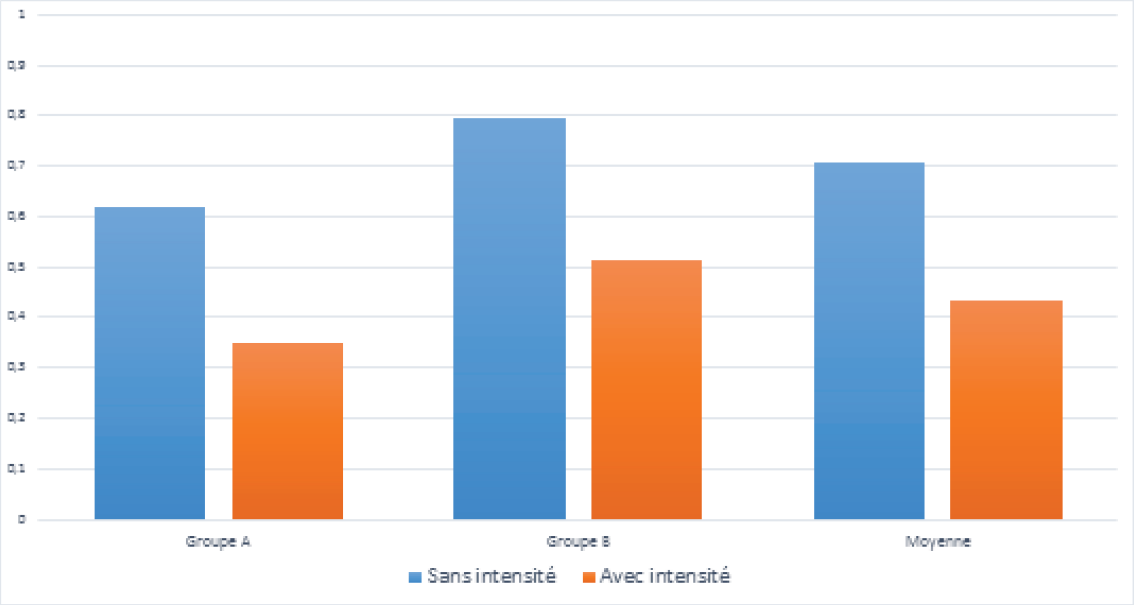


FIGURE 5.16 – Résultats globaux de l'impact du contexte émotionnel sur l'identification des émotions

Sur ce second graphique [Fig. 5.16], Nous pouvons observer que l'impact du contexte émotionnel est plus faible pour le groupe B. L'utilisation d'un trait de caractère positif (philie) engendre moins d'impact sur la reconnaissance qu'un trait de caractère négatif (phobie). Cependant, en moyenne la différence d'annotation entre les deux phases de notre questionnaire reste conséquente puisque nous obtenons une différence d'annotation supérieure à 55% en considérant l'intensité émotionnelle.

Détermination du seuil d'accord entre les participants

Le dernier résultat extrait de ce questionnaire correspond au niveau de concordance des réponses entre les participants pour les situations avec contexte. Nous souhaitons par cet intermédiaire identifier si l'ensemble des participants arrive à un accord sur l'émotion ressentie pour une situation donnée. Pour cela, dans chacun des deux groupes et pour chaque situation présentée, nous avons comptabilisé le nombre de

personnes ayant fourni la même émotion. Nous avons donc calculé le pourcentage d'accord sur chacune des images puis un pourcentage d'accord moyen sur les 32 situations présentées. La figure 5.17 présente les résultats obtenus.

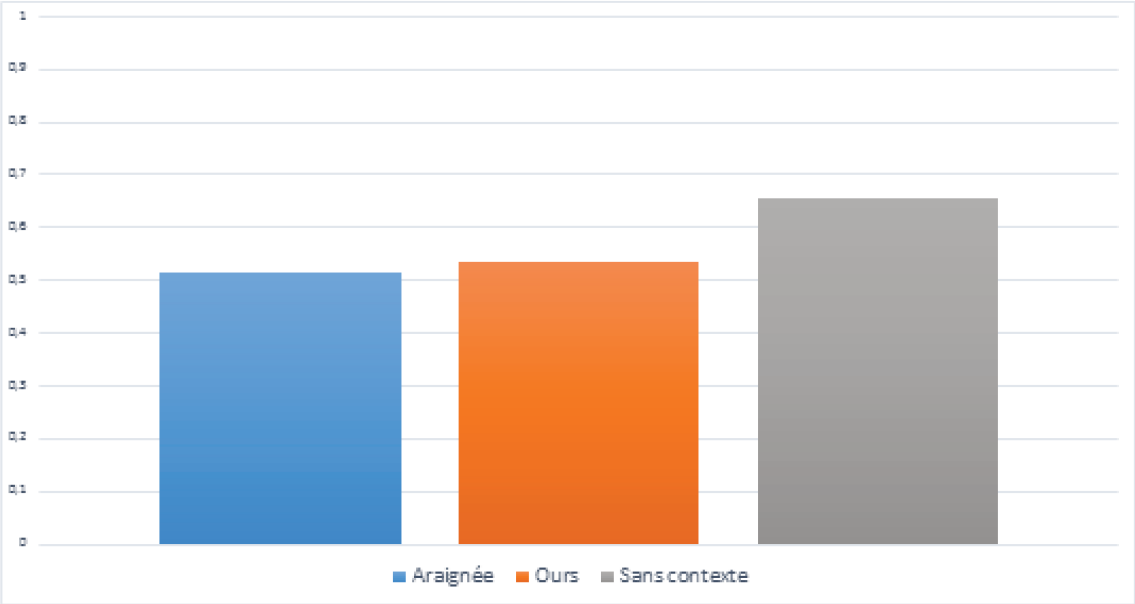


FIGURE 5.17 – Pourcentage d'accord de nos participants en fonction du contexte

Nous pouvons observer sur ces résultats que les deux groupes obtiennent un niveau d'accord légèrement supérieur à 50% en moyenne. Ce qui signifie que pour une situation donnée, seulement la moitié des participants arrive à un accord.

2.3 Discussion

Tout d'abord, ce que nous pouvons déduire du premier résultat de ce questionnaire, est que la reconnaissance des émotions même pour un être humain n'est pas certaine. Cependant, ce que nous souhaitons étudier grâce à ce questionnaire c'est l'importance du contexte émotionnel dans le cadre de l'interprétation de l'émotion. A l'échelle de notre expérimentation, nous pouvons dire qu'il existe une différence d'interprétation dès que l'on fournit aux participants une information sur le contexte émotionnel. De plus, cette différence reste modérée si l'on ne considère pas l'intensité émotionnelle. L'impact d'un phobie ou philie lorsque qu'une personne y est confrontée ne permet pas de modifier complètement l'émotion ressentie mais seulement de pondérer l'intensité de cette dernière. En moyenne, l'émotion proposée pour une expression faciale reste identique dans plus de 70% des cas. En revanche, l'intensité de l'émotion est modifiée dans plus de 55% des cas. Ce comportement observé chez nos participants nous permet de mettre en évidence l'intérêt de pondérer l'émotion en

fonction du contexte après l'interprétation des expressions faciales. De plus, le type de trait de caractères décrit modifie l'interprétation des émotions. Nous avons pu observer un impact plus important sur la reconnaissance des émotions de la part des phobies que des philies. Néanmoins, le calcul de la pondération n'est pas un élément évident. Le dernier résultat obtenu tend à montrer qu'il est difficile de définir un impact générique d'une phobie ou philie sur le ressenti de l'émotion. Nos participants n'ont pas réussi à se mettre d'accord sur l'émotion ressentie par notre sujet en connaissant le contexte. Seulement la moitié des participants arrive à un accord en considérant le contexte alors qu'ils sont plus de 65% à être en accord sans considérer le contexte. Ceci permet de dire que la définition de l'impact d'une philie ou phobie sur le ressenti émotionnel est un élément de personnalisation de notre système.

3 Observation de l'hystérésis

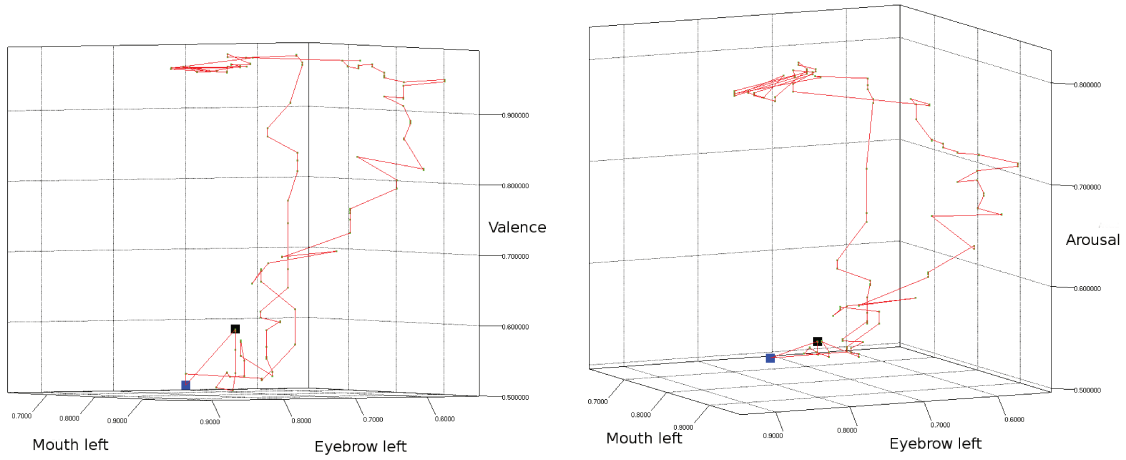
L'utilisation du modèle de Klaus Scherer comme base de notre système de classification des émotions est un choix risqué. En effet, à notre connaissance, aucune implémentation ni expérimentation n'a été réalisée sur cette approche. Un de nos objectifs était de mettre en évidence les hypothèses théoriques utilisées dans des conditions réelles. Ici, nous ne proposons pas une validation de ce modèle mais juste une observation montrant l'existence d'une caractéristique théorique dans des conditions expérimentales. Le phénomène que nous souhaitons mettre en évidence est l'hystérésis [Fig. 1]. Cette discontinuité est facilement identifiable par sa forme dans des espaces à trois dimensions. Au sein de nos cartes, l'utilisation d'un grand nombre de capteurs rend difficile l'observation de l'hystérésis même en effectuant des projections sur des espaces 3D. De plus, nos cartes ne comprenant pas de surfaces mais un nuage de points, il n'est pas certain d'identifier sa présence, même lorsqu'il existe une même expression de deux émotions différentes.

3.1 Observation

Le moyen d'observation que nous privilégions se concentre sur l'observation des conséquences d'une hystérésis [Fig. 1]. En particulier, lors de l'expérience émotionnelle, le chemin emprunté par l'expression de celle-ci est différent selon que l'on atteigne l'émotion ou que l'on revienne au calme. C'est cette différence que nous souhaitons observer au sein de notre système afin de mettre en évidence le phénomène d'hystérésis.

Pour ce faire, nous avons extrait de notre vidéo de test une séquence émotionnelle sur laquelle nous avons appliqué notre algorithme de classification. Pendant son

exécution, nous avons enregistré les états intermédiaires détectés par le système pour une séquence émotionnelle partant du calme pour atteindre la joie et revenir à son état initial. La figure 5.18, montre la représentation graphique de ces enregistrements avec en bleu le point de départ de l'émotion, en noir l'état final et en rouge le parcours effectué sur nos cartes.



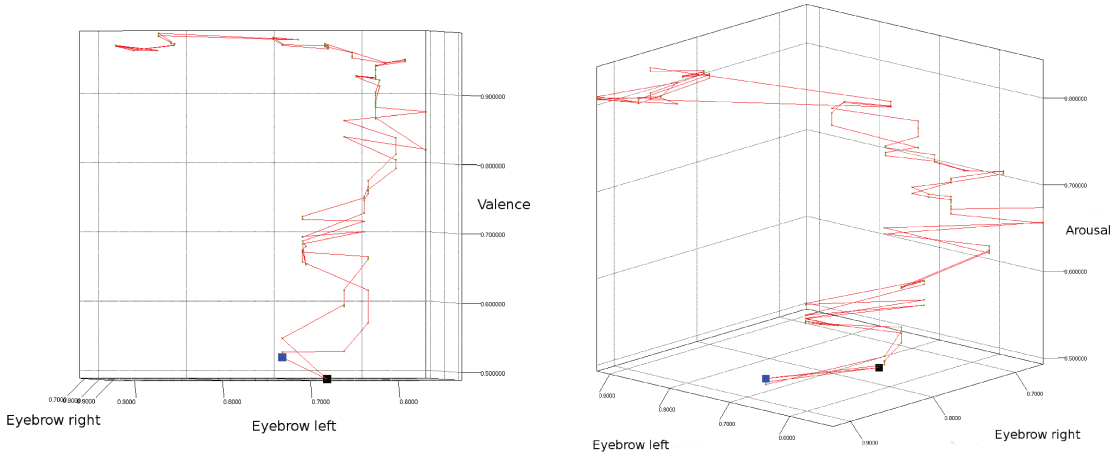
(a) Présence d'une hystérésis sur l'axe de valence (b) Présence d'une hystérésis sur l'axe d'éveil

FIGURE 5.18 – Observation d'une hystérésis lors de la détection de la joie sur les deux axes du modèle de Russell

La principale information que nous pouvons tirer de ces graphiques est la présence de deux chemins distincts ; le premier correspondant à l'apparition de la joie et le second à la disparition. Cette information met en évidence la différence d'expressions de l'émotion.

Afin de montrer que ce double chemin n'est pas le résultat de notre algorithme conservatif, nous avons pris la même séquence que nous avons scindée au moment de l'émotion la plus intense. Nous l'avons dupliquée et inversée afin d'obtenir une séquence symétrique comportant la même expression lors de l'apparition de l'émotion que lors de la disparition. Les résultats obtenus montrent la présence de deux chemins légèrement distincts.

Su la figure 5.19, nous pouvons observer la grande similarité entre l'apparition de l'émotion et sa disparition. La différence entre les tracés effectués sur une séquence normale par rapport à une séquence symétrique met en évidence la différence d'expressions existant dans l'expression émotionnelle ainsi que le faible impact de notre algorithme conservateur dans l'apparition de ce phénomène.



(a) Absence d'hystérésis sur l'axe de valence (b) Absence d'hystérésis sur l'axe d'éveil

FIGURE 5.19 – Observation du chemin de détection sur une vidéo symétrique exprimant la joie

3.2 Discussion

Bien que l'utilisation d'un algorithme conservateur incite à penser qu'il favorise le phénomère d'hystérésis, nous avons pu montrer que dans le cas d'une vidéo symétrique les deux chemins sont très similaires, ce qui n'est pas le cas pour une expression normale d'une émotion. Ce phénomène met en évidence l'existence de séquences d'expressions différentes pour des changements d'états émotionnels.

La présence d'une hystérésis dans la détection d'un changement émotionnel tend à apporter une réalité pratique aux fondements théoriques du modèle de Klaus R. Scherer. Les résultats obtenus précédemment sur la détection des émotions ainsi que l'observation d'une hystérésis nous incite à penser que l'utilisation des CEPs peut proposer une analyse pertinente basée sur les systèmes complexes. En effet, la possibilité de comparer les cartes ou d'observer des phénomènes de type hystérésis peut fournir des informations sur la généralisation des expressions émotionnelles ou sur l'établissement de certaines règles permettant de mieux comprendre l'expression émotionnelle.

4 Synthetisation des expressions émotionnelles

Dans cette partie, nous souhaitons mettre en avant les capacités de génération d'expressions émotionnelles à partir de nos CEPs. L'algorithme présenté dans le cha-

pitre 3 permet de retrouver les expressions émotionnelles à partir d'une annotation brute.

4.1 Méthodologie

Afin de quantifier les performances de notre système, nous avons mis en place une routine de comparaison des expressions mimées par un acteur face aux expressions fournies par notre avatar [Fig. 5.20]. La méthodologie présentée permet d'utiliser les annotations en sortie de notre système pour les réutiliser dans la synthétisation des expressions.

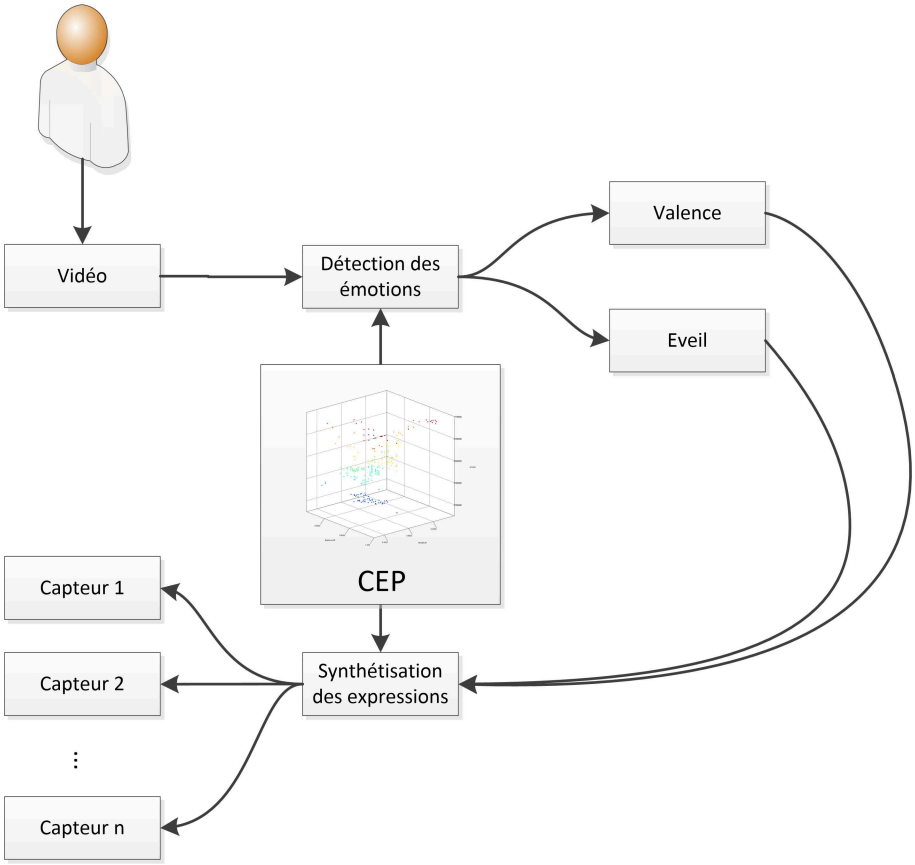
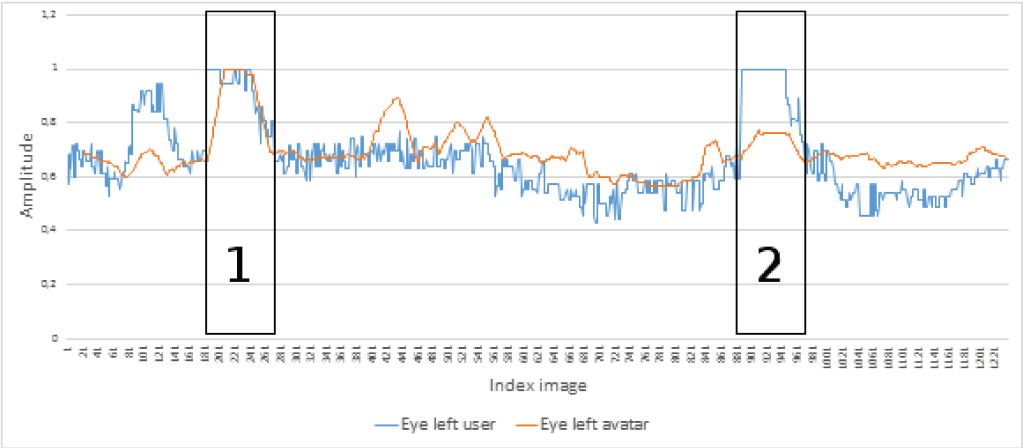


FIGURE 5.20 – Méthodologie d'évaluation des performances de l'algorithme de synthétisation des expressions faciales

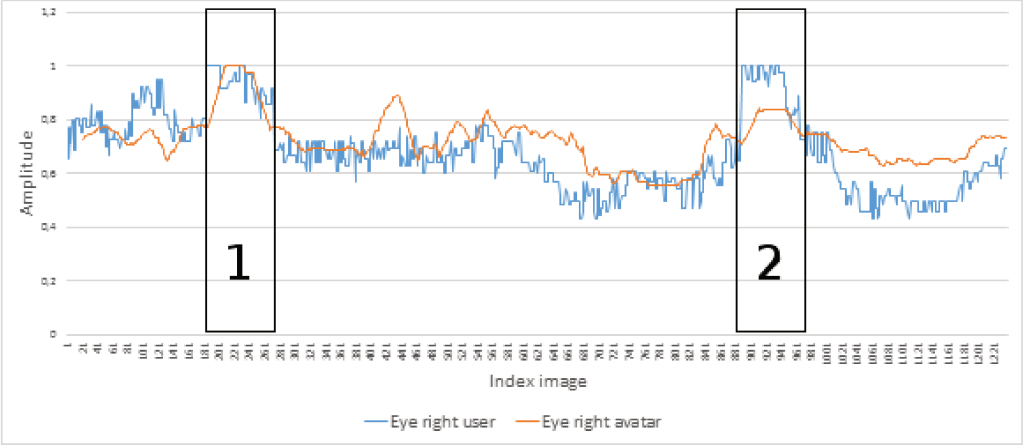
Afin d'extraire les informations utilisées pour la comparaison, nous proposons de réutiliser notre vidéo d'apprentissage sur laquelle nous appliquons notre algorithme de détection des émotions. En sortie de nos CEPs, nous capturons l'annotation brute afin de la réinjecter dans notre algorithme de synthétisation. Enfin, nous enregistrons à la fois les données capturées sur le sujet et celles calculées par notre générateur d'expressions.

4.2 Résultats

Les figures suivantes montrent un tracé comparant capteur par capteur les deux types de données avec en orange les expressions de l'avatar et en bleu celles de l'utilisateur.



(a) Sourcil gauche



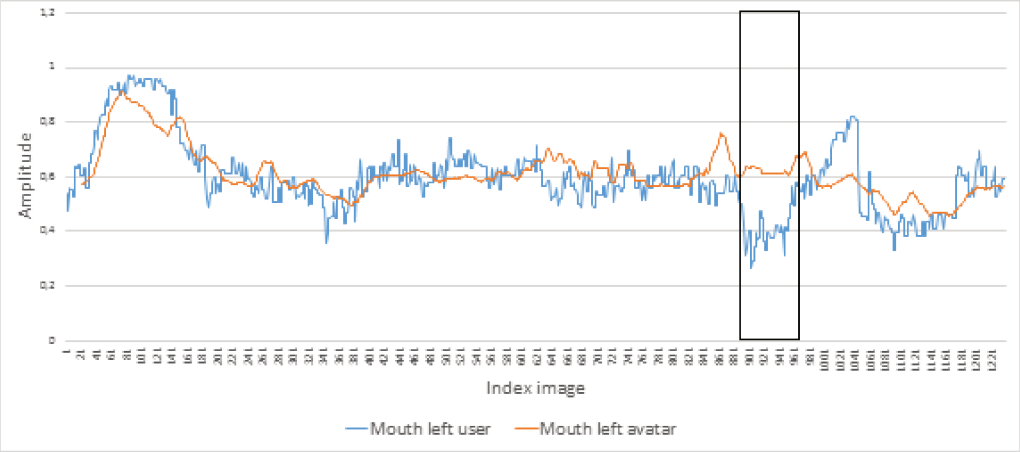
(b) Sourcil droit

FIGURE 5.21 – Tracés des deux capteurs de sourcils

Les deux premiers tracés présentés par la figure 5.21 représentent à gauche le capteur de sourcil gauche et à droite le sourcil droit. L'objectif ici est d'observer la similitude des deux tracés afin que l'avatar exprime l'émotion la plus proche de celle de l'utilisateur. La première constation que nous pouvons faire c'est que les tendances des deux tracés sont semblables. Cependant, il existe de nombreuses différences montrant une confusion dans le type d'émotion à exprimer. Le premier encadré des deux tracés représente une expression de la peur. Pour cette émotion, l'avatar propose bien une élévation des sourcils comme l'exprime l'utilisateur. En revanche, sur le second encadré, l'utilisateur exprime la surprise avec une forte élévation des

sourcils. Ici, l'avatar propose une expression où cette élévation est très atténuée par rapport à l'expression d'origine.

Nous observons ce comportement sur la bouche également où les tracés bien que plus semblable, comportent des variations atténuées ou contradictoires avec l'expression souhaitée. Les parties encadrées montrent une nouvelle fois une difficulté de notre système à identifier la bonne expression pour la surprise.

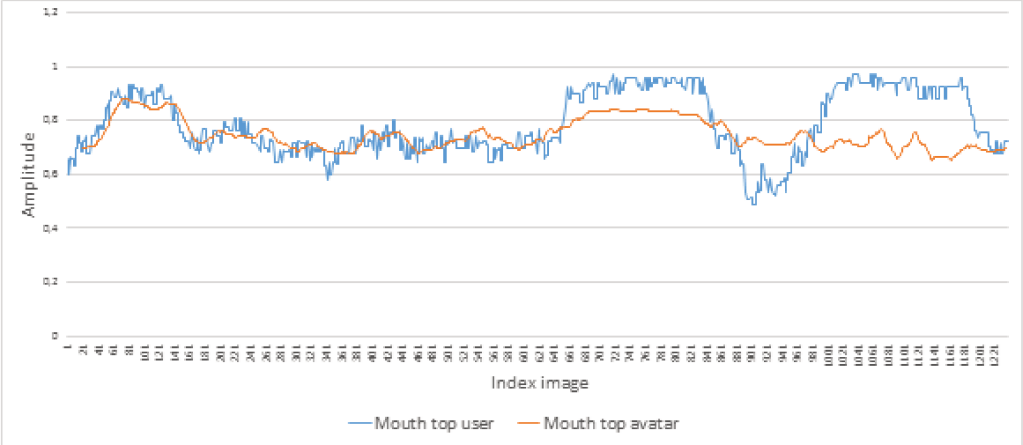


(a) Commissure gauche

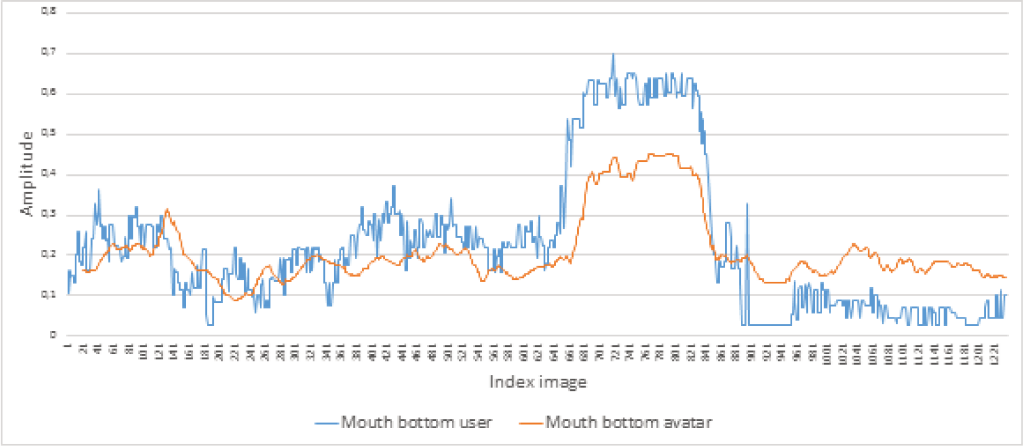
(b) Commissure droite

FIGURE 5.22 – Tracés des deux capteurs de la bouche gauche et droite

Enfin pour les bords de la bouche, les tendances sont proches et il n'existe que très peu de confusion dans la synthétisation. Néanmoins, l'atténuation reste très forte et ne permet pas d'identifier certaines expressions qui pourront alors être interprétées comme une séquence neutre.



(a) Bord haut



(b) Bord bas

FIGURE 5.23 – Tracés des deux capteurs de la bouche haut et bas

4.3 Discussion

Bien que les tendances de ces différents tracés montrent des simulitudes importantes, le ressenti face à l'avatar n'est pas suffisamment expressif à cause des confusions entre émotions, mais également à cause de l'atténuation des expressions. Nous pouvons expliquer ce comportement par deux aspects :

- l'utilisation du modèle de Russell a l'avantage d'être simple et accepté par la communauté. Néanmoins, certains défauts sont connus en particulier sur la discrimination des émotions comme la peur et la colère ou encore entre le dégoût et la tristesse. Les confusions que nous observons dans le cadre de cette exprience sont directement liées à cette caractéristique du modèle. Dans les exemples présentés précédemment, nous avons pris la séquence exprimant la

surprise. L'annotation de cette émotion sur la dimension d'éveil est très semblable à la peur ou à la colère. Nous observons donc une difficulté à identifier la bonne expression. L'utilisation d'un modèle plus expressif comme le modèle Pleasure, Arousal, Dominance (PAD) [Mehrabian 74] afin de mieux discriminer les émotions.

- la stratégie de sélection des expressions que nous utilisons est basée sur une moyenne des expressions retournées par nos cartes. L'atténuation observée dans nos résultats est causée par ce choix, mais c'est aussi la cause de certaines confusions lorsque deux émotions possèdent une même annotation, cela concerne des expressions antagonistes comme la peur et la colère, où les sourcils sont dans un premier cas, levés, et dans un second, baissés.

Les choix effectués ne permettent pas d'obtenir des résultats pouvant égaler les outils spécialisés dans ce traitement. Cependant, avec de faibles moyens et des solutions naïves, nous arrivons à obtenir des tendances proches de ce qu'exprime notre utilisateur. Ces résultats nous encourage à continuer l'exploration dans cette voie afin non pas de concurrencer ce qui existe déjà, mais d'apporter une nouvelle information permettant de personnaliser l'expression des avatars aux expressions de l'utilisateur.

5 Conclusion

Dans ce chapitre, nous avons présenté l'ensemble des résultats permettant d'évaluer les performances de notre réalisation. Nous avons testé notre système sur plusieurs points en proposant à chaque fois une quantification de ces résultats ou une observation de certains phénomènes significatifs. Nous avons commencé par présenter la méthodologie mise en oeuvre pour l'évaluation du coeur de notre système qui est la détection des émotions, ainsi que la définition de nos jeux de tests. La première étape de cette étude a été de montrer la pertinence de l'utilisation de l'algorithme KNN en définissant la meilleure valeur de la constante K à partir de nos vidéos d'apprentissage. La détermination de cette constante se base sur l'analyse de ce que nous avons défini, comme la déviation, et permet d'atteindre un taux de déviation par rapport à une annotation manuelle de l'ordre de 3% pour la meilleure valeur de K.

Nous avons poursuivi par deux expérimentations sur des séquences vidéos inconnues, afin d'observer le comportement du système sur des séquences n'ayant pas été utilisées pour l'apprentissage de nos cartes. La première a permis de mettre en

évidence la capacité du système à identifier à chaque fois l'apparition des émotions sur notre jeu de test. Cependant, sur les 17 séquences analysées, nous avons obtenu des erreurs d'annotations sur 3 d'entre elles, dont 2 concernant le dégoût, qui est interprété comme de la colère ou de la peur. Enfin, nous avons analysé la pertinence de la catégorisation et de la contextualisation de ces annotations au travers d'un questionnaire. Nous avons pu déduire de cette étude que le contexte émotionnel et plus particulièrement les phobies et philies, modifie l'interprétation des expressions faciales de manière importante. Cela justifie la pondération que nous effectuons au travers de l'ontologie émotionnelle. De plus, nous avons pu déterminer que les participants n'ont pas fourni une interprétation générique en fonction du contexte. Nous en concluons que la description de l'impact des phobies et philies sur le ressenti émotionnel est une personnalisation du système qui doit rester à la charge de l'utilisateur final.

Nous avons également présenté les observations que nous avons pu effectuer sur la présence d'une hystérésis pendant la détection des émotions. L'observation de ce phénomène permet d'apporter une indication quant à la validité du modèle proposé par Klaus Scherer.

Nous avons terminé par l'évaluation de la synthétisation des expressions émotionnelles. La comparaison entre les expressions mimées et générées montre une réelle similitude, bien que nos CEPs ne soient pas conçues pour cela. Cependant, un nombre trop important de divergences nous empêche de rivaliser avec les avatars émotionnels déjà proposés par la communauté.

Conclusion

Solutions et contributions de nos travaux

Dans ce document, nous avons présenté nos travaux sur la détection des émotions. En introduction, nous présentions trois problématiques sur lesquelles nous souhaitions proposer une solution. De plus, nous avons également présenté des solutions ou observations dépassant le cadre strict de ce questionnement, ceci en raison de l'utilisation d'un modèle jamais implémenté à notre connaissance ou par la structuration de notre détecteur permettant de développer de nouvelles fonctionnalités à moindre coût. Cependant, ce que nous avons présenté correspond à une étape préliminaire dans la réalisation d'un détecteur émotionnel. A partir des travaux existants, nous voulions proposer d'intégrer le contexte dans l'identification des émotions, mais aussi d'explorer de nouvelles voies en basant notre modélisation sur les systèmes complexes. Pour conclure la présentation de nos travaux, nous reviendrons dans un premier temps sur les apports d'une telle modélisation, ainsi que sur les performances de cette dernière, calculées à partir de notre prototype. Nous terminerons enfin par les améliorations et perspectives visées à partir de ce travail.

Comment concevoir un système capable d'unifer les modèles de description des émotions ?
La solution que nous proposons pour cette problématique est apportée par notre ontologie des émotions. Cette dernière correspond au module de catégorisation et de contextualisation dans notre architecture globale. L'approche continue que nous avons employée lors de la première phase d'interprétation (CEP) est alors catégorisée par notre ontologie. Cette catégorisation correspond à une traduction de l'annotation dimensionnelle en un terme du modèle catégoriel décrit dans l'ontologie. L'utilisation d'un tel outil nous permet une grande flexibilité et évolutivité dans la description à la fois des modèles dimensionnels et catégoriels. De plus, les standards employés issus du W3C favorisent l'intéropérabilité entre les bases de connaissances.

Comment concevoir un système capable de s'abstraire des capteurs utilisés ?

Tout d'abord, nous n'avons pas pu rendre compatible notre modèle avec l'ensemble des capteurs au sens large utilisable dans le cadre de la détection des émotions. Nous avons donc restraint cette compatibilité à un sous ensemble de capteurs répondant à certaines contraintes. Afin de déterminer les limitations que nous imposons à l'utilisateur de notre modèle, nous avons identifié les points communs couramment utilisés pour une détection en temps réel des émotions. Nous avons donc défini les capteurs compatibles comme l'ensemble des capteurs fournissant des données numériques et continues. Ceci permet d'englober la plupart des capteurs physiologiques comme le rythme cardiaque ou les reflexes galvanométriques. Cependant, dans notre prototype, nous avons pu montrer que certains capteurs a priori non compatibles avec notre description pouvait être modifiés afin d'être utilisés dans notre système. C'est le cas en particulier de la caméra qui est utilisée afin de suivre les déplacements de certains points caractéristiques du visage.

Néanmoins, le traitement initial effectué sur les capteurs afin d'améliorer la qualité du signal ou de le rendre compatible avec notre système est séparé de notre phase d'interprétation des signaux. Cette séparation permet d'abstraire notre modèle des considérations liées aux différents capteurs et permet d'uniformiser les données obtenues. Une fois le traitement effectué, nous utilisons nos CEPs afin d'effectuer l'interprétation de ces signaux. Ces cartes basées sur les systèmes complexes et le modèle dynamique des émotions de Scherer nous permettent de manipuler un nombre théoriquement illimité de capteurs afin de prédire l'émotion ressentie. Nous avons pour cela développé un algorithme de regression permettant à partir de nos CEPs de retrouver l'émotion correspondante à l'expression.

C'est grâce à la séparation des considérations avec le traitement individuel de chacun des capteurs puis l'utilisation de notre modèle capable de manipuler plusieurs capteurs que nous avons pu développer un détecteur capable de fonctionner avec des capteurs variables.

Comment utiliser les données contextuelles dans la détection des émotions ?

La principale difficulté rencontrée correspond à la nature même de cette information. Les informations d'expression changent rapidement et, dans le cadre de nos travaux, sont numériques et continues. En revanche, le contexte émotionnel ne peut être représenté par une donnée de ce type. Il est nécessaire dans ce cas de raisonner sur les liens sémantiques entre le contexte émotionnel (stimulus déclencheur) et les connaissances pérennes du sujet vis à vis de ce stimulus. Pour notre réalisation, nous nous sommes concentrés sur une partie du contexte émotionnel correspondant

aux philies et phobies d'un individu. Nous avons pu montrer que l'utilisation d'une ontologie permet de modéliser ces connaissances, mais aussi, grâce aux capacités d'inférences et de raisonnement de cette dernière, de retrouver les liens sémantiques entre une description d'une philie ou phobie et le stimulus présenté au sujet. En nous basant sur les travaux du web-sémantique, la conception d'une ontologie des émotions nous permet d'identifier ces liens mais aussi de décrire l'impact de ces traits de caractères sur la détection de l'émotion. En nous appuyant sur des bases de connaissances existantes comme dbpedia, nous avons pu réaliser un module pour notre système capable de pondérer la prédiction d'émotions en fonction d'un stimulus quelconque. L'aspect générique et interopérable de ce module permet d'une part de prendre en compte le contexte émotionnel dans la détection, mais aussi d'être étendu à d'autres aspects du contexte comme l'humeur ou les traits de caractères.

Vision globale de nos travaux

De manière plus générale, nous avons réutilisé une approche standard dans le domaine, à savoir la classification des expressions corporelles. Nous l'avons cependant adaptée à notre vision continue des émotions, puis étendue pour prendre en compte le contexte. Tout d'abord, nous avons conçu notre système à partir du fonctionnement humain en nous basant sur la théorie des deux facteurs. Cela nous a permis d'identifier les deux niveaux d'interprétation que nous avons créés pour notre système. De plus, afin de faciliter l'observation de plusieurs modalités, nous avons proposé de séparer les problématiques de traitement des capteurs dans un module à part afin d'uniformiser les données en entrée de notre système.

A partir de cette architecture, nous avons souhaité exploiter de nouvelles approches, en particuliers, pour l'interprétation des expressions corporelles. Notre hypothèse initiale était que nous pouvons ressentir une infinité d'émotions mais que notre vocabulaire nous limite dans l'expression de cette variété. C'est pourquoi nous avons voulu représenter cette vision en utilisant une approche continue pour décrire nos émotions avant de les catégoriser. Nous avons donc inscrit notre modèle dans une logique de régression afin de prédire une annotation continue des émotions. Nous avons également voulu construire notre base d'apprentissage, nécessaire à l'éxecution de notre algorithme de régression, sur des bases théoriques capables de nous fournir des outils d'analyses. Nous nous sommes tournés vers le modèle dynamique des émotions de Scherer basé sur les sytèmes complexes et les travaux de Thom. Pour ce dernier, n'ayant à notre connaissance jamais eu d'implémentation, nous avons voulu observer de manière expérimentale le fonctionnement d'un tel modèle. Une implémentation directe n'étant pas possible, de par le cadre applicatif que nous nous sommes fixés, nous en avons proposé une version adaptée à notre système sous le nom de CEP. Ces modifications se concentrent en partie sur les sources de données

utilisées et sur la description continue des émotions. La réalisation d'un prototype basé sur ce modèle a permis de mettre en évidence des capacités de détections importantes mais aussi d'observer le phénomène d'hystérésis utilisé par Scherer pour présenter son modèle.

La seconde partie de notre travail s'est basée sur la manière d'exploiter des connaissances sur un individu avec le contexte émotionnel. Nous avons proposé de pondérer notre interprétation des expressions corporelles en fonction des philies ou phobies de la personne qui les exprime. Nous avons présenté pour cela l'ontologie émotionnelle que nous avons réalisée permettant de pondérer l'expression corporelle mais aussi de catégoriser cette information afin de la rendre compréhensible par un humain. Nous avons alors réalisé une expérimentation afin de mettre en évidence l'importance du contexte dans la détermination de l'émotion, mais aussi de montrer que ce type de trait de caractères n'est pas généralisable.

Perspectives

Les travaux présentés dans cette thèse représentent un travail préliminaire permettant de fixer les bases d'un système évolutif et adaptable aux évolutions du domaine. Nous avons pu réaliser un détecteur en temps réel capable de manipuler des sources de données divergeantes afin d'englober un maximum d'informations dans notre système. Maintenant, nous souhaitons développer notre modèle et analyser les performances de ce dernier dans un cadre plus large et plus précis que ce nous avons pu faire jusqu'à présent. Pour cela, nous avons identifié trois voies de recherche que nous voulons explorer afin de valoriser ce travail.

Notre vision du contexte émotionnel reste limitée. Cependant, notre ontologie a permis de mettre en évidence l'intérêt d'un tel outil dans la modélisation des phobies et philies. Nous souhaiterions maintenant étendre notre champ d'observation. Nous avons pu voir dans les définitions des émotions qu'il existe des ambiguïtés en particulier avec les traits de caractères. Nous pensons que cette notion peut faire partie d'un contexte plus général. Tout comme les phobies et philies, les traits de caractères (big five []) sont des informations durables voire potentiellement immuables. De plus, le lien entre émotion et personnalité est déjà identifié [Mardaga 09]. Nous pensons donc qu'intégrer la personnalité au sein de notre ontologie permettrait d'affiner davantage la détection et ainsi correspondre encore plus finement avec le ressentie réel des utilisateurs.

Au vu des travaux effectués, nous avons pu montrer une première implémentation de notre modèle CEP. Les résultats obtenus grâce à notre prototype montrent qu'un tel système est capable d'identifier les émotions à partir de l'observation d'une

modalité. En nous basant dans le domaine des détecteurs multimodaux, nous pensons que l'augmentation du nombre de modalités observées permet d'améliorer la qualité de la détection. C'est pourquoi nous aimerions développer notre prototype en ajoutant de nouveaux capteurs en particulier d'ECG et d'EEG afin de voir l'impact de cette modification sur les performances globales de notre système. Cependant, une telle réalisation nécessite une expérimentation d'envergure. Comme nous le présentions, mimer des émotions au travers du visage est une chose facile, en revanche demander à un utilisateur d'exprimer une émotion réelle est beaucoup plus difficile. Nous avons néanmoins plusieurs protocoles et outils à notre disposition afin de mener à bien une étude plus large sur le sujet.

Enfin, nous avons basé la conception de nos cartes sur des théories mathoématiques capables de nous fournir des outils d'analyse mathématiques performant. En particulier, les travaux de Thom sur lesquels repose notre modèle, permettent d'étudier les discontinuités. Nous avons pu mettre en évidence la présence d'une hystérésis pendant notre phase de validation. Cependant, la présence de tels phénomènes est vue selon Scherer comme la modélisation d'une réalité de l'expérience émotionnelle. Une cartographie de ces discontinuités, ainsi qu'une comparaison des cartes de plusieurs utilisateurs ou d'un même utilisateur dans le temps, est une voie que nous souhaitons prendre. En effet, si nous pouvons identifier des similitudes entre deux utilisateurs, pouvons nous également définir une fonction de transition entre ces derniers ? Peut-on également transformer nos cartes personnalisées en cartes génériques paramétrables ? En fonction des traits de caractères ? De l'humeur ?

Cette dernière perspective laisse de nombreuses questions en suspens. De plus, la réponse à ces questions permettrait grandement, de simplifier la détection des émotions et d'envisager une utilisation à plus grande échelle.

Bibliographie

[Anttonen 05] Jenni Anttonen & Veikko Surakka. *Emotions and Heart Rate while Sitting on a Chair.* 2005.

[Arnold 60] Magda B Arnold. *Emotion and personality. Vol. I. Psychological aspects.* 1960.

[Banse 96] Rainer Banse & Klaus R Scherer. *Acoustic profiles in vocal emotion expression.* Journal of personality and social psychology, vol. 70, no. 3, page 614, 1996.

[Belzung 07] Catherine Belzung. Biologie des émotions. De Boeck, 2007.

[Berners-Lee 01] Tim Berners-Lee, James Hendler, Ora Lassila*et al. The semantic web.* Scientific american, vol. 284, no. 5, pages 28–37, 2001.

[Berners-Lee 07] Tim Berners-Lee. *Giant global graph.* online posting, Networks," Proceedings of the National Academy of Sciences of the United States of America, vol. 98, pages 404–9, 2007.

[Beyer 99] Kevin Beyer, Jonathan Goldstein, Raghu Ramakrishnan & Uri Shaft. *When is "nearest neighbor" meaningful ?* In Database Theory—ICDT'99, pages 217–235. Springer, 1999.

[Boiten 94] Frans A Boiten, Nico H Frijda & Cornelis JE Wientjes. *Emotions and respiratory patterns : review and critical analysis.* International Journal of Psychophysiology, vol. 17, no. 2, pages 103–128, 1994.

[Bos 06] Danny Oude Bos. *EEG-based Emotion Recognition.* 2006.

[Bradley 94] Margaret M Bradley & Peter J Lang. *Measuring emotion : the self-assessment manikin and the semantic differential.* Journal of behavior therapy and experimental psychiatry, vol. 25, no. 1, pages 49–59, 1994.

[Bradley 99] Margaret M Bradley & Peter J Lang. *Affective norms for English words (ANEW) : Instruction manual and affective ratings.* Rapport technique, Technical Report C-1, The Center for Research in Psychophysiology, University of Florida, 1999.

[Bradley 07] Margaret M Bradley & Peter J Lang. *The International Affective Digitized Sounds (; IADS-2) : Affective ratings of sounds and instruction manual.* University of Florida, Gainesville, FL, Tech. Rep. B-3, 2007.

[Cannon 27] W. B. Cannon. *The James-Lange theory of emotions : A critical examination and alternative theory.* American Journal of Psychology, vol. 39, pages 106–124, 1927.

[Chanel 05] Guillaume Chanel, Julien Kronegg, Didier Grandjean & Thierry Pun. *Emotion Assessment : Arousal Evaluation Using EEG's and Peripheral Physiological Signals.* 2005.

[Chaouachi 12] Maher Chaouachi & Claude Frasson. *Mental workload, engagement and emotions : an exploratory study for intelligent tutoring systems.* In Intelligent Tutoring Systems, pages 65–71. Springer, 2012.

[Coan 07] James A Coan & John JB Allen. Handbook of emotion elicitation and assessment. Oxford university press, 2007.

[Cowie 05] Roddy Cowie & Marc Schröder. Piecing together the emotion jigsaw, volume 3361 of *Lecture Notes in Computer Science*, pages 305–317. Springer Verlag, 2005. Book Title : Machine Learning for Multimodal Interaction, First International Workshop, MLMI 2004. Martigny, Switzerland, June 21-23, 2004, Revised Selected Papers.

[Dahmane 11] Mohamed Dahmane & Jean Meunier. *Emotion recognition using dynamic grid-based hog features.* In Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on, pages 884–888. IEEE, 2011.

[Darwin 02] C. Darwin, P. Ekman & P. Prodger. The expression of the emotions in man and animals. Oxford University Press, 2002.

[Ekman 71] Paul Ekman. *Universals and cultural differences in facial expressions of emotion.* In Nebraska symposium on motivation. University of Nebraska Press, 1971.

[Ekman 78] Paul Ekman, Wallace V Friesen & Joseph C Hager. *Facial action coding system (FACS).* A technique for the measurement of facial action. Consulting, Palo Alto, 1978.

[Ekman 92] Paul Ekman. *Are there basic emotions.* Psychological Review, 1992.

[Ekman 94] P. Ekman & R.J. Davidson. The nature of emotion : Fundamental questions. Series In Affective Science. Oxford University Press, 1994.

[Ekman 99] Paul Ekman. Basic emotions. Sussex, UK : John Wiley & Sons, Ltd., 1999.

[Fontaine 07] Johnny RJ Fontaine, Klaus R Scherer, Etienne B Roesch & Phoebe C Ellsworth. *The world of emotions is not two-dimensional.* Psychological science, vol. 18, no. 12, pages 1050–1057, 2007.

[Friesen 83] Wallace V Friesen & Paul Ekman. *Emfacs-7 : emotional facial action coding system.* Unpublished manuscript, University of California at San Francisco, vol. 2, 1983.

[Frijda 87] N.H. Frijda. The emotions. Studies in Emotion and Social Interaction. Cambridge University Press, 1987.

[Goleman 99] Daniel Goleman. Emotional intelligence. PBS Home Video, 1999.

[Grassi 09] Marco Grassi. *Developing HEO human emotions ontology.* In Biometric ID Management and Multimodal Communication, pages 244–251. Springer, 2009.

[Grenon 04] Pierre Grenon, Barry Smith & Louis Goldberg. *Biodynamic Ontology : Applying BFO in the Biomedical Domain.* In Stud. Health Technol. Inform, pages 20–38. IOS Press, 2004.

[Gross 95] James J. Gross & Robert W. Levenson. *Emotion Elicitation Using Films.* 1995.

[Gruber 95] Thomas R Gruber*et al. Toward principles for the design of ontologies used for knowledge sharing.* International journal of human computer studies, vol. 43, no. 5, pages 907–928, 1995.

[Gunes 07] Hatice Gunes & Massimo Piccardi. *Bi-modal emotion recognition from expressive face and body gestures.* Journal of Network and Computer Applications, vol. 30, no. 4, pages 1334–1345, 2007.

[Haag 04] Andreas Haag, Silke Goronzy, Peter Schaich & Jason Williams. *Emotion Recognition Using Bio-Sensors : First Steps Towards an Automatic System.* 2004.

[Hamm 11] Jihun Hamm, Christian G Kohler, Ruben C Gur & Ragini Verma. *Automated facial action coding system for dynamic analysis of facial expressions in neuropsychiatric disorders.* Journal of neuroscience methods, vol. 200, no. 2, pages 237–256, 2011.

[Hastings 11] Janna Hastings, Werner Ceusters, Barry Smith & Kevin Mulligan. *Dispositions and processes in the Emotion Ontology.* 2011.

[Healey 98] Jennifer Healey & Rosalind Picard. *Digital processing of affective signals.* In Acoustics, Speech and Signal Processing, 1998. Proceedings of the 1998 IEEE International Conference on, volume 6, pages 3749–3752. IEEE, 1998.

[Healey 02] Jennifer Healey & Rosalind W. Picard. *Eight-emotion Sentics Data.* http ://affect.media.mit.edu, 2002.

[Heraz 09] Alicia Heraz & Claude Frasson. *Predicting Learner Answers Correctness through Brainwaves Assesment and Emotional Dimensions.* In AIED, pages 49–56, 2009.

[Homma 08] Ikuo Homma & Yuri Masaoka. *Breathing rhythms and emotions.* Experimental physiology, vol. 93, no. 9, pages 1011–1021, 2008.

[Hoque 12] M Hoque, Daniel McDuff & R Picard. *Exploring temporal patterns in classifying frustrated and delighted smiles.* 2012.

[Izard 77] Carroll Izard. *Human Emotions.* 1977.

[James 84] William James. *What Is an Emotion ?* Mind, 1884.

[Jones 08] Christian Jones & Jamie Sutherland. *Acoustic emotion recognition for affective computer gaming.* In Affect and emotion in human-computer interaction, pages 209–219. Springer, 2008.

[Jraidi 12] Imène Jraidi, Pierre Chalfoun & Claude Frasson. *Implicit strategies for intelligent tutoring systems.* In Intelligent Tutoring Systems, pages 1–10. Springer, 2012.

[Kahneman 66] Daniel Kahneman & Jackson Beatty. *Pupil Diameter and Load on Memory.* Science, vol. 154, no. 3756, pages 1583–1585, 1966.

[Kanade 00] Takeo Kanade, Jeffrey F Cohn & Yingli Tian. *Comprehensive database for facial expression analysis*. In Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on, pages 46–53. IEEE, 2000.

[Kassam 13] Karim S Kassam, Amanda R Markey, Vladimir L Cherkassky, George Loewenstein & Marcel Adam Just. *Identifying Emotions on the Basis of Neural Activation*. PLOS ONE, vol. 8, no. 6, page e66032, 2013.

[Kemper 81] Theodore David Kemper. *Social constructionist and positivist approaches to the sociology of emotions*. 1981.

[Kleinginna 81] P R Kleinginna & A M Kleinginna. *A Categorized List of Emotion Definitions with Suggestions for a Consensual Definition*. Motivation and Emotion, 1981.

[Lang 05] Peter J Lang, Margaret M Bradley & Bruce N Cuthbert. *International Affective Picture System (IAPS) : Instruction Manual and Affective Ratings*. 2005.

[Lazarus 91] R.S. Lazarus. Emotion & adaptation. Oxford Univesity Press, 1991.

[Li 09] Mu Li & Bao-Liang Lu. *Emotion classification based on gamma-band EEG*. In Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE, pages 1223–1226. IEEE, 2009.

[López 08] Juan López, Rosa Gil, Roberto García, Idoia Cearreta & Nestor Garay. *Towards an ontology for describing emotions*. Emerging Technologies and Information Systems for the Knowledge Society, pages 96–104, 2008.

[Mardaga 09] S Mardaga & Michel Hansenne. *Personality modulation of P300 wave recorded within an emotional oddball protocol*. Neurophysiologie Clinique/Clinical Neurophysiology, vol. 39, no. 1, pages 41–48, 2009.

[Marr 82] David Marr. Vision : A computational investigation into the human representation and processing of visual information. W. H. Freeman and Company, NY, 1982.

[Masinter 05] Larry Masinter, Tim Berners-Lee & Roy T Fielding. *Uniform resource identifier (URI) : Generic syntax*. 2005.

[Mathieu 05] Yvette Yannick Mathieu. *Annotation of emotions and feelings in texts.* In Affective Computing and Intelligent Interaction, pages 350–357. Springer, 2005.

[Mehrabian 74] Albert Mehrabian & James A Russell. An approach to environmental psychology. the MIT Press, 1974.

[Metri 12] Priya Metri, Jayshree Ghorpade & Ayesha Butalia. *Facial Emotion Recognition Using Context Based Multimodal Approach.* International Journal of Emerging Sciences, vol. 2, no. 1, page 171, 2012.

[Mistry 09] Pranav Mistry & Pattie Maes. *SixthSense : a wearable gestural interface.* page 11, 2009.

[Nicholson 00] Joy Nicholson, Kazuhiko Takahashi & Ryohei Nakatsu. *Emotion recognition in speech using neural networks.* Neural Computing & Applications, vol. 9, no. 4, pages 290–296, 2000.

[Obrenovic 05] Zeljko Obrenovic, Nestor Garay, Juan Miguel López, Inmaculada Fajardo & Idoia Cearreta. *An ontology for description of emotional cues.* In Affective Computing and Intelligent Interaction, pages 505–512. Springer, 2005.

[Palomba 97] Daniela Palomba, Alessandro Angrilli & Alessio Mini. *Visual evoked potentials, heart rate responses and memory to emotional pictorial stimuli.* International Journal of Psychophysiology, vol. 27, no. 1, pages 55–67, 1997.

[Panksepp 82] Jaak Panksepp. *Toward a general psychobiological theory of emotions.* Behavioral and Brain Sciences, vol. 5, pages 407–422, 1982.

[Partala 03] Timo Partala & Veikko Surakka. *Pupil size variation as an indication of affective processing.* International Journal of Human-Computer Studies, vol. 59, no. 1, pages 185–198, 2003.

[Philippot 01] Pierre Philippot & Alexandre Schaefer. *Emotion and memory.* 2001.

[Picard 95] R W Picard. *Affective Computing.* Pattern Recognition, 1995.

[Picard 01] Rosalind W. Picard, Elias Vyzas & Jennifer Healey. *Toward machine emotional intelligence : Analysis of affective physiological state.* Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 23, no. 10, pages 1175–1191, 2001.

[Plutchik 01] Robert Plutchik. *The Nature of Emotions.* American Scientist, 2001.

[Schachter 62] Stanley Schachter & Jerome E. Singer. *Cognitive, social, and physiological determinants of emotional state.* Psychological Review, vol. 69, no. 5, pages 379–399, 1962.

[Schachter 79] S Schachter & J Singer. *Cognitive, social and physiological determinants of emotional state.* Psychological Review, 1979.

[Scherer 00] Klaus R. Scherer. *Emotions as episodes of subsystem synchronization driven by nonlinear appraisal processes.* 2000.

[Scherer 05] Klaus R Scherer. *What are emotions ? And how can they be measured ?* Social science information, vol. 44, no. 4, pages 695–729, 2005.

[Scherer 09] Klaus R Scherer. *Emotions are emergent processes : they require a dynamic computational architecture.* Philosophical Transactions of the Royal Society B : Biological Sciences, vol. 364, no. 1535, pages 3459–3474, 2009.

[Sénéchal 13] Thibaud Sénéchal, Jay Turcot & Rana El Kaliouby. *Smile or smirk ? automatic detection of spontaneous asymmetric smiles to understand viewer experience.* In Automatic Face and Gesture Recognition, 2013.

[Skinner 38] B. F. Skinner. The behavior of organisms : An experimental analysis. Century psychology series. D. Appleton-Century Company, incorporated, 1938.

[Smith 04] Barry Smith & Pierre Grenon. *The Cornucopia of Formal-Ontological Relations.* Dialectica, vol. 58, no. 3, pages 279–296, 2004.

[Susi 07] Tarja Susi, Mikael Johannesson & Per Backlund. *Serious games : An overview.* 2007.

[Sutherland 64] Ivan E Sutherland. *Sketch pad a man-machine graphical communication system.* pages 6–329, 1964.

[Takahashi 04] Kazuhiko Takahashi. *Remarks on emotion recognition from multimodal bio-potential signals.* In Industrial Technology, 2004. IEEE ICIT'04. 2004 IEEE International Conference on, volume 3, pages 1138–1143. IEEE, 2004.

[Thom 72] René Thom. Stabilité structurelle et morphogénèse. WA Benjamin Reading, Mass, 1972.

[Tomkins 80] Silvan Tomkins. *Affect as amplification : Some modifications in theory.* Emotion : Theory, research, and experience, 1980.

[Wessberg 00] Johan Wessberg, Christopher R. Stambaugh, Jerald D. Kralik, Pamela D. Beck, Mark Laubach, John K. Chapin, Jung Kim, S. James Biggs, Mandayam A. Srinivasan & Miguel A. L. Nicolelis. *Real-time prediction of hand trajectory by ensembles of cortical neurons in primates.* Nature, vol. 408, no. 6810, pages 361–365, 2000.

Annexe A

Algorithme de détection des émotions

```
1
2  fonction  classification (carte, donnees, emotionPrecedente)
3
4    annotation = 0
5    confiance = 0
6
7    // Etape 1 : parcours des sequences de la carte
8    pour i = 0 jusqu'a i = taille(carte)
9
10     sequence = carte[i]
11     distances[K]
12     annotations[K]
13
14     // Etape 2 : Recherche des K voisins les plus proches des nouvelles donnees
15     // completees par l'annotation emotionnelle precedente
16     pour j = 0 jusqu'a j = taille(sequence)
17       etat = sequence[j];
18       distance = calculerDistance(etat, donnees, emotionPrecedente)
19       pour k = 0 jusqu'a k < K
20         si distance < distances[k] alors
21           ajouter(distance, distances)
22           ajouter(etat, annotations)
23         fin si
24       fin pour
25     fin pour
```

```
26
27      // Etape 3 : Recherche de la confiance la plus faible pour le calcul du
28      // barycentre
29      confianceCourante = calculerConfianceMoyenne(distances)
30      si confianceCourante > confiance alors
31        confiance = confianceCourante
32        annotation = calculerBarycentre(annotations, distances)
33      fin si
34    fin pour
35
36    retourner confiance, annotation
37  fin fonction
```

Annexe B

Ontologie des émotions

```
1  <?xml version="1.0"?>
2
3  <!DOCTYPE rdf:RDF [
4      <!ENTITY xsd "http://www.w3.org/2001/XMLSchema#" >
5      <!ENTITY rdfs "http://www.w3.org/2000/01/rdf-schema#" >
6      <!ENTITY rdf "http://www.w3.org/1999/02/22-rdf-syntax-ns#" >
7      <!ENTITY emotion "http://www.i3s.unice.fr/~berthelo/ontology/2013/01/28/emotion.rdfs#
8  ]>
9
10 <rdf:RDF xmlns="http://www.i3s.unice.fr/~berthelo/ontology/2013/01/28/emotion.rdfs#"
11  xml:base="http://www.i3s.unice.fr/~berthelo/ontology/2013/01/28/emotion.rdfs"
12  xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
13  xmlns:emotion="http://www.i3s.unice.fr/~berthelo/ontology/2013/01/28/emotion.rdfs#"
14  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
15  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
16
17 <!--
18 ###########
19 # Classes #
20 ###########
21 -->
22 <rdfs:Class rdf:about="&emotion;Coordinate" />
23
24 <rdfs:Class rdf:about="&emotion;Radius" />
25   <rdfs:subClassOf rdf:resource="&emotion;Coordinate" />
26 </rdfs:Class>
27
```

```
28 <rdfs:Class rdf:about="&emotion;Component" />
29   <rdfs:subClassOf rdf:resource="&emotion;Coordinate" />
30 </rdfs:Class>
31
32 <rdfs:Class rdf:about="&emotion;Arousal">
33   <rdfs:subClassOf rdf:resource="&emotion;Component" />
34 </rdfs:Class>
35 <rdfs:Class rdf:about="&emotion;Valence">
36   <rdfs:subClassOf rdf:resource="&emotion;Component" />
37 </rdfs:Class>
38
39 <rdfs:Class rdf:about="&emotion;Emotion" />
40
41 <rdfs:Class rdf:about="&emotion;Joy">
42   <rdfs:subClassOf rdf:resource="&emotion;Emotion" />
43 </rdfs:Class>
44 <rdfs:Class rdf:about="&emotion;Fear">
45   <rdfs:subClassOf rdf:resource="&emotion;Emotion" />
46 </rdfs:Class>
47 <rdfs:Class rdf:about="&emotion;Sadness">
48   <rdfs:subClassOf rdf:resource="&emotion;Emotion" />
49 </rdfs:Class>
50 <rdfs:Class rdf:about="&emotion;Anger">
51   <rdfs:subClassOf rdf:resource="&emotion;Emotion" />
52 </rdfs:Class>
53 <rdfs:Class rdf:about="&emotion;Disgust">
54   <rdfs:subClassOf rdf:resource="&emotion;Emotion" />
55 </rdfs:Class>
56 <rdfs:Class rdf:about="&emotion;Surprise">
57   <rdfs:subClassOf rdf:resource="&emotion;Emotion" />
58 </rdfs:Class>
59
60 <rdfs:Class rdf:about="&emotion;Impact" />
61 <rdfs:Class rdf:about="&emotion;Person" />
62 <rdfs:Class rdf:about="&emotion;Stimulus" />
63 <rdfs:Class rdf:about="&emotion;Trait" />
64
65 <!--
66 ##############
67 # Properties #
```

```
68  ##############
69  -->
70
71  <rdf:Property rdf:about="&emotion;hasEffect">
72    <rdfs:domain rdf:resource="&emotion;Impact" />
73    <rdfs:range rdf:resource="&xsd;double" />
74  </rdf:Property>
75
76  <rdf:Property rdf:about="&emotion;hasImpact">
77    <rdfs:domain rdf:resource="&emotion;Trait" />
78    <rdfs:range rdf:resource="&emotion;Impact" />
79  </rdf:Property>
80
81  <rdf:Property rdf:about="&emotion;hasMaximum">
82    <rdfs:domain rdf:resource="&emotion;Coordinate" />
83    <rdfs:range rdf:resource="&xsd;double" />
84  </rdf:Property>
85
86  <rdf:Property rdf:about="&emotion;hasMinimum">
87    <rdfs:domain rdf:resource="&emotion;Coordinate" />
88    <rdfs:range rdf:resource="&xsd;double" />
89  </rdf:Property>
90
91  <rdf:Property rdf:about="&emotion;hasTrait">
92    <rdfs:domain rdf:resource="&emotion;Person" />
93    <rdfs:range rdf:resource="&emotion;Trait" />
94  </rdf:Property>
95
96  <rdf:Property rdf:about="&emotion;involves">
97    <rdfs:domain rdf:resource="&emotion;Stimulus" />
98  </rdf:Property>
99
100 <rdf:Property rdf:about="&emotion;isDefinedBy">
101   <rdfs:domain rdf:resource="&emotion;Emotion" />
102   <rdfs:range rdf:resource="&emotion;Coordinate" />
103 </rdf:Property>
104
105 <rdf:Property rdf:about="&emotion;pertainsTo">
106   <rdfs:domain rdf:resource="&emotion;Impact" />
107   <rdfs:range rdf:resource="&emotion;Component" />
```

```
108 </rdf:Property>
109
110 <rdf:Property rdf:about="&emotion;relatesTo">
111   <rdfs:domain rdf:resource="&emotion;Trait" />
112 </rdf:Property>
113
114 </rdf:RDF>
```

Annexe C

Requête SPARQL de catégorisation et de contextualisation des annotations brutes

```
SELECT ?emotion
WHERE {
  ############################
  # Retrieve valence range #
  ############################
  ?emotion emotion:isDefinedBy ?arousal
  ?arousal rdf:type emotion:Arousal
  ?arousal emotion:hasMinimum ?minArousal
  ?arousal emotion:hasMaximum ?maxArousal

  ############################
  # Retrieve arousal range #
  ############################
  ?emotion emotion:isDefinedBy ?valence
  ?valence rdf:type emotion:Valence
  ?valence emotion:hasMinimum ?minValence
  ?valence emotion:hasMaximum ?maxValence

  {
    ##############################
    # Retrieve emotion in range #
```

```
22      ##############################
23      select  ((xsd:double(?minArousal)  xsd :double(if (bound( ?effectArousal), ?effectArousal,
0))) AS ?minA)((xsd :double( ?maxArousal)  xsd:double(if  (bound(?effectArousal),
24                ?effectArousal ,  0)))  AS ?maxA)
25              ((xsd:double(?minValence)  xsd :double(if (bound( ?effectValence), ?effectValence,
0))) AS ?minV)((xsd :double( ?maxValence)  xsd:double( if (bound(?effectValence),
26                ?effectValence ,  0)))  AS ?maxV)
27      where {
28        FILTER (?minA < eveilBrute && ?maxA > eveilBrute &&
29                 ?minV < valenceBrute && ?maxV > valenceBrute)
30          {
31            ##############################################
32            # Case of person with a trait  in  relation  with stimulus #
33            ##############################################
34            select  ?effectArousal  ?effectValence
35            where {
36              \textbf{ utilisateur } emotion:hasTrait ?trait
37              ?trait  emotion:relatesTo ?extTrait
38              \textbf{stimulus} emotion:involves ?extStimulus
39              { ?extStimulus rdf:type ?extTrait  }
40              UNION
41              { ?extStimulus rdfs:subClassOf ?extTrait }
42              ?trait  emotion:hasImpact ?impactArousal
43              ?impactArousal emotion:pertainsTo emotion:Arousal
44              ?impactArousal emotion:hasEffect ?effectArousal
45
46              ?trait  emotion:hasImpact ?impactValence
47              ?impactValence emotion:pertainsTo emotion:Valence
48              ?impactValence emotion:hasEffect ?effectValence
49            }
50          }
51          UNION
52          {
53            ##############################################
54            # Case of person without trait or without relation  with stimulus #
55            ##############################################
56            select  ("0"^^xsd:double AS ?effectArousal)
57                    ("0"^^xsd:double AS ?effectValence)
58            where {
59              FILTER NOT EXISTS {
```

```
60                  \textbf{ utilisateur } emotion:hasTrait ?trait
61                  ?trait  emotion:relatesTo ?extTrait
62                  \textbf{stimulus} emotion:involves ?extStimulus
63                  { ?extStimulus rdf:type ?extTrait }
64                  UNION
65                  { ?extStimulus rdfs:subClassOf ?extTrait }
66                }
67              }
68            }
69        }
70    }
71  }
```

Annexe D

Les six émotions de Ekman

FIGURE D.1 – Emotions correspondant aux six émotions d'Ekman