

Modélisation, détection et annotation des états émotionnels à l'aide d'un espace vectoriel multidimensionnel

Imen Tayari-Meftah

► To cite this version:

Imen Tayari-Meftah. Modélisation, détection et annotation des états émotionnels à l'aide d'un espace vectoriel multidimensionnel. Autre [cs.OH]. Université Nice Sophia Antipolis, 2013. Français. <NNT : 2013NICE4017>. <tel-00908233v2>

HAL Id: tel-00908233

<https://tel.archives-ouvertes.fr/tel-00908233v2>

Submitted on 26 Jun 2013

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
ÉCOLE DOCTORALE STIC
SCIENCES ET TECHNOLOGIES DE L'INFORMATION
ET DE LA COMMUNICATION

THÈSE

pour obtenir le titre de

Docteur en Sciences

de l'Université de Nice - Sophia Antipolis

Mention : INFORMATIQUE

Présentée et soutenue par

Imen TAYARI MEFTAH

Modélisation, détection et annotation des états émotionnels à l'aide d'un espace vectoriel multidimensionnel

préparée à l'I3S-INRIA- Sophia Antipolis ,

soutenue le 12 avril 2013

Jury :

<i>Rapporteurs :</i>	Zohra BELLAHSENE	-	Professeur à l'Université de Montpellier II
	Walid MAHDI	-	Maître de conférence à l'Université de Sfax
	Ladjel BELLATRECHE	-	Professeur à l'ENSMA
<i>Directeur :</i>	Chokri BEN AMAR	-	Professeur à l'Université de Sfax
	Nhan LE THANH	-	Professeur à l'Université de Nice
<i>Président :</i>	Andrea TETTAMANZI	-	Professeur à l'Université de Nice
<i>Examineurs :</i>	Serge MIRANDA	-	Professeur à l'Université de Nice
	Cécile BELLEUDY	-	Maître de conférence à l'Université de Nice

À mon mari Anis, pour sa patience, son aide, son soutien, ses encouragements quotidiens, son écoute,.... Je lui dois une très profonde gratitude.

À mes Chers parents qui m'ont toujours encouragé à atteindre mon but.

À mes adorables grands-parents que j'aime de tout mon coeur.

À ma fille Élè et ma soeur Marwa, dont la seule présence suffit à me remonter le moral.

Remerciements

J'adresse ces tous premiers remerciements à mes rapporteurs : Walid Mahdi (maître de conférences à l'Institut Supérieur d'Informatique et de Multimédia de Sfax), Zohra Bellahsene (professeur des universités à l'université de Montpellier II) et Ladjel Bellatreche (Professeur des universités, LISI / ENSMA) pour m'avoir fait l'honneur d'évaluer mes travaux de recherche. Je remercie aussi les autres membres de jury : Andrea Tettamanzi, Serge Miranda et Cécile Belleudy pour m'avoir faire l'honneur et le plaisir de participer et d'assister à ma soutenance de thèse.

Mes remerciements s'adressent ensuite à mes encadreurs Nhan Le Thanh et Chokri Ben Amar pour m'avoir confié ce sujet et pour m'avoir fait bénéficier de leurs connaissances et des meilleurs conditions de travail tant au niveau des moyens matériels que des relations dans le monde académique.

Je tiens ensuite à remercier sincèrement mes amis et tous les anciens et nouveaux membres du laboratoire et en particulier Catherine Faron, Nguyen Thi Hoa Hue, Fabien Gandon, Nadia Cerezo, Michel Buffa, Nafissa, Oumy, Franck, Wafa et Leonardo... ainsi que toutes les personnes que j'ai croisé dans les couloirs.

Merci à toutes les personnes avec qui j'ai pu travailler au cours de ces années de doctorat : Alain Giboin, Peter Sander, Docteur Marcel Chatel professeur de médecine et M. Alain Anfonso responsable du projet Gerhome.

Merci aussi à Céline Maurel pour son aide et sa gentillesse.

Je remercie aussi toutes les personnes qui m'ont aidé au niveaux administratifs je pense en particulier à Régine Saelens, Sarah Choulet, Micheline Hagnéré et Mireille Blay-Fornarino.

Je remercie également toute ma famille de l'autre côté de la Méditerranée, et surtout ma mère Amel de m'avoir toujours soutenu pendant toutes mes années d'études. Mes sincères remerciements vont aussi à mes frères Aymen et Amine et à mes soeurs Marwa et Lina.

Je ne remercierai jamais assez mon mari Anis, qui a eu la lourde tâche de me supporter pendant toutes ces années, sa patience, son aide et ses encouragements m'ont largement aidé à mener à bien cette thèse...Merci mille fois. Merci enfin à ma petite fille que j'aime très fort.

Table des matières

1	État de l'art	5
1.1	Introduction	5
1.2	Émotions en psychologie	6
1.2.1	Définition du terme émotion	6
1.2.2	Théories des émotions	7
1.2.2.1	Théorie de James-Lange	7
1.2.2.2	Théorie de Cannon-Bard	8
1.2.2.3	Théorie de l'évaluation cognitive	8
1.2.3	Description des émotions	9
1.2.3.1	Approche dimensionnelle	10
1.2.3.2	Approche catégorielle	10
1.2.3.3	Modèle de Plutchik	11
1.2.3.4	Discussion	13
1.3	Émotions en informatique	15
1.3.1	Introduction	15
1.3.2	E-learning	16
1.3.3	Robotique pour l'assistance	17
1.3.4	Jeux Vidéo	17
1.3.5	Systèmes de dialogues	18
1.3.6	Discussion	18
1.4	Reconnaissance des émotions	18
1.4.1	Les expressions faciales	19
1.4.2	Reconnaissance acoustique des émotions	20
1.4.3	Reconnaissance par les signaux physiologiques	22
1.4.3.1	Activité électrodermale (GSR)	22
1.4.3.2	Pression sanguine volumique (BVP)	23
1.4.3.3	Volume et rythme respiratoire (RESP)	23
1.4.3.4	L'activité électromyographique (EMG)	23
1.4.4	La reconnaissance des émotions à partir des signaux physiologiques	24
1.4.5	Classification automatique des émotions	26
1.4.5.1	K plus proches voisins KNN (K Nearest Neighbor)	27
1.4.5.2	Support Vector Machines : SVM	28
1.4.5.3	Discussion	30
1.5	Représentation et échange des émotions	30
1.5.1	Introduction	30
1.5.2	Emotion Markup Language : EmotionML	30
1.5.3	Langage EARL (Emotion Annotation and Representation Language)	33

1.5.4	Langage EMMA (Extensible MultiModal Annotation markup language)	34
1.5.5	Synthèse	36
1.6	Conclusion	37
2	Un modèle de représentation et d'échange des émotions	39
2.1	Introduction	39
2.2	La couche psychologique	40
2.3	La couche formelle de représentation et de traitement des émotions .	42
2.3.1	Les émotions de base	44
2.3.2	Opération sur les émotions	45
2.3.2.1	Addition	45
2.3.2.2	La multiplication par un scalaire	47
2.3.2.3	Décomposition des émotions	49
2.3.2.4	Projection	49
2.3.2.5	Calcul de distance entre émotions	49
2.4	La couche Langage	51
2.5	Discussion	56
2.6	Conclusion	57
3	Reconnaissance des émotions basée sur les signaux physiologiques	59
3.1	Introduction	59
3.2	Acquisition des données	61
3.3	Schéma général de la méthode	62
3.4	Apprentissage	64
3.4.1	Découpage	64
3.4.2	Lissage	65
3.4.3	Détection des pics	69
3.4.4	Extractions des vecteurs caractéristiques	70
3.4.5	Création de la base d'apprentissage	72
3.5	Détection	72
3.6	Expérimentation et résultats	74
3.6.1	Choix de la taille de la fenêtre de lissage	75
3.6.2	Choix de la valeur k de l'algorithme KNN	77
3.6.3	Résultats généraux et comparaison	81
3.7	Conclusion	85
4	Reconnaissance multimodale	87
4.1	Introduction	87
4.2	Fusion de données	88
4.2.1	Fusion de descripteurs	88
4.2.2	Fusion de décisions	88
4.2.2.1	Méthode de Vote	89
4.2.2.2	Méthode probabiliste : les réseaux Bayésiens	90

4.3	Approche proposée pour la reconnaissance multimodale des émotions	90
4.3.1	Résultats	91
4.4	Exemple d'application : Détection de la dépression	94
4.4.1	Introduction	94
4.4.2	Présentation du projet Gerhome	96
4.4.3	La Dépression	97
4.4.4	Méthode de prévention de la dépression	98
4.5	Conclusion	100
A	Glossaire	109
B	Un modèle logistique pour la modélisation de la dépression	111
B.1	Modélisation avec la régression Logistique- Principe	112
B.1.0.1	Estimation des paramètres par la maximisation de la vraisemblance	112
B.2	Jeu de données pour la régression logistique	113
	Bibliographie	115

Table des figures

1	Robots capables d'exprimer des émotions. À droite le robot domestique intelligent du film I-Robot. À gauche les robots Wall-E et Eve du film Wall-E	1
2	Exemples de robots personnels et sociaux	2
1.1	Théorie des émotion de James-Lange	8
1.2	Théorie des émotion de Cannon-Bard	9
1.3	Représentation dimensionnelle de quelques émotions [Posner 2005] . .	11
1.4	le modèle de Circumplex de Plutchik [Plutchik 1980]. Les mots indiqués dans chaque cadran n'ont qu'une valeur indicative. Ce sont des traductions du dictionnaire.	13
1.5	Cône des émotions de Plutchik tiré de [Plutchik 1980]. Les émotions élémentaires sont placées dans une roue. Les émotions secondaires correspondent à des mélanges d'émotions primaires. La roue peut être transformée en cône afin de représenter les différents degrés d'intensité des émotions primaires et secondaires.	14
1.6	Liste des unités d'action relatives aux 6 expressions faciales [Zhang 2005]	20
1.7	(a) LEM du visage, (b) le modèle de l'expression faciale [Gao 2003] .	20
1.8	capteur de respiration [Williams 2004]	23
1.9	capteur EMG appliqué sur le muscle de la mâchoire [Williams 2004]	24
1.10	L'environnement de l'acquisition utilisé dans [Scheirer 2002]	25
1.11	Exemple de classification par la méthode des k plus proches voisins : 3 classes, $k = 10$	28
1.12	Exemple de classification binaire linéaire et non-linéaire dans $\mathbb{R}^2$ [Mercier 2010]	29
1.13	Framework multimodal proposé par le W3C intégrant le langage EMMA [Rouillard 2008]	35
2.1	schéma global du modèle à 3 couches pour échanger les émotions . .	40
2.2	Système colorimétrique RVB	42
2.3	les combinaisons des émotions (Plutchik)	45
2.4	Exemple d'addition de deux émotions dans une application bimodale de détection des émotions	48
2.5	Emotions basiques, similarité, intensité [Plutchik 1980]	50
2.6	Un exemple d'échange émotionnel entre deux utilisateurs	54
2.7	Schéma global de notre modèle de représentation des émotions. . . .	58
3.1	Exemple d'une session de données recueillies dans une journée par les 4 capteurs : EMG, GSR, BVP et respiration	63
3.2	Schéma général de la méthode de détection monomodale d'émotions	63

3.3	Schéma général de l'étape d'apprentissage	64
3.4	Découpage du signal en portions correspondant aux huit émotions	65
3.5	Découpage du signal EMG : isolation de l'émotion colère	65
3.6	Des portions des signaux physiologiques correspondant aux émotions : la colère (à gauche) et la peine (à droite). Les signaux de haut en bas : electromyogram, pression de volume de sang , conductivité de peau et respiration.	66
3.7	Forme d'un filtre moyennneur de longueur égale à $l = 5$	67
3.8	Coefficients d'un filtre gaussien de longueur égale à $l = 5$	67
3.9	Coefficients d'un filtre de Hanning de longueur égale à 5	67
3.10	Lissage du signal EMG avec les différents filtres : filtre moyennneur de longueur 5, filtre gaussien de longueur 5, filtre moyennneur de longueur 100 et filtre de Hanning de longueur 100	68
3.11	Comparaison de la qualité de lissage avec le filtre de Hanning $l = 100$ et le filtre moyennneur $l = 100$. On remarque que la courbe (a) est bien lissée par-contre la deuxième courbe (b) présente des variations qui reflète l'existence du bruit dans le signal	69
3.12	Détection des maxima et minima locaux dans un signal donné	70
3.13	Détection des pics dans un signal	70
3.14	Formes des pics relatives aux trois émotions : haine, colère et joie	71
3.15	Calcul de la largeur et de la longueur d'un pic	72
3.16	Courbes de précision relatives aux émotions : amour platonique, colère, haine et peine, en utilisant les signaux physiologique EMG en fonction de la taille de la fenêtre de lissage	75
3.17	précision globale de classification en fonction de la taille de la fenêtre de lissage	76
3.18	Courbes de précision relatives aux émotions : amour platonique, colère, haine et peine, en utilisant les signaux physiologiques EMG en fonction de la taille de la fenêtre de lissage	77
3.19	précision globale de classification en fonction de la taille de la fenêtre de lissage	78
3.20	Précision globale de classification pour la modalité EMG en fonction de la valeur de k	79
3.21	Précision globale de classification pour les modalités RESPIRATION, GSR et BVP en fonction de la valeur de k	79
3.22	Exemple de recherche utilisant KNN avec $k = 3$, $k = 5$, $k = 10$. Les points rouges représentent la classe colère, les points bleus la classe joie et les points verts la classe haine.	80
3.23	Taux de reconnaissance finaux de la méthode de détection proposée pour la modalité EMG.	81
3.24	Taux de reconnaissance finaux de la méthode de détection proposée pour la modalité RESPIRATION.	82
3.25	Taux de reconnaissance finaux de la méthode de détection proposée pour la modalité GSR.	82

3.26	Taux de reconnaissance finaux de la méthode de détection proposée pour la modalité BVP.	83
3.27	La moyenne globale du taux de reconnaissance des modalités pour toutes les émotions	84
4.1	Organisation générale d'un processus multimodal basé sur la fusion de descripteurs	88
4.2	Organisation générale d'un processus multimodal basé sur la fusion de décisions	89
4.3	La topologie d'un réseau Bayésien pour la reconnaissance bimodale des émotions [Sebe 2005]	90
4.4	Reconnaissance multimodale en utilisant l'addition du modèle multidimensionnel	92
4.5	Comparaison entre les résultats de l'approche monomodale et de l'approche multimodale	93
4.6	Comparaison entre les résultats de l'approche monomodale et de l'approche multimodale	93
4.7	Extraction de la caractéristique d'un signal EMG. La moyenne est calculée par scénario (intervalle de temps de chaque émotion)	95
4.8	Appratement Gerhome	96
4.9	Architecture du système de suivi à distance d'une personne à domicile	97
4.10	Chaussure avec capteur de conductivité de la peau	98
4.11	Boucle d'oreille captant la pression sanguine	98
4.12	Boucle d'oreille captant la pression sanguine	99
4.13	Prévention de la dépression	101

Liste des tableaux

1.1	Ensemble d'émotions basiques selon différents auteurs. Tableau extrait de [Tato 1999]	12
3.1	Ensemble des images utilisées pour provoquer les huit émotions ainsi qu'une description de ces états émotionnels	62
3.2	Précisions obtenues en faisant varier la taille de la fenêtre de Hanning	76
3.3	Précisions obtenues en faisant varier la valeur k de l'algorithme KNN	78
3.4	Tableau comparatif du taux de classification des différents méthodes	84
4.1	Tableau comparative des taux de classification des différents méthodes	94
4.2	les 11 valeurs caractéristiques extraites pour les 8 émotions du travail de MIT[Healey 2000]. μ_ϵ la moyenne du signal EMG, $\tilde{\mu}_g$ la moyenne du signal GSR normalisée, $\tilde{\delta}_g$ l'écart type du signal GSR normalisé, μ_B la moyenne de BVP, δ_B l'écart type de BVP, $\tilde{\mu}_R$ la moyenne du signal RESPIRATION normalisée, $\tilde{\sigma}_R^2$ l'écart type de la RESPIRATION, p_1 , p_2 et p_3 représentent la densité spectrale de puissance du signal RESPIRATION	95

Introduction générale

La science-fiction est un genre narratif cinématographique structuré par des hypothèses sur ce que pourrait être le futur. En partant des connaissances actuelles, elle met en scène des univers où se déroulent des faits impossibles ou non avérés en l'état actuel de la civilisation, des techniques ou de la science, mais qui correspondent généralement à des découvertes scientifiques et techniques à venir. Plusieurs films de science-fiction ont marqué nos esprits. On peut citer par exemple le film Wall-E de Andrew Stanton et le film I. Robot avec Will Smith. Ces oeuvres, mêlant fiction et réalité, incarnent la vie future où le robot a sa place à part entière dans la vie quotidienne. Les robots sont à l'image de l'humain tant sur le plan physique que sur le plan des compétences (raisonnement, mémoire, conscience ...) et surtout possédant des émotions. Ce genre de film nous fait voyager dans le futur, laissant supposer la possibilité d'une parfaite coopération homme-machine, où le robot est capable d'exprimer ou de prendre en compte les émotions de l'homme. Aujourd'hui cet aspect de la fiction n'est pas aussi loin qu'on ne l'imagine de la réalité. En effet, plusieurs travaux de recherche se penchent sur l'intégration d'une dimension émotionnelle dans les personnages virtuels ou les robots, afin d'apporter une illusion de vie aux utilisateurs, mais aussi améliorer l'interaction humain-machine. Plusieurs robots expérimentaux, appelées robots de compagnie, peuvent aujourd'hui communiquer avec l'utilisateur tout en tenant compte de son état d'humeur et en exprimant des émotions. Nous pouvons citer par exemple le chien robot Aibo [Veloso 2006] de Sony ou le robot humanoïde Nao [Gouaillier 2008] développé par Aldebaran Robotics ou encore le visage du robot d'Einstein développé par David Hanson [Hanson 2005]. Ces robots peuvent avoir une personnalité, sont capables de converser et d'établir un contact visuel avec nous. Ils offrent une gamme complète d'expressions faciales. Ils arrivent à analyser la parole et à modéliser l'état émotionnel de la personne.

Ces exemples témoignent de l'importance du rôle que jouent les émotions dans la vie de tous les jours. En effet, elles sont au centre de notre vie. Elles exercent



(a) I Robot

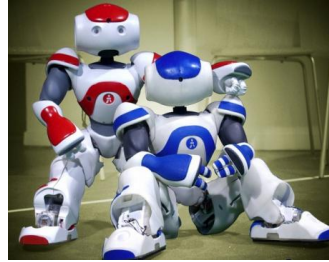


(b) Wall-E

FIGURE 1 – Robots capables d'exprimer des émotions. À droite le robot domestique intelligent du film I-Robot. À gauche les robots Wall-E et Eve du film Wall-E



(a) Le robot AIBO de sony



(b) Le robot humanoïde NAO

FIGURE 2 – Exemples de robots personnels et sociaux

une influence majeure sur les relations sociales entre les individus et jouent un rôle essentiel dans la prédiction (lecture) de l'état mental et des actions d'un humain. De nombreux domaines comme la philosophie, la psychologie, la sociologie, etc., se sont intéressés depuis plus d'un siècle aux émotions. Par contre, l'informatique ne s'est penchée que récemment, dans les années quatre-vingt-dix, sur les émotions, donnant naissance à l'informatique affective [Picard 1997a].

L'informatique affective cherche à donner aux machines (ordinateurs ou robots) la capacité de reconnaître, d'exprimer et de modéliser les émotions. Boosté par des domaines tels que la surveillance de santé de personnes, jeu vidéo interactif, communication ou sondage, le nombre de recherches scientifiques dans ce domaine a explosé. Ces recherches essaient entre autre de modéliser les états affectifs, de mieux reconnaître l'état émotionnel des individus et de définir les relations émotionnelles homme-machine. Une relation émotionnelle homme-machine se définit comme un ensemble d'interactions durant lesquelles l'utilisateur et la machine prêtent attention aux émotions de chacun, communiquent leurs émotions, expriment de l'empathie et établissent ainsi un lien. Pour cela, il est nécessaire que la machine soit dotée de capacités de perception, de compréhension et d'analyse des états émotionnels des utilisateurs pour rendre l'échange plus réaliste.

Généralement, les études portent sur des émotions primaires telles que la joie, la tristesse, la peur, etc. Mais en réalité, et dans la vie de tous les jours, les émotions sont souvent complexes et impliquent plusieurs émotions simultanées. Elles surviennent par exemple comme une succession rapide d'émotions, une superposition, une émotion masquée, supprimée ou au contraire exagérée. Par exemple, on peut ressentir un mélange de tristesse et de colère suite à une expérience de trahison par un ami ou un sentiment mixte de peur et de colère suite à une injustice infligée par son patron.

A l'heure actuelle, il n'existe pas de modèle générique pour échanger les émotions entre applications différentes. Les modèles des émotions développés en informatique sont très divers en termes de conception et de fonctionnement. L'utilité de ces modèles est limitée par les applications spécifiques car ils sont créés pour répondre à un contexte d'interactions prédéfini.

Ainsi, plusieurs questions scientifiques se posent et ouvrent sur les problé-

matiques suivantes : Comment modéliser les émotions en tenant compte de leur complexité? Comment effectuer l'échange émotionnel lors d'une interaction homme-machine ou machine-machine? Comment modéliser les émotions en informatique en se basant sur les différentes approches des émotions en psychologie ?

Notre travail s'inscrit dans le cadre de la modélisation informatique des émotions qui est une thématique reconnue dans le domaine de l'intelligence artificielle. L'objectif principal de cette thèse est d'essayer d'apporter des éléments de réponse relatifs aux différentes problématiques citées ci dessus et de contribuer à l'amélioration de l'interaction homme-machine. Notre but est de pouvoir identifier, décrire et échanger les qualités sémantiques des émotions issues des modalités multiples (parole, expression faciale, geste, etc.) lors d'une interaction homme-machine ou machine-machine.

Les principaux objectifs que nous nous sommes fixés sont les suivants :

- modélisation des émotions en tenant compte de leur richesse et de leur complexité ;
- proposer un système d'échange émotionnel complet permettant le partage aussi bien des émotions primaires que des émotions mixtes ou mélangés ;
- proposer un algorithme de détection des émotions en se basant sur notre modèle de représentation des émotions et sur les outils fournis par le domaine du traitement de signal.

Organisation du manuscrit

Le manuscrit est composé de 4 chapitres organisés de la façon suivante :

Le premier chapitre introduit en premier lieu des notions de base sur les émotions dans le domaine de la psychologie, il décrit les principales théories sur la représentation des émotions. Puis, il présente le rôle de l'émotion en intelligence artificielle et plus précisément en reconnaissance automatique de l'émotion. Nous présentons à cet égard, un état de l'art relatif à la reconnaissance automatique des émotions en se focalisant sur la reconnaissance des émotions à partir des signaux physiologiques. Nous clôturons le chapitre par une revue des travaux de modélisation et d'annotation des émotions.

Dans le deuxième chapitre, nous présentons notre modèle générique pour la représentation et l'échange des émotions entre applications multimodales. Nous décrivons en détail les trois couches qui le composent : la couche psychologique, la couche formelle de représentation et de traitement des émotions et la couche langage.

Le troisième chapitre présente une nouvelle approche de reconnaissance des émotions basée sur l'analyse des signaux physiologiques. Nous décrivons tout d'abord la base des signaux physiologiques qu'on a utilisés. Nous présentons ensuite les différentes étapes du traitement effectuées pour la reconnaissance des émotions. Enfin, nous exposerons les résultats obtenus.

Le quatrième chapitre présente une approche multimodale pour la reconnaissance des émotions. Nous décrivons, tout d'abord le principe de l'approche. Ensuite nous

présentons les résultats expérimentaux obtenus. Puis nous présentons un exemple d'application de détection de la dépression.

Ce rapport s'achève sur une conclusion générale, qui comporte une synthèse de notre contribution, dévoile l'originalité de notre travail et précise ses limites et les perspectives de recherche envisageables.

CHAPITRE 1

État de l'art

Sommaire

1.1	Introduction	5
1.2	Émotions en psychologie	6
1.2.1	Définition du terme émotion	6
1.2.2	Théories des émotions	7
1.2.3	Description des émotions	9
1.3	Émotions en informatique	15
1.3.1	Introduction	15
1.3.2	E-learning	16
1.3.3	Robotique pour l'assistance	17
1.3.4	Jeux Vidéo	17
1.3.5	Systèmes de dialogues	18
1.3.6	Discussion	18
1.4	Reconnaissance des émotions	18
1.4.1	Les expressions faciales	19
1.4.2	Reconnaissance acoustique des émotions	20
1.4.3	Reconnaissance par les signaux physiologiques	22
1.4.4	La reconnaissance des émotions à partir des signaux physiologiques	24
1.4.5	Classification automatique des émotions	26
1.5	Représentation et échange des émotions	30
1.5.1	Introduction	30
1.5.2	Emotion Markup Language : EmotionML	30
1.5.3	Langage EARL (Emotion Annotation and Representation Language)	33
1.5.4	Langage EMMA (Extensible MultiModal Annotation markup language)	34
1.5.5	Synthèse	36
1.6	Conclusion	37

1.1 Introduction

L'objectif de notre travail est de proposer de nouvelles approches pour la modélisation et l'annotation des états émotionnels afin de pouvoir effectuer des échanges

émotionnels entre différents applications. Un tel objectif ne peut être atteint qu'en se basant sur les travaux effectués précédemment sur les émotions. Pour cela, il est d'abord nécessaire de bien comprendre le phénomène émotionnel avant d'aborder un tel sujet. C'est pourquoi nous débutons ce manuscrit par présenter quelques notions concernant les émotions telles que leurs définitions, leurs différentes théories et leurs composantes. Nous présentons par la suite une vue d'ensemble sur les recherches menées en intelligence artificielle qui se sont intéressées aux émotions. Ensuite, nous présentons un état de l'art sur la reconnaissance des émotions. Enfin nous présentons des travaux de représentation et d'annotation des émotions.

1.2 Émotions en psychologie

1.2.1 Définition du terme émotion

L'émotion est une notion floue et elle est difficilement définissable. De ce fait, il existe plusieurs définitions du mot « émotion ». Mais le seul point sur lequel tous les psychologues s'accordent, c'est que le concept est difficile à définir. En effet, la complexité de la nature de l'émotion humaine ainsi que la diversité des formes avec lesquelles elle s'exprime ont, longtemps, empêché l'accès à une définition univoque du concept. Le mot « émotion » vient du latin « *emovere*, *emotum* » (enlever, secouer) et de « *movere* » (se mouvoir) signifiant mouvement vers l'extérieur. D'après l'étymologie, les émotions produisent donc des changements, psychiques ou de comportement, qui provoquent des états subjectifs internes. Ces états internes peuvent être positifs agréables comme la joie ou négatifs désagréables, comme la colère. Dans ce sens Ekman et Davidson [Ekman 1994] définissent l'émotion comme une réaction aiguë et transitoire, provoquée par un stimulus spécifique et caractérisée par un ensemble cohérent de réponses cognitives, physiologiques et comportementales [Petropoulou 2006].

Plusieurs définitions et rôles ont été donnés à l'émotion. Ces définitions diffèrent en fonction des différentes approches proposées. En 1879, Charles Darwin [Darwin 1872], fondateur de la théorie de l'évolution, la définit comme une qualité innée, universelle et communicative, liée au passé de l'évolution de notre espèce. Ekman [Ekman 1982], Izard [Izard 1977], Plutchik [Plutchik 1980], Tomkins [Tomkins 1980] et Mac Lean [Maclean 1993] ont développé la théorie des émotions de bases ou fondamentales, mais seules 6 émotions de base sont communes aux divers auteurs. Ces émotions sont : la Tristesse, la Colère, la Joie, le Dégoût, la Peur et la Surprise, connues sous le nom "Big Six" [Ekman 1994]. D'après James [William 1903] il n'y a pas de limite au nombre des différentes émotions de base qui peuvent exister, et c'est pourquoi les émotions des différentes théories peuvent varier indéfiniment à la fois quant à leur constitution et quant aux objets qui les engendrent, car il n'y a rien de fixe de toute éternité dans l'action réflexe [de Bonis 1996]. Ce point de vue se retrouve chez un certain nombre de spécialistes contemporains de l'émotion [Frijda 1986]. De leur côté, Kleinginna A.M et Klein-

ginna P.R. [Kleinginna 1981] ont recensé plus de 140 définitions, reflétant chacune différents aspects du processus émotionnel.

Malgré ces divergences, la majorité des psychologues s'accordent sur une définition consensuelle des états émotionnels. Selon celle-ci, l'émotion est réductible à trois composantes, plus ou moins indépendantes entre elles [Watson 1992] :

- une composante physiologique,
- une composante comportementale,
- une composante cognitive.

La *composante physiologique* de l'émotion se traduit par un changement du rythme cardiaque, une modification des réponses électrodermales, un changement de la température du corps, une variation de la fréquence respiratoire, etc. Ekman, Levenson et Friesen (1983) ont étudié les réactions physiologiques de six émotions (joie, tristesse, dégoût, colère, peur et surprise) à l'aide de cinq mesures physiologiques. Les résultats ont montré que chaque émotion est associée à des réactions spécifiques. Par exemple la colère est caractérisée par une forte augmentation des battements du cœur et de la température cutanée. Tandis que la peur n'est caractérisée que par une augmentation des battements du cœur. D'autres travaux ont montré que le niveau d'activité physiologique est lié à la valence émotionnelle (émotion positive/ négative) puisque les émotions négatives présentent des réponses du système nerveux autonome plus fortes [Cacioppo 1997].

La *composante comportementale* fait référence aux réactions comportementales et expressives. Par exemple, la peur qui est éprouvée lors d'une menace déclenche une séquence comportementale de défense ou de survie. La composante comportementale inclut l'expression faciale, les comportements gestuels (les changements de la posture, les mouvements du corps et des bras, etc.) et vocaux.

La *composante cognitive* comprend l'ensemble des processus mentaux qui apparaissent et qui se développent suite au déclenchement d'une émotion intervenant dans la perception de la situation, dans sa mémorisation et dans son traitement cognitif en mémoire. En d'autres termes, cette composante désigne la composante subjective de l'expérience émotionnelle. Dans l'approche cognitive de l'émotion, l'évaluation et la tendance à l'action sont deux processus essentiels. L'individu évalue si la situation est nouvelle ou si elle s'est déjà produite avant. Il évalue dans quelle mesure cette situation est plaisante ou déplaisante et quel contrôle il peut avoir sur elle. La composante cognitive-expérientielle évalue la nouveauté, la valence et le contrôle que peut avoir le sujet sur la situation. L'évaluation cognitive est souvent considérée comme un processus initial au déclenchement de l'émotion.

1.2.2 Théories des émotions

1.2.2.1 Théorie de James-Lange

Parmi les premières théories de l'émotion figure celle qu'a proposé en 1884 le fameux psychologue et philosophe américain, William James [William 1903], très proche de celle du psychologue danois, Carl Lange (1885). C'est pour cette raison,

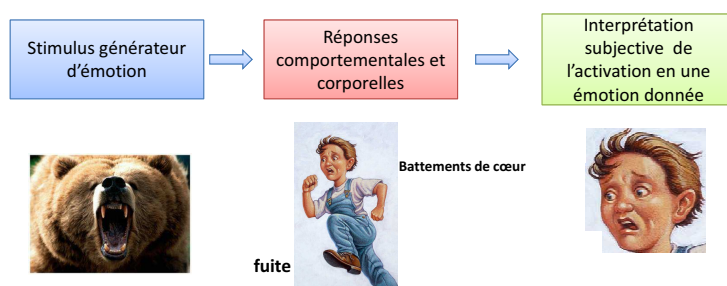


FIGURE 1.1 – Théorie des émotion de James-Lange

que dans la littérature, on parle souvent de la théorie périphérique de James-Lange. La théorie de James-Lange, énonce que l'émotion traduit la réponse aux modifications physiologiques intervenant dans le corps. Pour James William [William 1903] « L'émotion est notre perception des modifications qui surviennent [dans notre corps] ». Par exemple, on est triste parce que l'on pleure, plutôt que l'on pleure parce que l'on est triste. La FIGURE 1.1 montre un exemple de la théorie de James-Lang : lorsque une personne se retrouve en face d'un danger comme un animal effrayant par exemple, l'amène à courir pour fuir, et le fait de courir entraîne la peur. C'est-à-dire, un stimulus conduit à des réactions viscérales (battements de coeur, forte respiration) et comportementales, qui à leur tour, sont interprétées comme une émotion.

1.2.2.2 Théorie de Cannon-Bard

En 1927, le physiologiste américain Walter Cannon a publié un article critiquant la théorie de James-Lange et il a proposé une nouvelle théorie. La théorie de Cannon sera développée par Philip Bard (1934). Cannon et Bard ont noté que les réponses viscérales sont habituellement lentes et elles surviennent environ une à deux secondes après l'apparition du stimulus. A l'inverse, les réponses émotionnelles sont immédiates et précèdent souvent aussi bien les réactions viscérales que les comportements [Westen 2000]. Cannon et Bard ont donc proposé une théorie, selon laquelle les stimuli gèrent des émotions produisant simultanément une expérience émotionnelle et des réponses corporelles (cf. FIGURE 1.2).

1.2.2.3 Théorie de l'évaluation cognitive

Les théories de James-Lange et de Cannon-Bard mettent en évidence le fait que le phénomène émotionnel est accompagné de manifestations physiologiques. Cependant, les aspects sociaux ne sont pas abordés. L'environnement est ici envisagé uniquement comme stimulus. Or la particularité de l'environnement est qu'il peut affecter l'ampleur et l'intensité des manifestations physiologiques ressenties. Par exemple une personne qui rencontre un rat, ne dégage pas forcément un sentiment de dégoût et il peut même dégager des émotions positives d'admiration comme en Inde ou les rats peuvent être vénérés. Il est donc nécessaire de prendre en considération

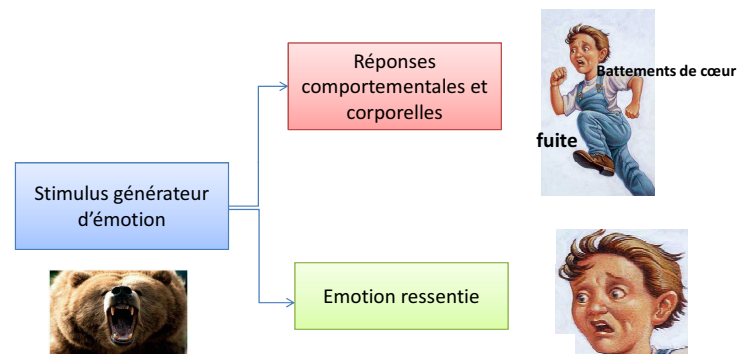


FIGURE 1.2 – Théorie des émotion de Cannon-Bard

le facteur environnemental déterminé par l'apprentissage social pour l'évaluation de l'événement stimulus. C'est de fait, ce que les théories cognitives ont développé dans leurs modèles.

Les modèles d'évaluation cognitive (appraisal theory of emotion) ont été introduits pour la première fois par Arnold en 1960 [Arnold 1960]. Ces modèles supposent qu'il est nécessaire de comprendre les évaluations que l'individu fait au sujet des événements de son environnement pour comprendre les émotions. Donc, une évaluation cognitive est définie comme un processus cognitif, rapide, automatique et inconscient dont la fonction est d'évaluer les stimuli perçus sur la base de critères particuliers. Selon cette théorie, une émotion est déclenchée par l'évaluation subjective d'un événement. La perception et l'évaluation cognitive d'un événement déterminent le type et l'intensité de l'émotion ressentie [Scherer 2000]. Plus précisément, le déclenchement d'une émotion est issu de l'évaluation d'un ensemble de variables (appelés variables d'évaluation). Les valeurs des variables d'évaluation dépendent autant des facteurs culturels que de l'état mental de l'individu (but, croyances, etc.) et de son profil (personnalité, préférences, etc.) [Lazarus 1991]. Ainsi, une même situation peut déclencher deux émotions distinctes chez deux individus différents.

Ainsi les modèles d'évaluation cognitive fournissent un cadre théorique qui permet de relier le processus cognitif de génération de l'émotion à la réaction émotionnelle produite. Du fait de cette richesse, elles ont été très utilisées dans le domaine de l'intelligence artificielle, principalement dans les travaux sur les modèles computationnels d'émotions qui essaient de simuler le processus de génération de l'émotion [Gratch 2004]. La plupart des modèles sont fondés sur les théories de l'évaluation cognitive de Ortony, Clore et Collins "le modèle OCC" [Ortony 1988] dans lequel 22 émotions types sont définies suivant leurs conditions de déclenchement.

1.2.3 Description des émotions

Plusieurs approches ont été présentées pour décrire l'ensemble des émotions. Les approches les plus connues sont :

- l'approche catégorielle,
- l'approche dimensionnelle.

1.2.3.1 Approche dimensionnelle

L'approche dimensionnelle est basée sur le principe que les émotions résultent d'un nombre fixé de concepts et propose donc de les représenter dans un espace multi-dimensionnels. Par exemple, les dimensions peuvent être un axe de plaisir et de déplaisir, d'éveil ou d'ennui, de nervosité, de puissance, de maîtrise de soi et bien d'autres au besoin du modèle.

Wundt (1896) [Wundt 1896] est le premier à élaborer un modèle dimensionnel des émotions. Il décrit la sphère de la conscience liée aux émotions et aux sentiments en fonction de trois dimensions : plaisir-déplaisir, relaxation-tension, calme-excitation. Par la suite, Schlosberg a proposé un modèle à deux (bonheur-colère, surprise-peur), puis trois dimensions (il a ajouté la dimension de l'activation « sommeil-tension ») [Schlosberg 1954].

Russel [Russell 1980], en 1980 propose de représenter les émotions dans un espace bi-dimensionnel où se croisent quatre oppositions bipolaires. Malgré le nombre variable de dimensions proposées par les différents auteurs, les dimensions valence et activation (ou arousal) sont les facteurs qui montrent le plus d'importance et de consistance dans le cadre des théories dimensionnelles. La valence permet de distinguer les émotions positives, agréables, comme la joie, des émotions négatives, désagréables, comme la colère. L'activation représente le niveau d'excitation corporelle, qui transparaît par nombre de réactions physiologiques, comme l'accélération du coeur, la transpiration, etc.

Certains travaux ajoutent toutefois une troisième dimension, trouvant les deux premières insuffisantes. Cette troisième dimension est nommée contrôle ou dominance [Russell 1997]. Elle correspond à l'effort du locuteur pour contrôler son émotion et permet de distinguer les émotions provoquées par le sujet lui-même ou par l'environnement. Elle permet de distinguer les émotions donnant lieu à des réactions d'approche et de combat comme la colère, de celles engendrant des comportements de fuite comme la peur.

1.2.3.2 Approche catégorielle

L'approche catégorielle est l'approche la plus répandue. Elle consiste à considérer les émotions comme des caractéristiques épisodiques et universelles [Darwin 1872]. Elle est basée sur un ensemble d'émotions dites « basiques », universelles, non réductibles et innées. Cette conception intègre une perspective évolutionniste issue des travaux de Darwin [Darwin 1872]. Le caractère universel des émotions entraîne la définition d'un petit nombre d'émotions basiques (la peur, la colère, etc.), qui ont pu être observées chez tous les individus, quelque soit leur nationalité ou leur culture.

Dans la continuité des travaux de Darwin, plusieurs théories qui font référence aux émotions basiques ont été développées. Parmi ces théories, on trouve le modèle de Tomkins [Tomkins 1962] qui se base sur 9 émotions de base dont 2 sont positives (la joie et l'intérêt), une qui est neutre la surprise et enfin six émotions négatives qui sont la colère, le dégoût, l'anxiété, la peur, la honte et le mépris. De même, Izard [Izard 1971] y ajoute la tristesse, la culpabilité et la timidité. Par contre il

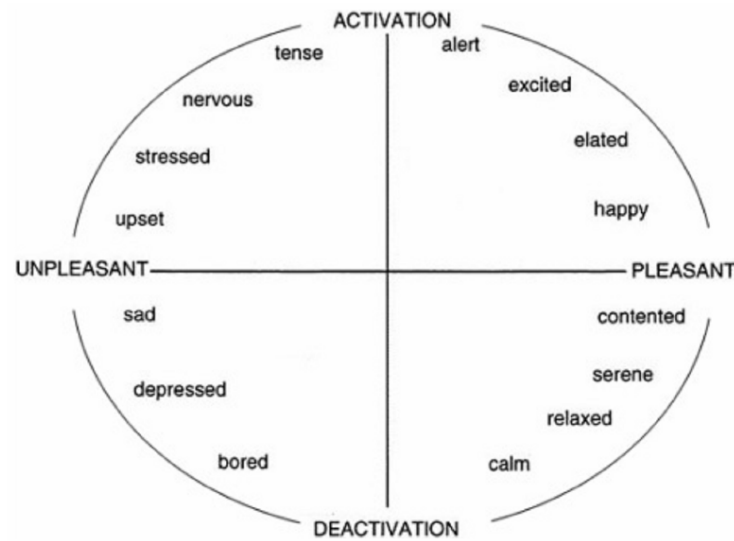


FIGURE 1.3 – Représentation dimensionnelle de quelques émotions [Posner 2005]

ne considère pas le mépris comme une émotion basique. Plutchik [Plutchik 1980] compare les émotions à une palette de couleurs. De la même façon que les couleurs, il a tenté d'établir un dictionnaire des émotions qui permettrait, sur la base d'émotions fondamentales, de dériver les différentes nuances de l'expérience émotionnelle. Dans la suite nous allons bien détailler l'approche de Plutchik, car c'est l'approche que nous allons utiliser pour définir notre modèle.

Dans ces travaux, Arnold (1960) [Arnold 1960] propose onze émotions primaires : la colère, l'aversion, le courage, le dégoût, le désir, la peur, la haine, l'espoir, l'amour, la tristesse et le désespoir. En 1982, Panksepp [Panksepp 1982] a défini quatre émotions fondamentales qui étaient initialement : l'espoir, la peur, la rage et la panique. En 1989, il a proposé quatre autres émotions fondamentales : la colère, le dégoût, la tristesse et l'intérêt.

Malgré la diversité des propositions, six émotions semblent apparaître dans la majorité de classifications proposées. Ces émotions sont les "Big Six" d'Ekman [Ekman 1999] (colère, peur, tristesse, joie, dégoût et surprise).

1.2.3.3 Modèle de Plutchik

« A une science de la colorimétrie, on pourrait faire correspondre une science de l'émotionométrie » [Plutchik 1980]

Le modèle de Plutchik a abordé le problème d'analogie avec les couleurs en proposant un dictionnaire d'émotions similaire au dictionnaire de couleurs. En effet, de la même façon qu'il existe des couleurs primaires et des nuances de ces couleurs qui varient en intensité et constituent le spectre des couleurs, il y aurait des émotions primaires, des nuances qui varieraient en intensité, des combinaisons possibles entre

TABLE 1.1 – Ensemble d'émotions basiques selon différents auteurs. Tableau extrait de [Tato 1999]

Auteurs	Emotions basiques
Ekman et al.	anger, disgust, fear, joy, sadness, surprise
Izard	anger, contempt, disgust, distress, fear, guilt, interest, joy, shame, surprise
Plutchik	acceptance, anger, anticipation, disgust, fear, joy, sadness, surprise
Tomkins	anger, interest, contempt, disgust, distress, fear, joy, shame, surprise
Arnold	anger, aversion, courage, dejection, desire, despair, fear, hate, hope, love, sadness
Fridja	desire, happiness, interest, surprise, wonder, sorrow
Gray	rage, terror, anxiety, joy
Mower	pain, pleasure
James McDougall	fear, grief, love, rage, anger, disgust, elation, fear, subjection, tender-emotion, wonder
Weiner and Graham	Happiness, sadness
Panksepp	expectancy, fear, rage, panic

certaines émotion de bases. Le modèle de Plutchik est défini par le circumplex qui est un modèle multidimensionnel basé sur des émotions primaires de base dont vont dériver des émotions secondaires. Selon Plutchik, il existe 8 émotions de base faites de 4 paires opposées :

- joie-tristesse,
- acception-dégoût,
- peur-colère,
- surprise et anticipation.

Les huit dimensions des émotions fondamentales sont disposées comme un cône, avec au sommet les termes qui désignent chaque émotion à son intensité maximale. La FIGURE 1.4 montre que l'axe vertical représente l'intensité et que la coupe horizontale montre les émotions intenses : la rage, la vigilance, l'extase, l'adoration, la terreur, l'étonnement, le chagrin et l'aversion. Si on regarde le cône de Plutchik on constate que plus on s'oriente vers le bas plus les différentes nuances des émotions sont difficiles à distinguer. Ainsi, par exemple la distinction entre l'ennui et la contrariété est plus difficile que son équivalent à l'étage supérieur : le dégoût et la colère.

Plutchik a défini des règles d'association des émotions fondamentales pour former des émotions mixtes. En effet, les émotions ne s'associent pas n'importe comment et leur combinatoire répond à des règles fondées sur la méthode des dyades et des triades [de Bonis 1996]. Plutchik a défini ainsi :

- les dyades primaires comme étant la combinaison de deux émotions adjacentes,
- les dyades secondaires comme étant la combinaison d'émotions proches à une émotion près
- et les dyades tertiaires comme étant la combinaison d'émotions voisines à deux

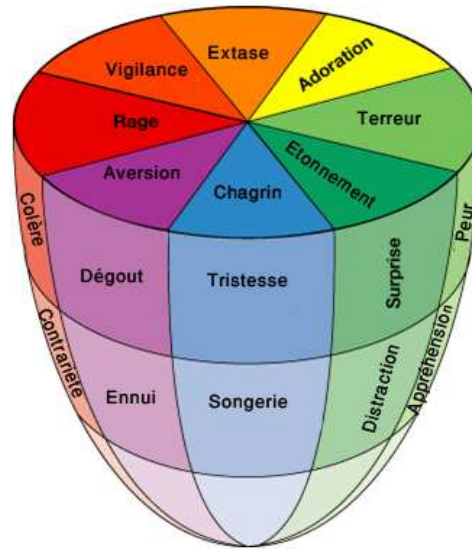


FIGURE 1.4 – le modèle de Circumplex de Plutchik [Plutchik 1980]. Les mots indiqués dans chaque cadran n'ont qu'une valeur indicative. Ce sont des traductions du dictionnaire.

émotions près.

Ainsi pour prendre quelques exemples, les deux émotions adjacentes joie et admiration, forment un composé primaire : l'amour (Fig. 1.5). Un composé secondaire : le désespoir, sera formé de la peur et de la tristesse, si l'on saute un élément. Enfin, un composé tertiaire sera formé de l'association de la peur au dégoût (si l'on saute deux éléments adjacents), composé qui aboutira à la honte.

1.2.3.4 Discussion

Nous avons vu dans cette section deux façons de représenter les émotions : l'approche dimensionnelle et l'approche catégorielle. cette dernière a été remise en question à plusieurs reprises [Ortony 1990] [Panksepp 1982]. En effet, la distinction des états émotionnels varie selon les cultures [Russell 1991]. Aussi, ce type de description discret s'avère insuffisant pour la description des états émotionnels les plus subtils tels que ceux présents dans la vie réelle, où les émotions peuvent s'avérer mélangées entre elles ou avec des manifestations correspondant plutôt à des attitudes [Devillers 2005]. Aussi l'approche catégorielle se base sur des labels pour représenter chaque émotion et chaque variation. Donc, le problème majeur de ce type d'approche vient du nombre de labels utilisés pour couvrir un maximum d'états émotionnels. De plus, chaque label étant un mot d'un langage particulier, il est difficile d'assurer que la traduction d'un terme d'une langue à une autre définit bien la même émotion. Il peut également être difficile d'inférer des transitions entre les différentes émotions. L'approche dimensionnelle a eu beaucoup de succès puisqu'elle est continue et permet la représentation d'une infinité d'émotions. Par contre, même s'il paraît facile

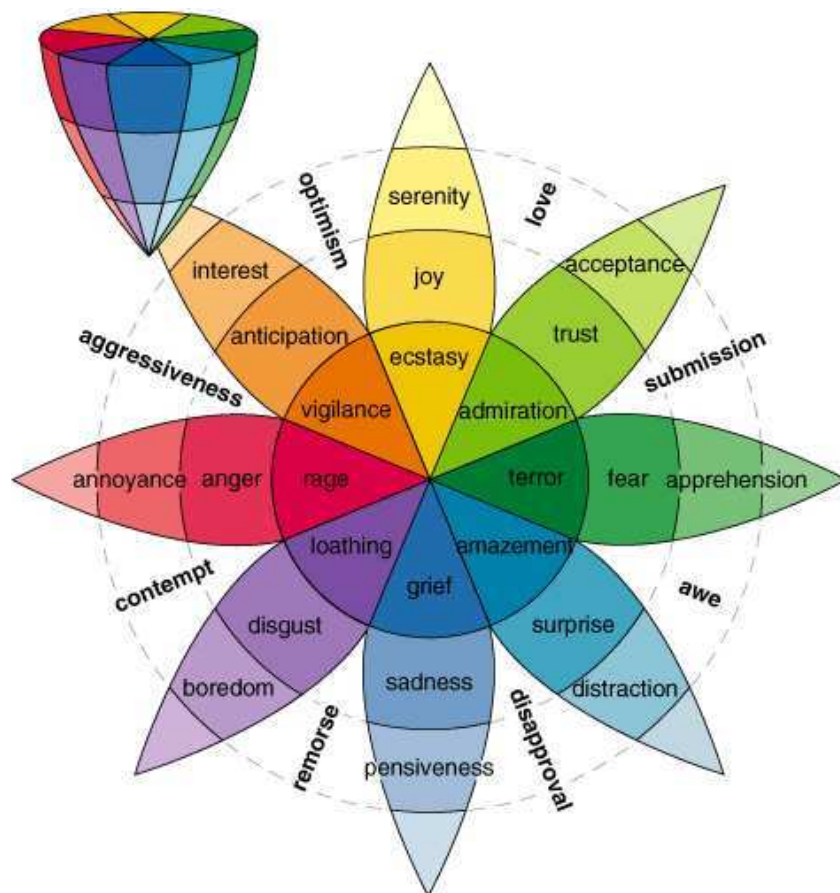


FIGURE 1.5 – Cône des émotions de Plutchik tiré de [Plutchik 1980]. Les émotions élémentaires sont placées dans une roue. Les émotions secondaires correspondent à des mélanges d'émotions primaires. La roue peut être transformée en cône afin de représenter les différents degrés d'intensité des émotions primaires et secondaires.

de positionner de manière approximative les émotions catégorielles sur un espace à plusieurs dimensions, l'inverse l'est beaucoup moins. En effet, deux émotions produisant des réactions éloignées peuvent se trouver très proches l'une de l'autre, par exemple la colère « upset » et la tristesse « sad » se caractérisent avec un fort mécontentement « unpleasant » mais une activation similaire, proche de 0 (cf. FIGURE 1.3).

Dans le cadre de nos recherches, nous adoptons l'approche de Plutchik qui permet de représenter le phénomène de combinaison d'émotions. Il semble en effet important d'utiliser une représentation permettant de prendre en compte ce phénomène très fréquent dans les relations interpersonnelles. Le modèle de Plutchik s'inscrit plus dans la famille des approches catégorielles que dimensionnelles. En effet, ce modèle considère que les expressions émotionnelles constituent quelques catégories émotionnelles de base comme la peur, la colère, la joie, la tristesse... , qui peuvent se combiner pour produire une grande variété d'expressions. Ce modèle peut être aussi considéré comme une approche dimensionnelle [Bruyer 2007], car Plutchik a également utilisé le modèle circumplex en 3 dimensions pour représenter les émotions. Il considère que les expressions sont des combinaisons variables, continues et très nombreuses de quelques dimensions de base qui sont : l'intensité, la similarité et la valence positive ou négative [Plutchik 1980].

1.3 Émotions en informatique

1.3.1 Introduction

Tout comme son importance dans le domaine des sciences cognitives et de la philosophie, l'émotion a aussi vu son intérêt croître dans le champ de l'intelligence artificielle. Plusieurs domaines tels que la surveillance de santé des personnes âgées, jeu vidéo interactif, communication ou sondage s'intéressent de plus en plus à l'état émotionnel des utilisateurs. En effet, la reconnaissance de l'état émotionnel des personnes permet de mieux gérer l'interactivité entre le système et les différents utilisateurs. Ainsi, un courant de recherche en intelligence artificielle est apparu : l'informatique affective, qui intègre de plus en plus la dimension émotionnelle dans les applications informatiques dont le but d'améliorer l'interaction homme-machine. Il vise à la modélisation des émotions et à la conception de systèmes capables de reconnaître, exprimer et synthétiser les émotions [Picard 1997a]. On distingue deux courants de recherche :

- la conception de systèmes capables de reconnaître les émotions de l'utilisateur,
- la modélisation de processus émotionnels dans des systèmes computationnels.

Le premier de ces deux courants de recherche, se concentre sur la modélisation des émotions dans des agents virtuels. Il a pour but de générer des émotions chez des agents virtuels pour améliorer l'interaction homme-machine. Ainsi, de nombreuses recherches ont été menées sur les avatars expressifs (voir par exemple [Neviarouskaya 2007], [Karpouzis 2003]). Ces avatars sont des personnages en trois dimensions capables d'exprimer des émotions selon les différents canaux de com-

munication affective, afin d'améliorer la communication avec un utilisateur humain. Le second de ces deux courants de recherche, utilise et analyse les indices physiologiques et somatiques pour déterminer l'émotion ressentie par l'utilisateur à l'aide de capteurs externes qui enregistrent des signaux physiologiques visuels et acoustiques. Dans le cadre de nos recherches, nous nous intéressons entre autres à ce deuxième courant de recherche. En effet, notre travail est divisé en deux parties : la reconnaissance des émotions et la représentation et l'échange des émotions entre applications. Dans cette section, nous présentons quelques travaux de l'état de l'art d'utilisation des émotions dans le domaine informatique. En effet, ces dernières années, les recherches dans le domaine de l'affective Computing ont donné naissance à des systèmes technologiques de plus en plus sophistiqués. Il existe de nombreux exemples d'utilisation des émotions en informatique tel que le E-learning, la médecine, la robotique, les jeux vidéo, etc.

1.3.2 E-learning

Plusieurs chercheurs soutiennent que les émotions sont fondamentales dans la communication et dans les interactions humaines en général. Picard [Picard 1997a], par exemple, estime que l'introduction d'agents émotionnels dans un système informatique peut permettre l'humanisation de ce dernier et ainsi faciliter les échanges Homme-Machine. Elle donne l'exemple d'un système d'apprentissage de piano assisté par ordinateur. Ainsi, le professeur peut analyser non seulement les notes jouées, la position du pianiste... mais également l'état émotionnel de son élève. Ce professeur serait bien plus efficace en étant capable de déterminer si son élève prend du plaisir à jouer, ou si au contraire il présente des signes de stress. Il pourrait en effet modifier son enseignement en conséquence en accélérant ou ralentissant le rythme d'apprentissage ou en fournissant des conseils encourageants et personnalisés. Pour l'élève, il est également bien plus agréable d'avoir un professeur non seulement parfait techniquement, mais également compréhensif et patient [Picard 2000].

On peut citer aussi le système d'apprentissage STEVE (SOAR Training Expert for Virtual Environments) [Elliott 1999] qui intègre un agent virtuel qui permet d'aider les étudiants à apprendre à exécuter des tâches physiques basées sur des procédures. Il peut montrer comment réaliser des tâches et répondre à des questions et fournir de l'aide pour résoudre certains problèmes. L'agent STEVE est appelé à contrôler et à agir sur l'état émotionnel de l'apprenant. Pour ce faire, il dispose de 26 types d'émotions et de 70 expressions faciales.

Johnson [Johnson 2004] a développé un système de tuteurs intelligents STI nommé DARWAR destiné à l'apprentissage d'une langue étrangère. L'apprenant pratique les concepts linguistiques nouvellement appris dans un environnement virtuel. Tout au long de son apprentissage, un agent virtuel l'aide, le conseille et le soutient émotionnellement en le motivant. Tout récemment, Ben Amar [Ben Ammar 2007] a développé un STI appelé EMASPEL intégrant plusieurs agents conçus pour la gestion des émotions. Ce système est capable de détecter certains échecs relatifs à l'approche pédagogique adoptée et, surtout, de mettre en action une approche

alternative d'apprentissage mieux adaptée aux conditions émotionnelles spécifiques de l'apprenant.

1.3.3 Robotique pour l'assistance

Wada et al. [Wada 2006] ont montré que les robots thérapeutiques ont un fort potentiel pour améliorer l'état de l'activité cérébrale chez les patients souffrant de démence. Il ont montré aussi que l'interaction avec un robot rend les personnes âgées plus actives et plus communicatives [Wada 2008]. Actuellement de nouveaux types de robots sont utilisés pour l'aide psychologique et cognitives. Ces robots sociaux sont essentiellement dédiés à la médiation ou à l'accompagnement et permettent une meilleure intégration sociale, un réconfort moral et psychologique ainsi qu'un développement cognitif de son utilisateur. Le robot Paro¹, par exemple a été conçu pour apporter un soutien émotionnel aux patients. Il a été développé par Takanori Shibata et ses collaborateurs du National Institute of Advanced Industrial Sciences and Techniques. Ce robot bébé phoque réagit aux stimulations externes, comme la voix et le contact, avec des sons et de petits mouvements du corps. Paro a prouvé son efficacité puisqu'il réduit le stress des patients et de leurs soignants. Il stimule l'interaction entre les patients et les soignants, améliore la relaxation des patients, leur motivation et la socialisation.

Le projet EmotiRob [Saint-Aime 2007] a pour objectif de réaliser un robot compagnon pour des enfants fragilisés (par exemple, des enfants en hospitalisation longue). Il s'agit d'un robot-peluche qui permet de réagir aux comportements de l'enfant en simulant des émotions par les mouvements du corps, les traits du visage et par l'émission de petits sons simples.

1.3.4 Jeux Vidéo

L'évolution des technologies du jeu vidéo ouvre de nouvelles perspectives. En effet, le jeu vidéo passe progressivement du statut de loisir pour devenir un vrai moyen d'expression artistique et certains créateurs n'hésitent pas à explorer de nouvelles façons de nous faire vivre des expériences émotionnelles inédites. Ils intègrent de plus en plus la notion d'affect au sein de leurs jeux. De ce fait, les jeux vidéo deviennent plus complexes. Ils emploient des systèmes d'affichages immersifs et intègrent des personnages virtuels intelligents qui tiennent compte de l'état émotionnel du joueur. Ochs [Ochs 2008], dans ses travaux de recherche a travaillé sur la modélisation de la dynamique des relations sociales de personnages virtuels dans les jeux vidéo. Ses travaux de recherche étaient de tenter d'améliorer la crédibilité des personnages non-joueurs (i.e. des personnages virtuels non contrôlés par le joueur) afin d'améliorer le sentiment d'immersion du joueur dans un jeu. On peut citer également les travaux de Jones et Sutherland [Jones 2008] qui ont porté sur la détection explicite des émotions du joueur afin de modifier dynamiquement l'environnement. Par exemple, un jeu sur ordinateur peut s'adapter et réagir en ajoutant de nouveaux adversaires pour

1. <http://paro.jp/english/index.html>

stimuler le joueur lorsqu'il commence à se lasser (émotion détectée : l'ennui), où au contraire, diminuer la difficulté du jeu si le joueur est trop excité, stressé ou fatigué.

1.3.5 Systèmes de dialogues

Les émotions interviennent également dans les systèmes de dialogues. En effet, les émotions font partie intégrante de la communication dans le sens où elles ont la faculté d'agir sur la sémantique du message transmis. Plus particulièrement, elles font partie de l'environnement de transmission d'un message qu'il soit oral ou écrit, et influent sur la compréhension de l'interlocuteur. Parmi les systèmes de dialogues développés, on peut citer les travaux de Devillers et Vidrascu [Devillers 2006], qui ont étudié les réactions de frustration ou d'irritation dans des interactions orales enregistrées dans des centres d'appels (financier et médical). On peut citer également les travaux de Chavel [Clavel 2007] qui a développé une application de surveillance. Ils se sont intéressés à la reconnaissance des émotions dans la parole afin d'identifier des situations de menace pour la vie humaine.

1.3.6 Discussion

Nous avons vu dans le paragraphe précédent que l'informatique affective a pris de l'importance avec le développement des systèmes technologiques. Cependant, la représentation des états émotionnels d'un utilisateur ou les états d'émotion à simuler par une interface utilisateur exige un format de modélisation adapté. En effet, les systèmes émotionnels basés sur les interactions homme-machine ou machine-machine ont besoin d'un format de représentation bien défini pour pouvoir échanger les informations émotionnelles. Notre travail s'inscrit dans le cadre de modélisation et annotation des émotions. Il s'agit de proposer une approche pour la représentation, la modélisation et l'annotation des émotions qui permet d'effectuer les échanges émotionnels entre différentes applications et d'améliorer la crédibilité des interactions homme-machine. Ce travail propose par conséquent un système d'échange émotionnel qui va de la reconnaissance des émotions jusqu'à l'échange entre applications. Notre travail est divisé en trois parties distinctes :

- la modélisation des émotions,
- la reconnaissance des émotions
- l'annotation et l'échange émotionnel entre applications.

Dans le paragraphe suivant, nous allons survoler les travaux sur la reconnaissance des émotions et nous allons nous focaliser sur la reconnaissance des émotions à partir des signaux physiologiques.

1.4 Reconnaissance des émotions

La reconnaissance des émotions est un axe de recherche important dans le domaine de la neurophysiologie ainsi que dans le domaine de l'intelligence artificielle

pour améliorer les interfaces homme-machine. Généralement, l'expérience émotionnelle déclenche des signaux moteurs et physiologiques (variations du pouls, augmentation de la température corporelle, etc). Ces signaux ont pour but de communiquer l'émotion ressentie. Donc, notre corps permet de transmettre de l'information par plusieurs canaux :

- les expressions faciales
- les intonations de la voix (la prosodie)
- les réactions du Système Nerveux Autonome (les signaux physiologiques)
- les positions et mouvements du corps

Dans le domaine de l'informatique, les émotions de l'utilisateur peuvent être reconnues à partir de capteurs comme les caméras, microphones, capteurs physiologiques. Dans les paragraphes qui suivent nous présentons la détection des émotions à partir des différentes modalités (expressions faciales, la voix, les signaux physiologiques,...). Dans la littérature, une modalité est simplement assimilée à un média, parfois appelée canal [Snoek 2003].

1.4.1 Les expressions faciales

Les expressions faciales constituent l'élément le plus important dans le processus de communication [Smith 1997] et elles présentent une source importante d'information concernant l'état émotionnel des personnes durant une interaction sociale. En effet, lors de l'expérience d'une émotion par un individu, des muscles spécifiques sont activés permettant par exemple de sourire ou de froncer les sourcils. L'expression faciale peut être mesurée par l'électromyogramme. Cette technique consiste à mesurer directement l'activité électrique des muscles faciaux par électrodes appliquées sur la surface cutanée et donne accès aux modifications latentes de l'activité faciale non visible. Ekman et Friesen à la fin des années 1970 ont développé des méthodes de mesure des comportements du visage [Ekman 1978]. En particulier, ils ont créé Le Facial Action Coding System (FACS) qui permet le codage de toutes les unités d'actions (ou Action Unit, AU) visibles sur un visage. Chaque AU peut correspondre à la contraction ou à la détente d'un ou plusieurs muscles qui se traduit par le mouvement d'une partie donnée du visage. L'expression d'anxiété correspond donc à la mise en jeu de plusieurs unités d'actions, par exemple : 1, lever des sourcils intérieurs (Inner Brow Raiser), 2, lever des sourcils extérieurs, 20, tension des lèvres et 26, abaissement de la mâchoire inférieure (cf. FIGURE 1.6).

De nombreuses études ont été effectuées en informatique sur la reconnaissance des émotions à partir des expressions faciales. En effet, le visage est le canal communiquant le plus d'informations émotionnelles. En général, on distingue trois étapes principales dans un système d'analyse d'expression faciale. La première étape permet de détecter le visage et limiter la zone de recherche. La deuxième étape consiste à extraire les informations nécessaires pour décrire au mieux l'expression. La dernière étape permet de classer l'image, c'est à dire l'affecter à une catégorie d'expressions.

Hammal et al. [Hammal 2007] proposent une méthode de classification basée sur les distances entre points d'intérêt du visage. Abboud et al. [Abboud 2004] proposent

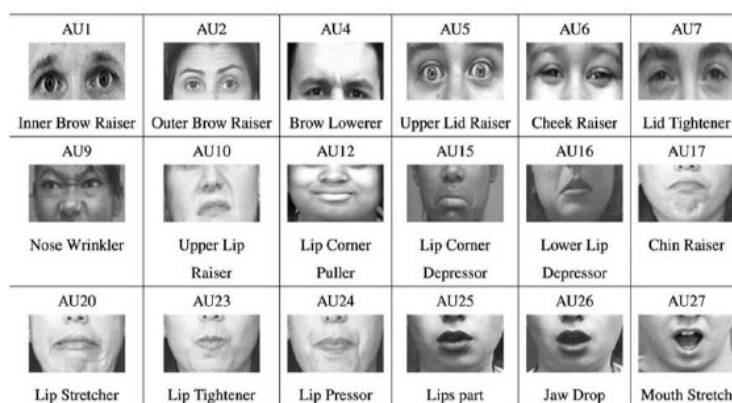


FIGURE 1.6 – Liste des unités d'action relatives aux 6 expressions faciales [Zhang 2005]

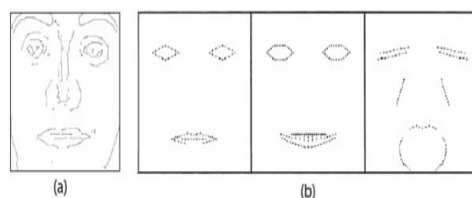


FIGURE 1.7 – (a) LEM du visage, (b) le modèle de l'expression faciale [Gao 2003]

un modèle automatique pour la classification des expressions faciales en se basant sur les Modèles d'Apparence Active (MAA). Ils ont représenté chaque image de visage par un vecteur MAA correspondant. Ensuite, ils mesurent la distance entre le vecteur MAA et le vecteur moyen dans l'espace de Fisher en utilisant la distance de Mahalanobis [Lauter 1985]. Pour chaque configuration, le visage testé est assigné à la classe qui a la moyenne la plus proche.

Gao [Gao 2003] propose une méthode de classification de trois expressions : neutre, sourire et colère. Son approche utilise la ligne du contour (LEM : Line Edge Map) [Gao 2002] comme une expression descriptive (cf. FIGURE 1.7-a). Ensuite il calcule la distance de Hausdorff dans la ligne directe du segment (la mesure de disparité définie entre 2 ensembles de lignes) entre le visage courant LEM et le modèle de la caricature de l'expression (cf. FIGURE 1.7-b).

1.4.2 Reconnaissance acoustique des émotions

Les émotions sont aussi communiquées à travers la voix. Certaines caractéristiques de la voix comme l'intonation, l'intensité, etc traduisent une émotion. Par exemple, la tristesse correspond à un débit de parole lent, une tonalité basse, une intensité faible et une articulation bredouillée [Picard 1997b]. Les premières études sur la parole émotionnelle se sont basées sur une analyse de la prosodie. La prosodie est un objet complexe de la parole qui regroupe divers phénomènes tels que le ton,

l'accent, la mélodie, le rythme ou encore le débit. Elle est décrite par trois paramètres acoustiques : la fréquence fondamentale, l'intensité et la durée [Dellaert 1996].

- La fréquence fondamentale est définie comme la fréquence de vibration des plis vocaux à un instant donné.
- L'intensité correspond à l'énergie présente dans le signal, c'est à dire la force avec laquelle un son est émis.
- La durée représente l'intervalle de temps nécessaire pour émettre un signal (nombre de syllabes réalisées par seconde).

Depuis des années, plusieurs études abordent le problème de la reconnaissance des émotions dans la parole. Un système de reconnaissance automatique des émotions repose classiquement sur quatre phases principales : l'extraction de descripteurs acoustiques, l'apprentissage, la classification et enfin la décision [Clavel 2010].

- L'extraction de descripteurs acoustiques : il s'agit d'analyser le signal de parole pour le transformer en une séquence de vecteurs acoustiques contenant les valeurs des différents descripteurs ou paramètres retenus. Cette étape permet d'obtenir une représentation compacte des principales caractéristiques acoustiques du signal de parole. L'intensité du signal de parole ou la fréquence centrale des formats sont deux exemples de tels descripteurs acoustiques
- L'apprentissage : permet de regrouper les vecteurs acoustiques correspondant aux sons d'une même classe dans un représentant ou un modèle caractéristique de cette classe. Ce modèle est traditionnellement obtenu à partir d'une base de données (dite d'apprentissage).
- La classification : il s'agit de comparer les vecteurs acoustiques du signal vocal à analyser aux représentants ou modèles de chaque classe.
- La décision : permet d'associer une classe à un segment de parole.

Plusieurs travaux se sont intéressés à la reconnaissance acoustique des émotions. Nous citons par exemple Istrate [Istrate 2003] qui s'est intéressé à la reconnaissance des émotions à partir des sons pour l'aide aux personnes âgées ou aux personnes hospitalisées. Il a proposé un système d'analyse sonore et des solutions aux problèmes spécifiques au traitement du son pour la télésurveillance médicale. Claval, dans sa thèse [Clavel 2007] a développé un système automatique de reconnaissance d'émotions de type peur en situation anormale en se basant sur les manifestations vocales de l'émotion. Attabi [Yazid 2008] s'est focalisé sur la détection automatique des appels problématiques dans un contexte réel de centres d'appels téléphoniques. Il a utilisé l'information sur l'état émotionnel du locuteur présente dans le signal acoustique pour détecter les problèmes de compréhension entre un locuteur et un système de dialogue homme-machine. Nous pouvons citer également les travaux de Lee [Lee 2003] qui consiste à la reconnaissance de sept émotions (ennui, dégoût, joie, colère, neutre, panique, tristesse). Il a utilisé les statistiques de la prosodie (la fréquence fondamentale, l'énergie, la durée). Il a utilisé deux types de classificateurs : le classificateur linéaire (LD) et les k-nearest neighbor (KNN). Chacun des classificateurs est entraîné séparément selon le sexe du locuteur (homme ou femme). Les meilleurs résultats ont été obtenus avec le classificateur LD pour les données des hommes (47,04%) et avec le classificateur KNN pour les données des femmes

(47.14%).

1.4.3 Reconnaissance par les signaux physiologiques

Différentes activations physiologiques sont associées aux états émotionnels. Elles concernent des modifications au niveau du Système Nerveux Autonome SNA (modification du rythme cardiaque, de la pression artérielle, de la température corporelle, etc). Ces modifications varient en fonction de l'émotion considérée et des sujets. Ceci induit un patron de réponse complexe permettant de distinguer les différentes émotions. Cependant il est très difficile de trouver un lien univoque et systématique entre une émotion donnée et une activation physiologique caractéristique. En effet, on peut observer qu'un grand nombre de ces modifications sont communes à plusieurs émotions. Par exemple, l'augmentation du rythme cardiaque semble identique dans la colère, la peur et la tristesse. Cependant, dans le cas de la colère, elle est associée à une forte augmentation de la température cutanée, ce qui n'est pas le cas dans la peur ou la tristesse (aucune modification dans le premier cas et une diminution dans le second). En outre, il convient de remarquer que des mesures plus précises du rythme cardiaque ont permis de mettre en évidence des variations subtiles du patron de la rythmicité cardiaque entre la peur et la colère [ABDAT 2010]. Sinha et al. [R. 1996] ont montré que la peur, la tristesse et la colère sont associées à une augmentation du rythme cardiaque plus que la joie. Le dégoût qui diminue le rythme cardiaque est associé à une baisse de la température cutanée. Aussi la conductance de la peau augmente après un état d'amusement et diminue après l'état neutre et reste la même après une tristesse. Les indices physiologiques qui sont couramment utilisés pour déterminer l'état émotionnel d'une personne sont :

- la réponse électrodermale RED (galvanic skin response (GSR)) ;
- le volume sanguin impulsionnel (Blood Volume Pulse BVP) ;
- le signal du volume respiratoire (VR) ;
- l'activité électromyographique (EMG) ;
- la température cutanée (Skin Temperature SKT) ;
- la fréquence cardiaque (Fc) ;
- le rythme électroencéphalogramme (EEG).

Dans notre travail, nous nous focalisons sur la détection des émotions en utilisant les quatre premiers signaux physiologiques, c'est pour cela que nous allons bien les détailler dans la suite de l'exposé.

1.4.3.1 Activité électrodermale (GSR)

L'activité électrodermale GSR (également appelée conductivité de la peau) est une donnée physiologique enregistrée à la surface de la peau et reflète le niveau de conductibilité électrique de la peau. Cette activité électrique de la peau varie très sensiblement dans les situations mettant en jeu les émotions. Cette mesure peut être prise sur la paume des mains ou bien sur la plante des pieds [Stemmler 2001]. Plusieurs travaux (Collet et al. [Collet 1997], Picard et al. [Picard 2001]) ont démontré



FIGURE 1.8 – capteur de respiration [Williams 2004]

qu'il est possible de différencier les émotions de base (joie, colère, tristesse, surprise, dégoût et peur) par les mesures physiologiques, notamment de la GSR, qui peut fournir une mesure indirecte de l'intensité émotionnelle.

1.4.3.2 Pression sanguine volumique (BVP)

La pression sanguine volumique est un indicateur de l'écoulement du sang à travers le corps humain. Elle permet de mesurer l'activation physique et le changement du rythme cardiaque. Le BVP diminue sous l'effet du stress et l'effort. Il est utilisé comme indicateur de la peur [Levenson 1992] et de la colère [daniel kahneman 1973, Levenson 1992]. La pression sanguine volumique se mesure grâce à un capteur appelé photoplethysmographe qui s'attache au pouce du sujet. Il mesure l'impulsion cardio-vasculaire à travers le corps humain. On peut en déduire la fréquence cardiaque (HR).

1.4.3.3 Volume et rythme respiratoire (RESP)

Le rythme respiratoire est défini par l'alternance régulière des mouvements d'inspiration et d'expiration, où le volume de la cage thoracique augmente à chaque pénétration de l'air (ou inspiration) et diminue à chaque rejet (ou expiration). Un état de repos est caractérisé par une respiration plus lente. Par contre, un état d'excitation émotionnelle provoque une respiration plus profonde. Aussi les émotions à valence négative causent généralement des respirations irrégulières [ABDAT 2010]. La respiration se mesure à l'aide d'un capteur qui se place au-dessus du sternum pour la surveillance thoracique. Il est constitué d'une ceinture Velcro entourant le torse et d'un plétismographe qui s'étend lorsque le sujet expire (cf. FIGURE 1.10).

1.4.3.4 L'activité électromyographique (EMG)

L'activité électromyographique représente l'enregistrement des courants électriques qui accompagnent l'activité musculaire. Cette mesure permet de déterminer l'état émotionnel du sujet. L'électromyographie s'effectue en utilisant des électrodes qui sont positionnées à la surface du corps, constituées de fines aiguilles (le plus souvent) que l'on enfonce dans le muscle que l'on désire étudier. Cette technique permet



FIGURE 1.9 – capteur EMG appliqué sur le muscle de la mâchoire [Williams 2004]

de détecter l'activité musculaire spontanément, au repos ou au moment de l'activité du muscle, ou bien lorsque l'on procède à la stimulation électrique (galvanique) musculaire nerveuse dans le cadre d'une épreuve ou d'un examen neurologique. L'EMG a été utilisé pour étudier les expressions faciales [Ekman 2006], les expressions gestuelles [Marrin 1998] et le concept de valence émotionnelle [Lang 1995].

1.4.4 La reconnaissance des émotions à partir des signaux physiologiques

La riche littérature existante dans le domaine de la reconnaissance des émotions à partir des signaux physiologiques permet de confirmer que ces derniers sont pertinents pour déterminer les émotions ressenties. Plusieurs travaux ont été réalisés dans ce domaine. Cependant, il est difficile de faire une étude comparative entre ces travaux, dû au fait qu'ils diffèrent dans la façon d'induire l'émotion et dans le choix de celles traitées. Par la suite, nous allons essayer de résumer quelques travaux dans le domaine.

Wagner et al. [Wagner 2005] se sont intéressés à mesurer les manifestations physiologiques (le rythme cardiaque, la respiration, l'électromyographie (EMG) et la réponse électromodale de la peau), et ensuite à les corrélérer avec les deux dimensions de valence et d'excitation.

Vrana et al. [Vrana 1986] ont opté pour l'image et la répétition d'un texte dans le silence pour induire la peur et la neutralité. Ils ont traité le rythme cardiaque avec le test de Newman-Keuls et la méthode d'ANOVA (ANalysis Of VAriance). Les auteurs ont trouvé que le rythme cardiaque s'accélère et s'élève durant l'induction de la peur par rapport à l'induction de l'état neutre.

Lisetti et Villon [Villon 2006] ont proposé le modèle Cartes Psycho-Physiologiques Émotionnelles (Psycho Physiological Emotion Map PPEM) qui constitue une représentation paramétrique permettant l'interprétation émotionnelle à partir des signaux physiologiques. Leur technique vise à interpréter psychologiquement les paramètres physiologiques (conductance cutanée et battements de cœur) et à produire une extraction continue de l'état affectif de l'utilisateur durant une interaction homme-machine.

Scheirer et al. [Scheirer 2002] ont utilisé une interface d'un jeu d'ordinateur lent

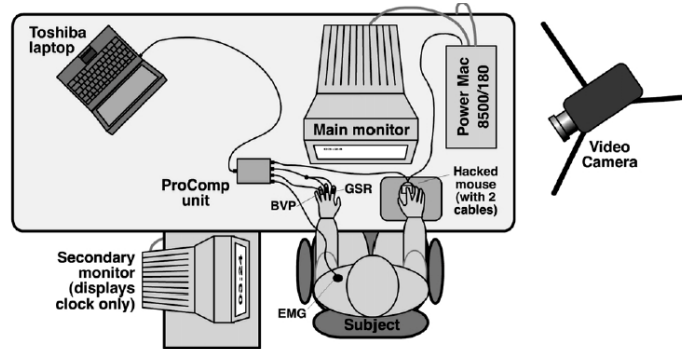


FIGURE 1.10 – L’environnement de l’acquisition utilisé dans [Scheirer 2002]

afin de provoquer la frustration à 24 personnes. Ils ont utilisé la conductivité de la peau et la pression du volume sanguin avec les chaînes de Markov cachées. Les résultats obtenus dépassent 50% pour 21 sujets parmi les 24.

Nasoz et al. [Nasoz 2003] ont développé un système multimodal qui reconnaît l’état émotionnel de l’utilisateur à partir de ses signaux physiologiques. Ils ont utilisé des films pour induire l’émotion d’un participant avant de la mesurer. Ensuite, ils ont utilisé l’algorithme des k plus proches voisins (KNN) et l’analyse factorielle discriminante (DFA : Discriminant Function Analysis) pour la reconnaissance des émotions. Pour les émotions frustration et surprise, le premier algorithme (KNN) a donné de meilleurs résultats, tandis que pour les autres émotions (peur, tristesse, colère et amusement) le deuxième algorithme (DFA) a donné de meilleurs résultats.

Deux autres travaux sont souvent cités dans la littérature. Le premier est celui de Haag et ses collègues [Williams 2004] qui ont utilisé des photos pour provoquer les émotions des participants. Ils se sont intéressés à mesurer les manifestations physiologiques (la température, la respiration, l’EMG, la conductance cutanée, le BVP, l’ECG (l’activité musculaire du cœur). Pour la reconnaissance, Haag et ses collaborateurs ont utilisé l’algorithme des réseaux de neurones artificiels. Le deuxième est celui de Healey et Picard [Healey 2000] qui se sont intéressés à détecter le niveau de stress lors de la conduite d’une voiture. Ils ont mesuré quatre signaux physiologiques (GSR "Galvanic Skin Response", respiration, électromyogramme, et battement de cœur) pour un acteur entrain de voir des images provoquant des émotions. Les émotions induites sont : pas d’émotion (neutre), la colère, la haine, la douleur, l’amour platonique, l’amour romantique, la joie et le respect. L’acteur n’exprime pas seulement chaque émotion de l’extérieur, mais il doit vivre la situation de chaque émotion de l’interne. Les expériences incluent des séances de 25 minutes quotidiennes sur une période de 20 jours qui sont issus de 4 capteurs (électrocardiogramme, électromyogramme EMG, résistance de la peau et respiration). Les signaux physiologiques ont été traités à l’aide d’une recherche séquentielle en avant (Sequential Floating Forward Search :SFFS), la projection de Fisher (Fisher projection : FP) et avec une méthode hybride combinant les deux méthodes (SFFS-FP). La classification hybride

a donné le meilleur résultat, 81% pour les huit catégories d'émotions. Prendinger et al. [Prendinger 2004] ont utilisé les signaux physiologiques pour déterminer l'état émotionnel d'un utilisateur afin de mieux l'orienter. Ils ont implémenté un agent intelligent qui suggère des actions appropriées à l'utilisateur, en se basant sur ses réactions émotionnelles. Il peut par exemple encourager l'utilisateur si ce dernier n'est pas motivé (faible activation "arousal") ou bien le féliciter si l'utilisateur dégage une forte activation positive. Un autre exemple est le travail de Conati et al. [Conati 2003] dans lequel ils ont utilisé les signaux physiologiques pour détecter l'état émotionnel d'un étudiant lors d'un jeu éducatif. Ils ont utilisé le réseau dynamique de décision (DDN : Dynamic Decision Network) pour combiner la cause de l'émotion et son effet sur l'utilisateur pour diminuer l'incertitude de leur modèle.

Chanel et al. [Chanel 2007] ont proposé un système pour la reconnaissance de trois états émotionnels positif, calme et négatif à partir des signaux physiologiques (la pression sanguine, la fréquence cardiaque, la respiration, la réponse galvanique de la peau et les signaux EEG). Ils ont demandé aux participants de vivre deux situations déjà vécues (une situation joyeuse et une situation douloureuse) pour induire les émotions positives et négatives. Pour l'état neutre, ils ont demandé aux participants de se relaxer et de se calmer. Les signaux physiologiques ont été traités à l'aide d'un filtre moyen en changeant la fenêtre pour chaque signal (512 pour RED, 128 pour la pression sanguine et 256 pour la respiration) afin d'enlever le bruit. Leurs résultats montrent l'efficacité du signal EEG par rapport aux autres signaux. Les auteurs ont obtenu un taux de 67% pour trois classes (neutre, positive, négative) et un taux de 79% pour deux classes avec les SVM.

Kim et al. [Kim 2002] ont proposé un système de reconnaissance des émotions basé sur les signaux physiologiques. Ils ont utilisé les signaux physiologiques pour la reconnaissance de quatre émotions (tristesse, colère, stress et la surprise). Les auteurs ont utilisé un protocole d'induction composé de trois aspects : l'audition, la vision et la cognition. Par exemple, pour induire la tristesse, ils utilisent une lumière bleue associée à une musique triste en racontant une histoire triste par un acteur qui joue sur sa voix. Le résultat obtenu est de 61.76% pour 4 émotions.

1.4.5 Classification automatique des émotions

Un système de reconnaissance des émotions comprend une phase de classification qui consiste à assigner des segments de signaux émotionnels à des catégories ou classes d'émotions i.e. colère, joie, tristesse, etc. En général, le système de classification automatique comporte deux étapes :

- La phase d'apprentissage : qui consiste à estimer la description de l'espace des observations qui traduit le mieux l'association avec les classes correspondantes. Il s'agit de créer un modèle caractéristique de chaque classe
- La phase de test : qui permet d'évaluer les performances du système de classification. Elle consiste à associer une classe d'émotion au vecteur de caractéristiques à l'aide d'une fonction de décision.

Pour pouvoir effectuer une classification et traduire les signaux acquis vers une émotion spécifique, une étape d'extraction des caractéristiques les plus pertinentes doit être faite. Cette étape permet de choisir les principaux critères ou paramètres caractéristiques qui doivent être pris en compte dans l'étape de classification. Parmi ces méthodes de sélection nous pouvons citer la recherche séquentielle en avant (Sequential Floating Forward Selection SFFS) et la discriminante linéaire de Fisher qui ont été utilisées par Picard et al. [Healey 2000, Picard 2001] pour réduire la dimensionnalité des données tout en conservant le maximum d'information possible.

Sequential Floating Forward Selection SFFS :

C'est une méthode de sélection de caractéristiques. Elle consiste à utiliser l fois l'algorithme SFS de manière à ajouter l variables, puis à utiliser r fois l'algorithme SBS afin d'en supprimer r . Ces étapes sont alors répétées jusqu'à l'obtention du critère d'arrêt.

Discriminante linéaire de Fisher ADF :

La technique de Fisher permet une réduction du nombre de dimensions avec une faible augmentation de la probabilité d'erreur. Le principe consiste à trouver la direction W qui discrimine au mieux les classes en projection. Le critère de Fisher J permet de maximiser le ratio de séparation des deux classes par rapport à leurs dispersions.

Une fois l'extraction/sélection des attributs achevée, il reste à effectuer la classification. Plusieurs systèmes de classification ont été développés : plus proches voisins (KNN), machines à vecteurs de support (SVM), arbres de décision, réseaux bayesiens,...

1.4.5.1 K plus proches voisins KNN (K Nearest Neighbor)

La méthode KNN est une méthode d'apprentissage supervisé. Pour classer une nouvelle observation x , l'idée consiste à faire voter les plus proches voisins de cette observation. La classe de x est déterminée en fonction de la classe majoritaire parmi les k plus proches voisins de l'observation x .

Algorithme 1-NN

La méthode du plus proche voisin est une méthode à base de voisinage et non-paramétrique [Mathieu-Dupas 2010]. Une nouvelle observation est classée dans la classe d'appartenance de l'observation de l'échantillon d'apprentissage qui lui est la plus proche. Le calcul de leur similarité est basé sur des mesures de distance.

Formellement, soit L l'ensemble des données d'apprentissage :

$$L \{(y_i, x_i), i = 1, \dots, n_l\} \quad (1.1)$$

où $y_i \in (1, \dots, c)$ dénote la classe de l'individu i et le vecteur $x_i = (x_{i1}, \dots, x_{ip})$ représente les variables prédictrices de l'individu i . Pour déterminer le plus proche voisin il faut se baser sur une fonction distance $d(., .)$. La distance la plus populaire

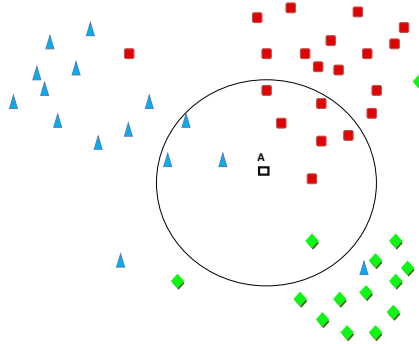


FIGURE 1.11 – Exemple de classification par la méthode des k plus proches voisins : 3 classes, $k = 10$

est la distance euclidienne.

Ainsi pour une nouvelle observation (y, x) , le plus proche voisin $(y_{(1)}, x_{(1)})$ dans les données d'apprentissage est déterminé par :

$$d(x, x_{(1)}) = \min_i (d(x, x_i)) \quad (1.2)$$

(avec $x_{(1)}$ représente le plus proche voisin de x et $y_{(1)}$ représente sa classe d'appartenance) finalement $\hat{y} = y_{(1)}$, la classe du proche voisin sera sélectionnée pour la prédiction de la classe y .

Algorithme KNN

La méthode des K plus proches voisins sélectionne les k plus proches observations. Ainsi la décision est en faveur de la classe majoritairement représentée par les k voisins. Le paramètre k doit être déterminé par l'utilisateur : $k \in \mathbb{N}$. Le meilleur choix de k dépend du jeu de données. En général, les grandes valeurs de k produisent un lissage qui réduit l'effet du bruit sur la classification et donc le risque de sur-apprentissage, mais rendent les frontières entre classes moins distinctes. Donc il faut faire un choix de compromis entre la variabilité associée à une faible valeur de k contre un "oversmoothing" ou surlissage pour une forte valeur de k .

1.4.5.2 Support Vector Machines : SVM

La méthode des Support Vector Machines est une méthode de classification binaire par apprentissage supervisé, introduite par Vapnik en 1995. Cette méthode repose sur l'existence d'un classificateur linéaire dans un espace approprié. Elle consiste à estimer un hyperplan qui sépare les deux classes, en garantissant que la marge entre le plus proche des deux classes soit maximale.

La marge est la distance entre la frontière de séparation et les échantillons les plus proches appelés vecteurs support. Dans un problème linéairement séparable les SVM trouvent une séparatrice qui maximise cette marge. Dans le cas d'un problème non linéaire on utilise une fonction noyau pour projeter les données dans un espace

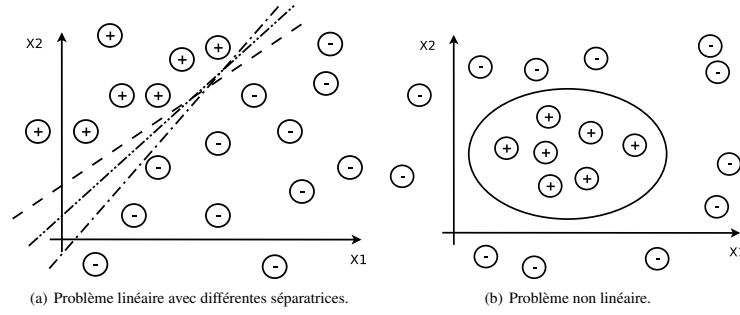


FIGURE 1.12 – Exemple de classification binaire linéaire et non-linéaire dans $\mathbb{R}^2$ [Mercier 2010]

de plus grande dimension où elles seront linéairement séparables (cf. FIGURE 1.12).

Dans le cas où les données sont linéairement séparables, étant donné un ensemble des exemples $\{(x_1, y_1), \dots, (x_n, y_n)\}$, avec x_i les vecteurs de caractéristiques de dimension D , $y_i \in \{-1, +1\}$, les étiquettes dans un problème à deux classes, $y_i = -1$ si l'exemple est négatif et $y_i = +1$ si l'exemple est positif. L'algorithme estime l'hyperplan optimal permettant de maximiser la marge entre les observations correspondant aux deux classes. L'hyperplan est donné par la solution du problème d'optimisation :

$$\min \frac{1}{2} \|w\|^2 \cdot w \quad (1.3)$$

avec les contraintes suivantes :

$$\forall i, y_i(w \cdot x_i + b) \geq 1 \quad (1.4)$$

L'équation 1.3 est résolue en introduisant les multiplicateurs de Lagrange $(\alpha_i)_{1 \leq i \leq l}$ et un lagrangien [Mahdhaoui 2010]. L'équation précédente s'écrit alors sous la forme suivante :

$$L(w, b, \alpha_i) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^l \alpha_i [y_i (w \cdot x_i + b) - 1] \quad (1.5)$$

Finalement la fonction de décision est définie par

$$f(x) = w \cdot x + b = \sum_{i=1}^m \alpha_i y_i (x \cdot x_i) + b \quad (1.6)$$

Cependant, dans le cas où les données ne sont pas linéairement séparables, les contraintes deviennent alors :

$$y_i(w \cdot x_i + b) \geq 1 - \xi_i \quad (1.7)$$

Pour résoudre ce problème, on change la fonction d'optimisation :

$$\frac{1}{2} \|w\|^2 + C \sum_{i=1}^l \xi_i \quad (1.8)$$

1.4.5.3 Discussion

La méthode des k plus proches voisins est une technique d'apprentissage statistique. Donc l'estimation de probabilité à posteriori permet d'intégrer et de combiner le classifieur KNN dans une méthodologie statistique. En effet, un modèle statistique peut être dérivé du classifieur KNN en considérant les classes d'appartenance des voisins. Les probabilités locales qui sont basées sur les classes des voisins sont définies par :

$$P_{knn}(C_n \setminus x) = \frac{k_n}{k} \quad (1.9)$$

où x représente le vecteur de caractéristiques de la base de test et k_n le nombre de voisins de la classe C_n parmi les k voisins les plus proches.

Le SVM standard ne permet pas d'aboutir directement à des probabilités à posteriori de classification. Cependant pour garder une approche statistique, la sortie du SVM doit être représentée sous forme de probabilité. La toolbox LIBSVM [Lib 2001] propose une méthode de calcul de ces probabilités à partir des paramètres obtenus par SVM. Ainsi les probabilités à posteriori sont déterminées à l'aide d'une fonction paramétrique sigmoïde.

1.5 Représentation et échange des émotions

1.5.1 Introduction

Les émotions sont une composante essentielle dans la communication humaine. Elles représentent de plus en plus un élément crucial pour le bon fonctionnement des systèmes interactifs homme-machine/ machine-machine. L'étude des émotions soulève plusieurs questions comme : Comment représenter et annoter de manière fiable les émotions ? A quels niveaux d'abstraction ? Comment annoter les émotions complexes ? Comment échanger les émotions entre applications ? Pour apporter une réponse à ces questions, de nombreuses recherches sont conduites dans le domaine des sciences affectives visant à modéliser les émotions pour une utilisation dans des contextes technologiques. Nous présentons dans ce qui suit quelques langages de représentation et d'annotation des émotions.

1.5.2 Emotion Markup Language : EmotionML

Emotion Markup Language (ou EmotionML) est un langage de balise pour décrire les émotions humaines. Il a été mis en place par le World Wide Web Consortium (W3C) en juillet 2010 [Baggia 2010]. Il permet d'exprimer les émotions via le langage XML. En effet, EmotionML permet de représenter et traiter les données émotionnelles en utilisant la technologie XML ce qui permet l'interopérabilité entre les différentes plateformes envisagées. Ce langage permet de faire une annotation des différents états émotionnels. Il est aussi capable de représenter plusieurs types d'émotions, non seulement les émotions basiques mais aussi les émotions secondaires. Ce langage est basé sur des balises xml permettant de classer

les émotions suivant les théories catégorielles, les théories dimensionnelles ou encore la théorie de l'évaluation. Ainsi, en utilisant la théorie catégorielle l'émotion « satisfaction » sera annotée avec EmotionML par : (sachant que l'émotion « satisfaction » doit figurée dans le vocabulaire des catégories des émotions présent dans : <http://www.example.com/category/everyday-emotions>).

```
<emotion category-set="http://www.example.com/emotion/category/
everyday-emotions.xml">
  <category name="satisfaction"/>
</emotion>
```

Le langage EmotionML suppose que l'excitation et la puissance sont des échelles unipolaires avec des valeurs comprises entre 0 et 1, et la valence est une échelle bipolaire avec des valeurs typiques comprises entre -1 et 1. Le code ci dessous montre un exemple d'une émotion présenté à l'aide de l'approche multi-dimensionnelle (combinaison de la valence, de l'excitation et de la puissance).

```
<emotion dimension-set="http://www.example.com/emotion/dimension/PAD.xml">
  <dimension name="arousal" value="0.3"/><!-- lower-than-average arousal ->
  <dimension name="pleasure" value="0.9"/><!-- very high positive valence ->
  <dimension name="dominance" value="0.8"/><!-- relatively high potency ->
</emotion>
```

L'exemple précédent présente un état d'excitation assez faible, de plaisir très élevé et de forte puissance. Autrement dit un état émotionnel positif avec contrôle de situation.

Le langage EmotionML prend en charge les différents types de modalités et permet également la représentation de l'intensité des émotions. En effet, il utilise la balise `<modality>` pour l'annotation des modalités et la balise `<intensity>` pour exprimer l'intensité à laquelle est attribuée une certaine confiance. Ces différentes balises offrent ainsi un langage générique, permettant d'annoter les émotions suivant les principales théories et de spécifier les modalités d'expression de ces émotions. Dans l'exemple ci dessous, l'émotion « faible surprise » a été reconnue à l'aide des expressions faciales.

```
<emotion category-set="http://www.example.com/emotion/category/everyday-
emotions.xml"> <modality="face">
  <intensity value="0.2"/>
  <category name="surprise"/>
</emotion>
```

EmotionML permet également de couvrir l'aspect temporaire en fournissant des informations sur le début et la fin du déclenchement d'une émotion et sa variation au cours du temps [Baggia 2010].

Tout document EmotionML doit commencer par l'élément `<emotionml>`, qui est l'élément racine du document. L'élément `<emotionml>` doit utiliser le nom de

domaine : EmotionML namespace : "http://www.w3.org/2009/10/emotionml".

Exemple :

```
<emotionml version="1.0" xmlns="http://www.w3.org/2009/10/emotionml">
...
</emotionml>
```

Annotation d'une seule émotion : l'élément `<emotion>`

L'élément `<emotion>` permet d'annoter une seule émotion qui peut être simple ou complexe. Pour annoter plusieurs émotions dans un contexte donné, il est nécessaire d'utiliser plusieurs éléments `<emotion>`. Ci dessous un exemple d'annotation d'un extrait du livre "Les Aventures d'Alice au pays des merveilles".

```
< emotionml xmlns="http://www.w3.org/2009/10/emotionml"
xmlns:meta="http://www.example.com/metadata"
category-set="http://www.w3.org/TR/emotion-voc/xml#everyday-categories">
<info>
<meta :doc>Example adapted from (Zhang, Black & Sproat 2003)
http://www.cs.cmu.edu/~awb/papers/eurospeech2003/esper.pdf
</meta :doc>
</info>
<emotion>
<category name="Disgust" value="0.82"/>
"Come, there's no use in crying like that!"
</emotion>
said Alice to herself rather sharply;
<emotion>
<category name="Anger" value="0.57"/>
"I advise you to leave off this minute!"
</emotion>
</emotionml>
```

L'élément `<emotion>` peut avoir un attribut *id* qui sert à référencer une annotation de manière unique. Les attributs *start*, *end*, *duration*, *etc* donnent des informations sur l'heure à laquelle s'est produite l'émotion et l'attribut *expressed-through* permet de préciser la ou les modalités à travers lesquelles l'émotion a été exprimée. L'élément `<emotion>` doit contenir au moins un élément enfant. Les éléments enfants sont détaillés ci dessous :

L'élément `<category>` : permet la description d'une émotion dans le modèle catégoriel. L'attribut "name" qui permet d'annoter le nom de l'émotion doit être défini dans le vocabulaire déclaré pour la représentation des émotions. Il existe quatre types de vocabulaire : "category vocabularies" pour une représentation des

émotions basées sur l'approche catégorielle, "dimension vocabularies" pour une représentation dimensionnelle, "appraisal vocabularies" pour une représentation basée sur les modèles de l'évaluation cognitive et le "action tendency vocabularies".

L'élément <dimension> : permet la description d'une émotion dans le modèle dimensionnel. L'attribut "name" décrit le nom de la dimension qui doit être déclaré dans le vocabulaire dimensionnel (dimension vocabularies).

L'élément <appraisal> : permet la description d'une émotion dans les modèles de l'évaluation cognitive.

L'élément <link> : permet de donner des informations sur le contenu annoté comme l'emplacement, le nom du fichier, etc.

1.5.3 Langage EARL (Emotion Annotation and Representation Language)

Le langage de représentation et d'annotation des émotions EARL [The HUMAINE Association 2006] a été proposé pour permettre l'échange des données émotionnelles et la réutilisation des ressources. C'est un langage pour la représentation des émotions basé sur XML. C'est une première tentative qui a été effectuée par le réseau d'excellence HUMAINE [The HUMAINE Association 2006] pour annoter les états émotionnels. Ce langage utilise une variété de représentations pour décrire les émotions. On trouve la description catégorielle, la description dimensionnelle et la description évaluation cognitive ce qui permet à l'utilisateur de choisir la représentation convenable à son approche. La possibilité de relier une représentation à une autre rend le format utilisable dans des environnements hétérogènes où il existe une variété de représentations des émotions.

Représentation catégorielle : c'est la représentation la plus simple, elle utilise un mot pour décrire un état émotionnel.

Exemple :

```
<emotion category="pleasure">Hello!</emotion>
```

Description dimensionnelle : permet de représenter les émotions dans l'espace dimensionnel (valence/arousal)

Exemple :

```
<emotion xlink:href="face.jpg" arousal="-0.2" valence="0.5" />
```

Le langage EARL permet également de représenter des émotions complexes. En effet, simultanément, il permet d'annoter deux ou plusieurs émotions qui se produisent ensemble. Il permet également d'annoter la variation des émotions dans le temps et la représentation multimodale des émotions.

L'exemple ci dessous montre l'annotation de deux émotions détectées par deux modalités différentes.

```
<complex-emotion xlink:href="clip23.avi">
<emotion category="peur" modality="face"/>
<emotion category="tristesse" modality="voice"/>
</complex-emotion>
```

EARL permet aussi d'annoter la durée de l'émotion en précisant l'heure de début et de fin pour les enregistrements vocaux ou vidéos.

Exemple :

```
<emotion start="0.4" end="1.3" category="pleasure" />
```

1.5.4 Langage EMMA (Extensible MultiModal Annotation markup language)

EMMA (Extensible MultiModal Annotation Markup Language) est un langage de balises pour l'annotation multimodale. Il fait partie des normes du W3C pour les interactions multimodales [W3C 2009]. Il est utilisé par les systèmes pour faire la représentation sémantique des variétés des informations récupérées en entrée (la parole, les gestes, link...) et pour les intégrer au sein d'une application multimodale. Voici un exemple d'annotation avec EMMA, les données sont récupérées en entrée par la voix, le code en rouge représente l'interprétation de l'interaction :

```
<emma :emma version="1.0" xmlns="http://www.w3.org/2003/04/emma">
<emma :interpretation id="interp1"
emma :start="1087995961542"
emma :end="1087995963542"
emma :tokens="flights from Nice to Paris"
emma :confidence="0.6"
emma :medium="acoustic"
emma :mode="speech"
emma :function="dialog"
emma :source="http://example.com/microphone/NC-61"
emma :signal="http://example.com/signals/sg23.wav"
emma :media-type="audio/dsr-202212;rate:8000;maxptime:40">
<origin>Nice</origin>
<destination>Paris</destination>
</emma :interpretation>
</emma :emma>
```

Le langage EMMA est utilisé comme une future norme d'échange de données entre les composants du système multimodale. Dans une architecture multimodale, chacune des entrées est reconnue et interprétée par des modalités spécifiques (reconnaissance vocale, gestuelle, de l'écriture, clavier, souris,...). Toutes ces informa-

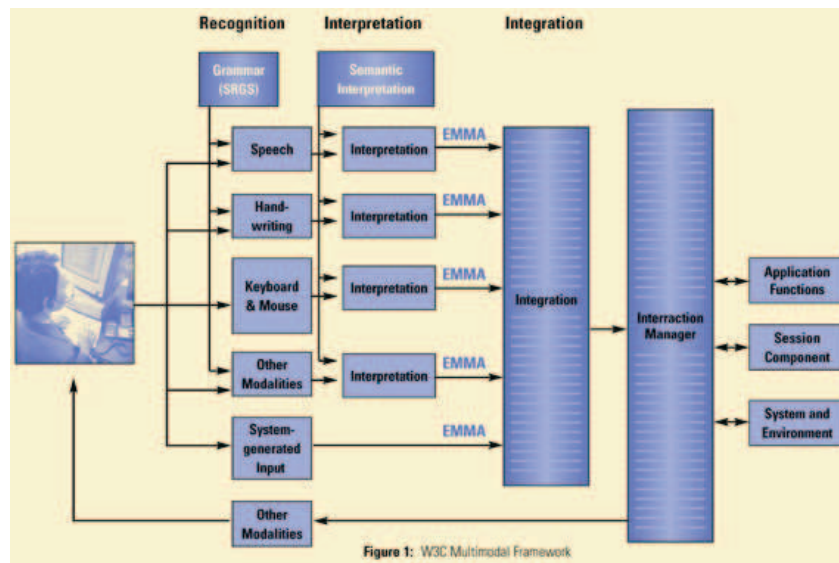


FIGURE 1.13 – Framework multimodal proposé par le W3C intégrant le langage EMMA [Rouillard 2008]

tions récupérées, qui sont propres à chacune des modalités, sont ensuite représentées dans un langage commun qui est EMMA. Selon le W3C, EMMA pourrait devenir un langage commun d'interopérabilité entre les différents composants d'un système multimodal. Ceci permettrait aux développeurs de créer des plates-formes multimodales plus fiables et moins chères pour l'intégration de chaque type de composants, mais à l'heure actuelle aucune implémentation de ses spécifications n'est disponible [Rouillard 2008].

EMMA a été développé pour permettre l'annotation des informations récupérées par les entrées de dispositifs hétérogènes. C'est un rassemblement d'informations multimédia, multimodales, multi plateformes ainsi que de systèmes d'information hétérogènes. En fait, c'est une approche pour fournir une plateforme commune pour des protocoles et des représentations différentes de l'information.

La FIGURE 1.13 présente l'architecture du framework multimodal proposé par le W3C qui intègre l'utilisation du langage EMMA. Les étapes principales sont :

- la reconnaissance d'une interaction,
- puis son interprétation,
- et enfin son intégration.

Afin d'illustrer l'utilisation de ce langage, imaginons qu'un utilisateur pointe son stylet ou son stylo électronique sur une carte affichée sur un écran et prononce la phrase « zoom avant, ici ». Le code suivant représente l'interaction vocale :

```

<emma :interpretation id="voice1"
emma :medium="acoustic"
emma :mode="voice">
<command>
<action>zoom</action>
<location emma :hook="ink">
</location>
</command>
</emma :interpretation>

```

Le code ci dessous représente l'interaction gestuelle :ink :

```

<emma :interpretation id="pen1"
emma :medium="tactile"
emma :mode="ink">
<location>
<points>42, 158 </points>
</location>
</emma :interpretation>

```

EMMA permet la fusion des données multimodales et les interactions vocales et gestuelles seront estampillées puis, la notation EMMA suivante sera générée [Rouillard 2008] :

```

<emma :interpretation emma :medium="acoustic tactile"
emma :mode="voice ink"
<command>
<action>zoom avant</action>
<location>
<point> 42, 158 </point>
</location>
</command>
</emma :interpretation>

```

Cette commande EMMA sera ensuite transférée au manager d'interaction chargé de déterminer la ou les actions à effectuer. Dans notre exemple, il pourrait s'agir de l'envoi d'une commande appropriée au navigateur web puis la génération d'une synthèse vocale annonçant la bonne exécution de la commande de l'utilisateur.

1.5.5 Synthèse

L'étude bibliographique des travaux concernant la modélisation et l'annotation des états émotionnels, nous a permis de montrer que le problème d'échange d'émotions entre applications n'est pas suffisamment traité et qu'il n'existe pas de modèle formel de représentation des émotions. En effet, les langages {EMMA, EARL, Emo-

tionML} proposés dans la littérature présentent des limites pour la représentation, l'analyse et l'annotation des émotions.

le langage EMMA est utilisé comme une norme d'échange de données entre les composants du système multimodale. Mais il ne standardise pas la représentation des données entrantes d'interprétations des utilisateurs. En effet, il permet seulement l'annotation de différentes données entrantes des utilisateurs, mais ne permet pas de définir les notions annotées. C'est un langage générique pour l'annotation des données. Ce langage est aussi monodispositif et plus orienté vers les entrées que vers les sorties [Rouillard 2008]. D'après les travaux de José Rouillard, rien n'est prévu dans l'architecture proposée par le W3C pour supporter l'interaction dans des systèmes à la fois multimodaux, multidispositifs et multicanaux.

EARL et EmotionML quant à eux, permettent la représentation et l'échange des états émotionnels, mais ils ne sont pas encore standardisés et n'expriment pas les nuances sémantiques et pragmatiques des émotions. En fait, ces deux langages d'annotation des émotions ne permettent pas d'identifier les qualités sémantiques des données émotionnelles et par conséquent, il ne permettent pas l'échange du concept émotionnel entre applications. Ils utilisent le dictionnaire de langues naturelles pour la définition des émotions. Par exemple, le mot joie avec EARL ou EmotionML va être interprété comme une chaîne de caractères {j, o, y } et pas comme un concept qui peut être traduit en plusieurs langages, joie en français, farah en arabe, gioia en italien, etc. Les langages EmotionML et EARL ne permettent pas de gérer les émotions. En effet, ils ne proposent pas des règles pour traiter et analyser les états émotionnels.

1.6 Conclusion

Dans ce chapitre, nous avons présenté dans un premier temps un état de l'art sur l'émotion humaine avec ses définitions, ses théories et ses modèles computationnels existants. Nous avons pu constater qu'il n'existe pas de consensus ni pour la définition de l'émotion, ni pour la modélisation de celle-ci. Puis, dans un second temps, nous avons cité quelques travaux intégrant l'émotion dans des applications informatiques pour l'interaction, le contrôle et l'apprentissage des émotions. Nous avons présenté un état de l'art des travaux en reconnaissance automatique des émotions. en se focalisant sur la reconnaissance des émotions à partir des signaux physiologiques. Enfin, nous avons étudié les travaux bibliographiques concernant la modélisation et l'annotation des états émotionnels.

Dans le chapitre suivant, nous présentons une nouvelle vision de la modélisation des états émotionnels qui se base sur la représentation algébrique des émotions dans un espace vectoriel.

Un modèle de représentation et d'échange des émotions

Sommaire

2.1	Introduction	39
2.2	La couche psychologique	40
2.3	La couche formelle de représentation et de traitement des émotions	42
2.3.1	Les émotions de base	44
2.3.2	Opération sur les émotions	45
2.4	La couche Langage	51
2.5	Discussion	56
2.6	Conclusion	57

2.1 Introduction

Certes les travaux actuels proposent des mécanismes permettant de capturer les états émotionnels, mais ils présentent des limitations soit sur le nombre d'émotions traitées puisqu'ils se limitent aux émotions de base comme la colère, la peur, la tristesse, la joie, etc., soit au niveau des outils de captures utilisés. L'échange émotionnel entre applications soulève des problèmes dus à l'hétérogénéité de ces applications, la complexité des états émotionnels, la diversification des outils de capture des émotions et la dépendance du traitement par rapport à ces capteurs physiques. Les systèmes interactifs actuels permettent la communication de plusieurs façons (chat, mail, etc.), mais ne prennent que très peu en compte la communication émotionnelle. On peut en effet se servir d'émoticones (smileys) animés censés traduire les émotions des utilisateurs mais cela reste encore très basique et ne permet pas de traduire tout le panel d'émotions que peut ressentir un humain. Dans ce contexte, notre objectif est de pouvoir modéliser les émotions les plus complexes y compris les émotions mélangées, masquées, simulées, etc.

Notre travail s'inscrit dans le cadre du projet Emotica. Il s'agit de définir un modèle pour la représentation et l'échange des émotions entre des applications multimodales hétérogènes. Dans notre cas on s'intéresse aux jeux sérieux (ou « Serious Games » en anglais). Notre but est de pouvoir transférer les informations émotionnelles entre les différents joueurs utilisant des plateformes et des modalités

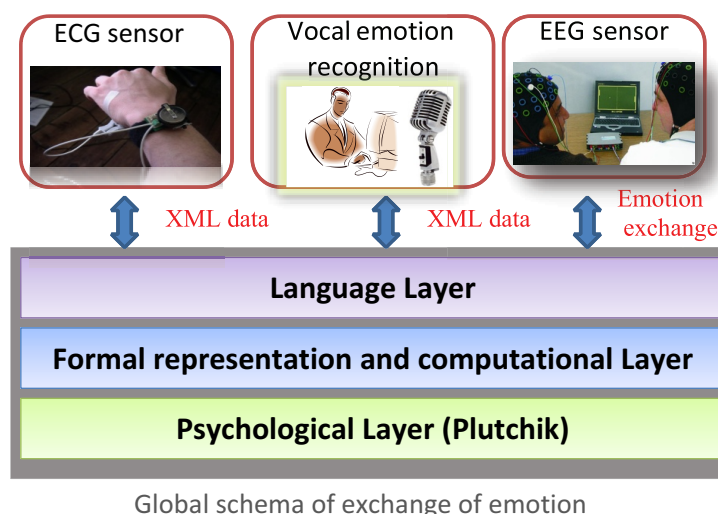


FIGURE 2.1 – schéma global du modèle à 3 couches pour échanger les émotions

différentes dans le cadre d'un jeu multi-joueurs. Dans ce genre de jeux, il est intéressant de connaître l'émotion de chaque joueur (frustration, ennui, confusion, plaisir, intérêt) afin d'améliorer l'interactivité entre eux. Cela permet aux joueurs de se sentir efficace et de maintenir leur intérêt pour le jeu. Il s'agit de capturer les traits physiologiques du joueur pour déterminer son état émotionnel et de les transférer par la suite à son ou à ses partenaires de jeu utilisant la même ou d'autres plateformes/modalités. Dans ce type de scénario, chacune des entrées est reconnue et interprétée par des modalités spécifiques à chaque application (reconnaissance vocale, gestuelle, expression faciale). Toutes ces informations récupérées qui sont propres à chacune des modalités/plateformes vont être ensuite représentées par un modèle générique commun qui doit être interprétable par les différentes machines afin d'effectuer des traitements appropriés pour l'échange émotionnels entre applications.

Dans ce chapitre, nous présentons un modèle générique pour la représentation et l'échange des émotions entre applications multimodales. Il s'agit d'un modèle de représentation hiérarchique composé de trois couches, comme le montre la figure cf. 2.1 : la couche psychologique, la couche formelle de représentation et de traitement des émotions et la couche langage.

2.2 La couche psychologique

La couche psychologique est la première couche de notre modèle. Elle présente le modèle psychologique que nous avons adopté pour décrire les émotions. En psychologie, il est généralement admis que les émotions sont l'une des facettes centrales de la psyché humaine [Gross 2011]. La plupart des théories des émotions admettent l'existence de plusieurs composantes émotionnelles. Par exemple, les processus cognitifs, les réactions physiologiques périphériques, les changements motivationnels, l'expression motrice, et le sentiment subjectif associé [Scherer 2001a]. Nous avons

vu dans le chapitre précédent, les différentes théories proposées dans la littérature pour la définition de l'émotion. Nous avons remarqué que ces théories diffèrent sur leurs hypothèses, et sur comment différencier les différents états émotionnels à partir de leurs schémas expressifs [Scherer 2001b]. Donc, Il n'existe pas de consensus sur la définition des émotions, sur leur nature exacte, ou sur les processus responsables de nos réactions émotionnelles. Pour cela et après une étude sur les différentes théories et modèles proposées en psychologie, nous avons choisi le modèle émotionnel qui nous semble le plus pertinent dans la littérature.

Donc, dans notre travail, et pour la couche psychologique nous avons adopté la théorie de Plutchik [Plutchik 1980], qui est défini par le circumplex qui est un modèle multidimensionnel basé sur huit émotions primaires de base dont vont dériver des émotions secondaires. Ce choix s'explique par les différents avantages et les spécificités de ce modèle. En effet, le modèle de Plutchik vérifie plusieurs conditions importantes pour l'élaboration de notre modèle final. D'abord, comme nous l'avons vu auparavant, le modèle de Plutchik est basé sur huit émotions primaires englobant les six émotions de base autour des quelles la plupart des chercheurs se sont mis d'accord et qui sont connues sous le nom "Big Six" [Ekman 1994] (la tristesse, la colère, la joie, le dégoût, la peur et la surprise). Ensuite, le modèle de Plutchik prend en considération l'intensité des émotions (le niveau d'excitation), ce qui le rend continu et extensible. La dimension verticale du modèle du circumplex de Plutchik (cf. FIGURE 1.4) représente l'intensité ou le niveau d'éveil, les émotions s'ordonnent du sommet à la base suivant un état d'excitation maximale jusqu'à un état de profond sommeil [Plutchik 1980]. La FIGURE 1.4 montre que la coupe horizontale présente les émotions intenses : la rage, la vigilance, l'extase, l'adoration, la terreur, l'étonnement, le chagrin et l'aversion [de Bonis 1996]. Sur l'axe vertical, chaque émotion primaire se subdivise en trois émotions d'intensité croissante. Ainsi, par exemple, on passe de la terreur (intensité forte) à la peur (intensité moyenne) puis à l'appréhension (intensité faible), ou au chagrin (intensité forte) à la tristesse (intensité moyenne) et au souci (intensité faible). De plus, le modèle de Plutchik a abordé le problème d'analogie avec les couleurs en proposant un dictionnaire d'émotion similaire au dictionnaire de couleur. " *A une science de la colorimétrie, on pourrait faire correspondre une science de l'émotionométrie*". Comme il existe des couleurs primaires et des nuances de ces couleurs qui varient en intensité et en saturation et qui constituent le spectre des couleurs, il y aurait des émotions primaires, des nuances qui varient en intensité, des combinaisons possibles entre certaines émotions de base. Enfin, le modèle de Plutchik est un modèle intuitif et riche puisqu'il donne, grâce à son circumplex, une représentation claire des émotions complexes, résultant de la combinaison de certaines émotions basiques. Ces combinaisons ne sont pas le fruit d'un choix arbitraire, elles sont dégagées d'une analyse empirique [de Bonis 1996] et le combinatoire des émotions répond à des règles assez précises, des procédures de validation expérimentale fondées sur la méthode des dyades et des triades permettant de confirmer l'existence de ces règles de combinaison. Plutchik distingue trois cas : les dyades primaires (= mélange de deux émotions adjacentes comme par exemple la peur et la surprise qui, combinées, produisent l'effroi ou encore la joie

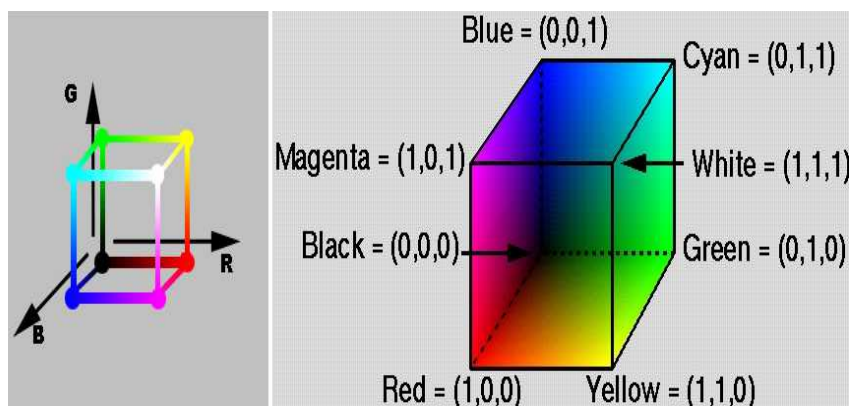


FIGURE 2.2 – Système colorimétrique RVB

et l'acceptation qui, combinées, donnent l'amour), les dyades secondaires comme étant la combinaison d'émotions proches à une émotion près, et les dyades tertiaires comme étant la combinaison d'émotions voisines à deux émotions près comme par exemple l'anxiété qui est la résultante de la peur et de l'anticipation).

2.3 La couche formelle de représentation et de traitement des émotions

La couche formelle de représentation et de traitement des émotions est la deuxième couche de notre modèle. Elle représente le modèle multidimensionnel correspondant à l'approche psychologique défini dans la première couche de notre modèle. Il s'agit d'une nouvelle vision de la modélisation des états émotionnels différente des approches traditionnelles qui utilisent une représentation ontologique. Notre approche propose une représentation algébrique des émotions. Il s'agit d'un modèle multidimensionnel qui se base sur la représentation algébrique des émotions dans un espace vectoriel. Dans cette couche, nous représentons chaque émotion par un vecteur dans un espace à huit dimensions dont chacun représente une émotion de base défini par Plutchik. Cette représentation nous permet, à la fois, de représenter une infinité d'émotions, et d'autre part de proposer des outils mathématiques très performants pour l'analyse et le traitement de ces émotions. Le modèle de représentation des émotions que nous proposons est similaire au modèle de représentation des couleurs qui se base sur trois couleurs de bases (Rouge, Vert, Bleu) pour exprimer toutes les couleurs du spectre visible (cf. FIGURE 2.2).

Notre modèle, par analogie, va se baser sur un ensemble d'émotions basiques qui vont être utilisées pour exprimer toutes les émotions possibles. Après avoir étudié le modèle de Plutchik, nous proposons de reprendre les mêmes émotions basiques pour définir la dimension de notre base. Par conséquent, la dimension de notre base d'émotion est égale à huit et elle est définie formellement par $(B) = (\text{Joie}, \text{Tristesse}, \text{Acceptation}, \text{Dégout}, \text{Peur}, \text{Colère}, \text{Surprise}, \text{Anticipation})$. Ainsi toute émotion

2.3. La couche formelle de représentation et de traitement des émotions 43

peut être réalisée à partir d'une combinaison des autres émotions fondamentales qui définissent notre Base (B).

Ainsi toute émotion (e) s'écrit sous forme unique comme combinaison linéaire des u_i = (Joie, Tristesse, Acceptation, Dégoût, Peur, Colère, Surprise, Anticipation). C'est-à-dire que toute émotion (e) représentée par le vecteur E s'écrit sous la forme suivante :

$$(e) = \sum_{i=1}^8 \langle E, u_i \rangle u_i \quad (2.1)$$

Donc $(e) = \alpha_1 Joy + \alpha_2 sadness + \alpha_3 trust + .. + \alpha_7 Surprise + \alpha_8 anticipation$
Où α_i sont des scalaires et $u_i (i = 1..8)$ sont des vecteurs de la base (B). Les composantes de E peuvent aussi être représentées en colonne sous forme d'une matrice. La matrice est appelée matrice colonne des composantes (ou des coordonnées) ou vecteur colonne des composantes de E . Pour α_i $i \in [1..8]$, le scalaire α_i est appelé la i ème composante ou i ème coordonnée du vecteur E dans la base $B = (Joye, Tristesse, Acceptation, Dégoût, Peur, Colère, Surprise, Anticipation)$.

$$E = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ . \\ . \\ \alpha_8 \end{pmatrix}_B$$

Les coordonnées (ou les composantes) α_i sont comprises dans l'intervalle $[0,1]$ et dénotent la valeur de l'émotion i de la base (B) dans l'émotion composée (e). Plus la valeur de α_i se rapproche de 1, plus l'émotion est ressentie. Plus elle est proche de 0, moins l'émotion est ressentie. Le calcul et la détermination des valeurs de α_i seront abordés dans le chapitre suivant.

En algèbre linéaire, une base d'un espace vectoriel est une famille de vecteurs de cet espace telle que chaque vecteur de l'espace puisse être exprimé de manière unique comme combinaison linéaire de vecteurs de cette base. Dans le paragraphe suivant, nous allons montrer que (B) est une base d'un espace vectoriel. Pour cela, nous devons montrer que (B) = (Joie, Tristesse, Acceptation, Dégoût, Peur, Colère, Surprise, Anticipation) est une famille de vecteurs à la fois libre et génératrice d'un espace vectoriel. Nous commençons tout d'abord par démontrer que (B) est une famille génératrice.

D'après la théorie de Plutchik, toutes les émotions aussi complexes qu'elles soient sont des combinaisons d'émotions basiques. Donc on peut obtenir toutes les émotions possibles à partir de ces huit émotions de bases. Donc notre base (B), constituée par les huit émotions de bases définies par Plutchik, est génératrice. Pour montrer que notre base (B) est libre, on va raisonner par l'absurde.

Hypothèse : Supposons que notre base n'est pas libre. Donc on aura forcément deux décompositions différentes d'une même émotion (e). Alors :

$$(e) = \alpha_1 Joy + \alpha_2 sadness + \alpha_3 trust + .. + \alpha_7 Surprise + \alpha_8 anticipation \quad (2.2)$$

et

$$(e) = \gamma_1 Joy + \gamma_2 sadness + \gamma_3 trust + .. + \gamma_7 Surprise + \gamma_8 anticipation \quad (2.3)$$

Avec

$$\begin{pmatrix} \alpha_1 \\ \alpha_2 \\ . \\ . \\ \alpha_8 \end{pmatrix}_B \neq \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ . \\ . \\ \gamma_8 \end{pmatrix}_B$$

(2.2)-(2.3) donne

$$\begin{aligned} 0 &= (\alpha_1 - \gamma_1)Joy + (\alpha_2 - \gamma_2)sadness + (\alpha_3 - \gamma_3)trust \\ &+ .. + (\alpha_7 - \gamma_7)Surprise + (\alpha_8 - \gamma_8)anticipation \end{aligned} \quad (2.4)$$

Donc

$$\alpha_1 = \gamma_1; \alpha_2 = \gamma_2; \alpha_3 = \gamma_3; ...; \alpha_7 = \gamma_7 \text{ et } \alpha_8 = \gamma_8 \quad (2.5)$$

Alors notre hypothèse de début est fausse. Donc notre base est libre.

2.3.1 Les émotions de base

Les émotions de base sont les émotions primaires de base dont vont dériver des émotions secondaires. Une émotion de base est décrite par un vecteur qui contient un seul coefficient non nul. Ainsi un vecteur représente une émotion basique si et seulement s'il vérifie la propriété suivante :

$$\forall i \in [1..8], \exists \alpha_i \text{ avec } \frac{\alpha_i}{\sum_{j=1}^8 \alpha_j} = 1 \quad (2.6)$$

Un vecteur représentatif d'une émotion de base doit contenir un seul coefficient non nul. Ci dessous la représentation vectorielle de quelques émotions de base :

$$\begin{aligned} E_{joie} &= \begin{pmatrix} \alpha_1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}_B & E_{tristesse} &= \begin{pmatrix} 0 \\ \alpha_2 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}_B & E_{colre} &= \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ \alpha_6 \\ 0 \\ 0 \end{pmatrix}_B & E_{surprise} &= \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ \alpha_7 \\ 0 \end{pmatrix}_B \\ &\text{avec } \alpha_1 \neq 0, \alpha_2 \neq 0, \alpha_6 \neq 0 \text{ et } \alpha_7 \neq 0 \end{aligned}$$

Pour illustrer le vecteur représentatif d'une émotion de base, prenons l'exemple E_{joie} , qui est la représentation vectorielle de l'émotion (*joie*), une émotion de base. Donc le vecteur E_{joie} doit nécessairement vérifier la propriété précédente 2.6, avec

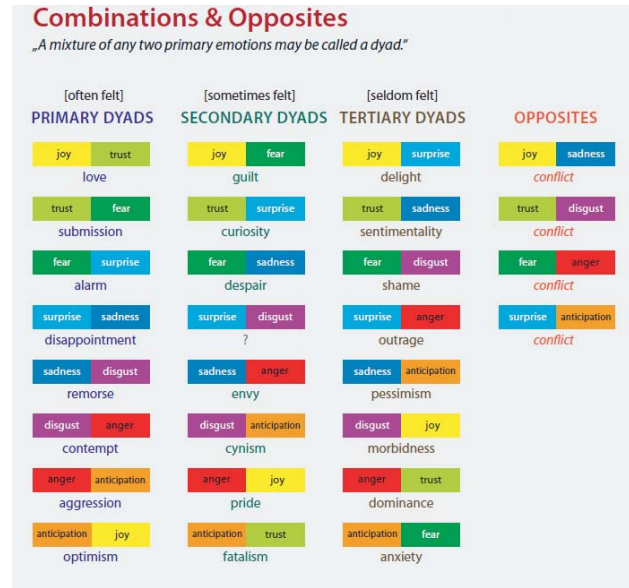


FIGURE 2.3 – les combinaisons des émotions (Plutchik)

α_i représente la première composante du vecteur E_{joie} « $\alpha_i = \alpha_1$ ».

La structure de notre modèle s'appuie sur la prise en considération de la propriété de l'intensité des émotions. En effet la composante α_i détermine l'intensité de l'émotion basique appropriée, d'où selon la valeur de la variable α_i on peut faire la différenciation entre appréhension, peur et terreur, ou bien entre rage, colère et contrariété. Donc on peut dire que la rage est l'émotion basique colère avec une forte intensité ($0.75 < i < 1$). Ainsi les différents valeurs de α_i permet de distinguer les émotions d'un état d'excitation maximal à un état de profond sommeil.

Une fois que les émotions seront représentées par des vecteurs dans un espace vectoriel approprié, on pourra les traiter grâce aux opérations usuelles sur les vecteurs. Les opérations de base qu'on va détailler dans notre travail sont l'addition vectorielle, la multiplication par un scalaire, la projection et le calcul de distance entre vecteurs.

2.3.2 Opération sur les émotions

2.3.2.1 Addition

Nous avons vu dans ce qui précède que les émotions complexes résultent de la combinaison de plusieurs émotions simples. Selon le schéma, proposé par Plutchik (cf. FIGURE 2.3), l'amour, par exemple, serait un mélange de joie et d'acceptation, la culpabilité, un mélange de joie et de peur et aussi la soumission est un mélange de peur et d'acceptation [Plutchik 1980].

Dans cette partie, nous définissons la combinaison entre les émotions comme étant l'addition des vecteurs représentant de ces émotions. Cette addition est

défini par la valeur maximale des composants (terme à terme) de chaque vecteur (les deux vecteurs ont la même dimension puisqu'ils sont exprimés dans la même base). Nous avons choisi le maximum au lieu de la somme classique parce que, d'une part l'intensité α_i ne doit pas dépasser la valeur 1, et d'autre part en raison de la spécificité de chaque modalité. Par exemple, lorsque nous effectuons une reconnaissance bimodale. Chaque modalité va nous donner un vecteur émotionnel à huit composantes pour la même émotion. Donc, pour le même axe (la même composante), nous gardons la valeur la plus grande car chaque modalité peut mieux détecter une émotion donnée qu'une autre. Par exemple la joie est mieux reconnue en expression faciale qu'en vocale [Sander 2010]. Mais la colère est reconnue aussi bien en vocal qu'en facial.

Soient E_{1u} et E_{2u} deux vecteurs qui représentent deux émotions exprimées respectivement dans la base (B) par $(\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_8)$ et $(\lambda'_1, \lambda'_2, \lambda'_3, \dots, \lambda'_8)$.

L'addition de ces deux vecteurs est défini par le vecteur E' .

$$E' = E_{1u} \oplus E_{2u} = \max(\lambda_i, \lambda'_i) \text{ for } 0 \leq i \leq 8 \quad (2.7)$$

D'où la représentation vectorielle de l'émotion Amour, qui a été défini par Plutchik, comme l'association des deux émotions basiques : joie et admiration est défini par le vecteur E_{amour} trouvé en faisant l'addition de E_{joie} et $E_{admiration}$ en appliquant la formule 2.7 :

$$E_{amour} = E_{joie} \oplus E_{admiration}$$

$$\text{donc } E_{amour} = \begin{pmatrix} \alpha_1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}_B \oplus \begin{pmatrix} 0 \\ 0 \\ \alpha_3 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}_B = \begin{pmatrix} \alpha_1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}_B$$

avec $\alpha_1 \neq 0$ et $\alpha_3 \neq 0$

De la même façon on obtiendra la représentation vectorielle des autres émotions secondaires définies par Plutchik. La FIGURE 2.3 montre toutes les combinaisons possibles des émotions définies par ce psychologue américain.

Cependant, Plutchik dans son modèle, considère les associations entre les émotions opposées comme par exemple l'association entre joie et tristesse, confiance et dégoût, etc comme étant des combinaisons conflictuelles interdites (FIGURE 2.3). Ce qui va l'amener à ne pas les considérer dans sa théorie. Cependant, dans notre modèle on va les prendre en considération. On va donc étendre le modèle de Plutchik en y ajoutant les combinaisons qu'il appelle de conflit car l'association de deux émotions opposées n'est pas seulement possible mais en plus elle a du sens.

Prenons comme exemple une personne qui est triste mais qui quand même sourit ou bien quelqu'un qui simule un sentiment alors qu'il ressent le sentiment opposé.

Dans ces cas de figures, on peut avoir en même temps une association de deux sentiments opposés qui donne naissance à un sentiment composé complexe et non pas un conflit. En effet, les expressions faciales et verbales sont des indicateurs peu dignes de confiance [Dave 1997]. Les expressions faciales étant le principal canal de communication émotionnelle, nous apprenons rapidement à les contrôler. L'expression verbale (ce que nous disons) est tout à fait contrôlable [Clay 2009]. Par contre, les signaux physiologiques sont très difficiles voire impossibles à contrefaire, bien que nous puissions être conscients de tels signaux. Par exemple, un individu rougissant de honte aura conscience de son rougissement mais ne pourra l'arrêter volontairement. De même, le rythme cardiaque est quasiment impossible à contrôler. Certaines intonations de la voix peuvent se situer dans cette catégorie, la réaction émotionnelle mettant en jeu les muscles utilisés pour parler ainsi que les cordes vocales. Ainsi, il peut être très difficile de cacher les effets d'une profonde tristesse dans notre voix.

Donc notre modèle est générique et permet de représenter tous les types d'émotions : émotions primaires, émotions composées, émotions simulées, émotions cachées. En appliquant la formule précédente 2.7, nous obtenons une émotion secondaire, qui sera représentée par un vecteur contenant deux composantes α_i non nulles. D'autre part, les émotions composées ne sont pas limitées à l'association de seulement deux émotions de base. En effet, le phénomène émotionnel est tellement riche qu'on peut avoir des états émotionnels combinant plusieurs émotions de base. Ainsi on peut avoir des vecteurs où tous les α_i sont non nulles, ce qu'on appelle les émotions plus complexes [Ortony 1990]. Donc, en utilisant notre modèle, ces émotions complexes seront aussi faciles à manipuler que les émotions basiques.

La FIGURE 2.4 montre un exemple d'addition de deux émotions dans une application de détection des émotions. Dans ce cas, la détection de l'émotion est effectuée à l'aide de plusieurs modalités. Chaque modalité va donner son propre vecteur représentant l'émotion détectée. Par exemple on va avoir un premier vecteur E_1 comme émotion captée par les expressions faciales et un autre vecteur E_2 représentant l'émotion trouvée en utilisant la modalité : signaux physiologiques. D'où l'émotion globale sera l'addition des deux vecteurs trouvés à partir des deux différentes modalités. Cet exemple montre la facilité avec laquelle on trouve l'émotion globale avec notre modèle et qui consiste simplement à faire une opération mathématique sans aucune intervention humaine. Ce qui n'est pas le cas avec d'autres modèles continus comme le modèle dimensionnel valance/arousal où l'addition des valeurs de deux points différents représentant deux émotions différentes donne un autre point qui n'a pas de sens émotionnel dans ce modèle.

2.3.2.2 La multiplication par un scalaire

La multiplication par un scalaire consiste à multiplier chacune des composantes du vecteur représentant l'émotion par un scalaire. Cette multiplication va produire un nouveau vecteur émotionnel mais avec des nouvelles valeurs α_i qui seront toujours comprises entre 0 et 1. Cette opération de multiplication des composantes de notre vecteur correspond à l'amplification ou l'affaiblissement de l'émotion. Ce

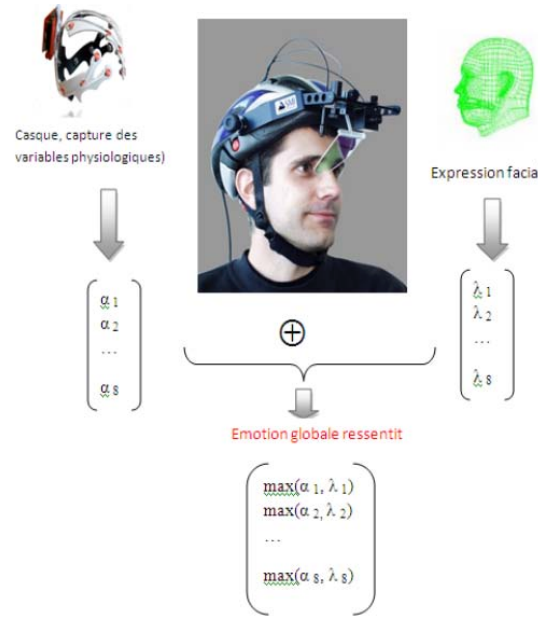


FIGURE 2.4 – Exemple d'addition de deux émotions dans une application bimodale de détection des émotions

qui correspond physiquement à l'augmentation ou l'affaiblissement de l'intensité de

l'émotion. La multiplication par un scalaire β d'un vecteur $E = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \cdot \\ \cdot \\ \alpha_8 \end{pmatrix}_B$ donne

un vecteur

$$\beta \cdot \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \cdot \\ \cdot \\ \alpha_8 \end{pmatrix}_B = \begin{pmatrix} \beta \cdot \alpha_1 \\ \beta \cdot \alpha_2 \\ \cdot \\ \cdot \\ \beta \cdot \alpha_8 \end{pmatrix}_B$$

Ainsi nous obtiendrons un nouveau vecteur qui représente l'émotion amplifiée ou bien affaiblie, et ceci dépend du contexte. En effet l'intensité de l'émotion déclenchée dépend de l'intensité du stimulus et de l'événement approprié. Cette opération (multiplication par un scalaire) peut être utilisée par exemple dans une application d'agent émotionnel où l'intensité de l'émotion décroît, voire croît au cours du temps, selon la désirabilité de l'événement et sa probabilité. Aussi elle peut être utilisée comme une fonction d'affaiblissement des émotions, c'est à dire la manière dont le personnage va rejoindre son état stable après un stimulus émotionnel.

2.3.2.3 Décomposition des émotions

Il s'agit de décomposer une émotion complexe en émotions de base, en effet les émotions complexes sont des émotions mixtes, qui sont composées de plusieurs expériences émotives de base. Elles contiennent habituellement une ou plusieurs émotions. Donc pour s'informer correctement sur l'émotion, il faut donc la décomposer en émotions élémentaires, afin de pouvoir ressentir la ou les émotions et traiter adéquatement les autres expériences qui la composent. Pour cela nous allons utiliser la projection pour la décomposition des émotions.

2.3.2.4 Projection

Soit E le vecteur représentant d'une émotion (e) dans la base (B)

$$E = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ . \\ . \\ \alpha_8 \end{pmatrix}_B$$

On définit la projection du vecteur E selon la composante i par

$$\Pi_i(E) = \alpha_i, \quad i \in [1..8] \quad (2.8)$$

Exemple :

$$\Pi_4 \begin{pmatrix} 0.2 \\ 0.1 \\ 0 \\ 0.8 \\ 0.3 \\ 0.4 \\ 0 \\ 0.1 \end{pmatrix}_B = 0.8$$

Donc, la projection nous permet d'isoler une composante donnée du vecteur E , et nous donner par la suite l'intensité de la composante choisie. Ainsi cette opération nous permettra de faciliter l'interprétation des vecteurs représentant des émotions.

2.3.2.5 Calcul de distance entre émotions

Le modèle que nous proposons est un modèle continu, c'est à dire que nous pouvons représenter une infinité d'émotions dans notre base. Pour faciliter l'interprétation de ces vecteurs, nous proposons une méthode utilisant un outil mathématique qui est la distance euclidienne, qui permettra de déterminer l'émotion correspondante dans le domaine discret (ensemble fini d'émotions connues).

La première étape de notre méthode consiste en la création d'un dictionnaire d'émotions. Ce dictionnaire est formé par des vecteurs représentant l'ensemble des



FIGURE 2.5 – Emotions basiques, similarité, intensité [Plutchik 1980]

émotions proposées par Plutchik.

La FIGURE 2.5 montre l'ensemble des émotions qui définissent une première partie du dictionnaire des émotions défini par Plutchik. L'autre partie est composée des émotions illustrées dans la FIGURE 2.3 et qui résultent de la combinaison des différentes émotions des bases.

Nous avons vu précédemment, que les émotions présentent de multiples nuances, et que la composante α_i détermine l'intensité de l'émotion basique appropriée. Donc selon la valeur de la variable α_i on peut faire la différenciation entre extase (extasy), joie (joy) et sérénité (serenity). Les valeurs de la variable α_i vont être calculées dans le chapitre suivant. Mais pour bien illustrer notre modèle nous proposons d'affecter des valeurs aux émotions .

L'ensemble des nuances des émotions basiques définies par Plutchik sont :

- Extase (extasy) - joie (joy) - sérénité (serenity) .
- Adoration (admiration) - sympathie (trust) - résignation (acceptance).
- Terreur (terror) - peur (fear) - appréhension (apprehension).
- Etonnement (amazement) - surprise (surprise) - distraction (distraction).
- Chagrin (grief) - tristesse (sadness) - songerie (pensiveness).
- Aversion (loathing) - dégoût (disgust) - ennui (boredom).
- Rage (rage) - colère (anger) - contrariété (annoyance).
- Vigilance (vigilance) - excitation (anticipation) - intérêt (interest).

Donc nous proposons de diviser l'intervalle $[0,1]$ en trois parties égales et prendre le centre de chaque intervalle pour représenter les nuances des émotions. Ainsi les émotions extase, joie et sérénité seront représentées dans notre dictionnaire respectivement par les vecteurs suivants :

$$E_{extase} = \begin{pmatrix} 0.83 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}_B \quad E_{joie} = \begin{pmatrix} 0.5 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}_B \quad E_{srnit} = \begin{pmatrix} 0.17 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}_B$$

Au final, notre base sera composée d'environ soixante dix émotions qui peut être enrichi en y ajoutant d'autres vecteurs représentant d'autres émotions non présentes dans notre base. Lors de l'interprétation d'un vecteur représentant une émotion quelconque, nous allons calculer la distance entre son vecteur et tous les autres vecteurs de notre dictionnaire. L'émotion la plus proche de notre vecteur (émotion recherchée) sera l'émotion qui va minimiser cette distance. Ainsi, nous cherchons à mesurer la ressemblance ou la similitude d'un vecteur quelconque à tous les vecteurs de la base d'émotions, pour déterminer l'émotion qui correspond le mieux à ce vecteur. En effet lors de la détection des émotions, nous allons obtenir un vecteur qui représente l'émotion ressentie par le sujet. Pour déterminer la signification émotionnelle de ce vecteur nous allons calculer la distance euclidienne pour trouver l'émotion la plus proche à cette émotion.

La distance euclidienne (ou 2-distance) entre un vecteur $E = \begin{pmatrix} x_1 \\ x_2 \\ . \\ . \\ x_8 \end{pmatrix}_B$ et un autre

vecteur $Y = \begin{pmatrix} y_1 \\ y_2 \\ . \\ . \\ y_8 \end{pmatrix}_B$ est défini par :

$$d(E, Y) = \sqrt{\sum_{i=1}^n (y_i - x_i)^2} \quad (2.9)$$

Donc l'interprétation émotionnelle de n'importe quel vecteur exprimé dans notre base se fait d'une façon automatique sans aucune intervention d'expert et cela à l'aide des outils mathématiques qui sont faciles à implémenter et qui donnent de très bons résultats.

2.4 La couche Langage

C'est la dernière couche de notre modèle. Elle a pour rôle de coder les informations émotionnelles afin d'effectuer l'échange des données émotionnelles entre les

52 Chapitre 2. Un modèle de représentation et d'échange des émotions

utilisateurs lors des différentes interactions. Notre stratégie d'annotation se base sur l'utilisation du modèle multidimensionnel défini dans le paragraphe précédent pour représenter les émotions permettant ainsi la représentation d'une infinité d'émotions. Dans ce contexte, nous proposons d'utiliser le langage de balise XML (eXtensible Markup Language) pour décrire les émotions. Le langage XML est un langage informatique de balisage générique qui permet de décrire, stocker et échanger des données structurées sous la forme d'une arborescence. La technologie XML a été choisie dans cette partie car elle possède déjà tous les outils permettant de traiter les informations qu'elle véhicule, ainsi que pour assurer l'interopérabilité entre les différentes plateformes envisagées.

Nous proposons donc de définir une balise `<emotion-vector>` pour représenter les émotions dans notre modèle vectoriel multidimensionnel. Entre les éléments `<emotion-vector>` et `< \emotion-vector>` nous décrivons la matrice colonne des composantes des émotions des utilisateurs.

Par exemple, le code suivant sera généré dans le cas de la reconnaissance de deux émotions en utilisant des modalités différentes : le rythme cardiaque (modifications physiologiques) et les expressions faciales. Ainsi, chaque modalité va donner un vecteur émotionnel avec des coefficients différents.

```
<emotion-vector>
<modality set= "multi-modality" count="2">
<vector mode="heart-rate">
<intensity axis="joy">0.0</intensity >
<intensity axis="sadness">0.4</intensity >
<intensity axis="trust">0.0</intensity >
<intensity axis="disgust">0.0</intensity >
<intensity axis="fear">0.8</intensity >
<intensity axis="anger">0.2</intensity >
<intensity axis="surprise">0.0</intensity >
<intensity axis="anticipation">0.0</intensity >
</vector>
<vector mode="face">
<intensity axis="joy">0.0</intensity >
<intensity axis="sadness">0.7</intensity >
<intensity axis="trust">0.0</intensity >
<intensity axis="disgust">0.0</intensity >
<intensity axis="fear">0.1</intensity >
<intensity axis="anger">0.0</intensity >
<intensity axis="surprise">0.0</intensity >
<intensity axis="anticipation">0.0</intensity >
</vector>
</emotion-vector>
```

Dans l'exemple ci dessus, chaque modalité a donné son propre vecteur représen-

tant l'émotion détectée.

En appliquant l'addition des vecteurs émotionnels trouvés à partir des deux différentes modalités, nous obtenons le vecteur de l'émotion globale ressentie.

$$E = \begin{pmatrix} 0 \\ 0.7 \\ 0 \\ 0 \\ 0.8 \\ 0.2 \\ 0 \\ 0 \end{pmatrix}_B$$

L'étape suivante consiste à appliquer l'algorithme de distance pour déterminer l'émotion la plus proche à ce vecteur dans la base de données.

Une autre représentation du résultat de la reconnaissance des émotions peut être donnée par le code xml suivant en encapsulant directement le vecteur émotionnel global résultant de la sommation des vecteurs générés par chaque modalité.

```
<emotion-vector>
<modality set= "multi-modality" >
<vector mode="multi-modality">
<intensity axis="joy">0.0</intensity >
<intensity axis="sadness">0.7</intensity >
<intensity axis="trust">0.0</intensity >
<intensity axis="disgust">0.0</intensity >
<intensity axis="fear">0.8</intensity >
<intensity axis="anger">0.2</intensity >
<intensity axis="surprise">0.0</intensity >
<intensity axis="anticipation">0.0</intensity >
</vector>
</emotion-vector>
```

L'interprétation et le résultat final des deux exemples donnent exactement le même résultat, c'est à dire le même vecteur de l'émotion globale. Nous allons opter pour la première représentation xml parce qu'elle offre plusieurs avantages par rapport à la deuxième représentation. En effet, elle permet de donner une description détaillée des différentes modalités utilisées tels que le nombre et le type des modalités utilisées pour la reconnaissance de l'émotion. Aussi, cette représentation permet de nous donner plus d'information sur l'émotion en donnant une décomposition détaillée des différents vecteurs émotionnels qui la constituent. En effet, les canaux expressifs peuvent être contradictoires ou complémentaires et les émotions peuvent entraîner des expressions très subtiles ou contrôlées et cette représentation nous permet de savoir si l'émotion a été simulée ou masquée lors de l'interaction.

La FIGURE 2.6 montre un exemple d'échange émotionnel entre applications.

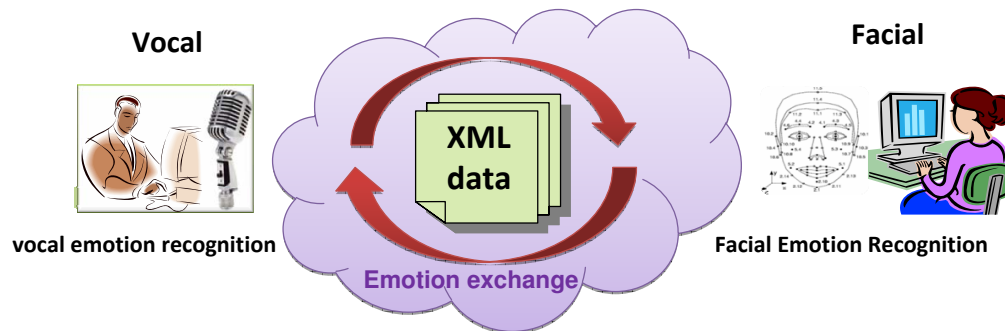


FIGURE 2.6 – Un exemple d'échange émotionnel entre deux utilisateurs

En effet le premier utilisateur utilise une caméra pour la détection des émotions. L'émotion détectée à l'aide des expressions faciales sera représentée par notre modèle vectoriel, et transmis par la suite à l'autre utilisateur sous forme de données XML. La machine du deuxième utilisateur analyse les données XML et détermine l'émotion correspondante en utilisant l'algorithme de la distance euclidienne pour trouver l'émotion correspondante. Le deuxième utilisateur va réagir en voyant l'émotion de l'autre utilisateur affichée à l'aide d'un avatar ou d'autre techniques de restitution d'émotion, et l'émotion ressentie va être captée à l'aide de son micro (la tonalité de sa voix) et va être envoyé de la même façon sous forme de données XML au premier utilisateur.

Notre modèle permet d'améliorer l'interaction entre différents utilisateurs utilisant différents modalités. Grâce à la représentation vectorielle adoptée, l'interprétation émotionnelle devient très simple et l'échange automatisé des émotions entre différents utilisateurs devient plus facile.

Utilisation de EmotionML Dans la couche langage, nous pouvons également utiliser le langage EmotionML pour la description des émotions mais nous devons tout d'abord ajouter un nouvel élément `<plutchik-dimension>` pour représenter les émotions dans notre modèle vectoriel. EmotionML, comme nous l'avons vu dans le chapitre précédent, est un langage d'annotation et de représentation des émotions. Ce langage contient des éléments pour représenter les émotions dans l'approche catégorielle, l'approche dimensionnelle, l'approche d'évaluation cognitive (appraisal) et l'approche de tendance à l'action [Frijda 1986]. Afin de pouvoir représenter les émotions dans notre modèle vectoriel nous devons modifier le vocabulaire de EmotionML et ajouter la balise suivante :

```
<vocabulary type="plutchik-dimension" id="intensity-dimension">
<item name="joie"/>
<item name="tristesse"/>
<item name="acceptation"/>
<item name="dégoût"/>
```

```

<item name="peur"/>
<item name="colère"/>
<item name="surprise"/>
<item name="anticipation"/>
</vocabulary>

```

Dans l'exemple suivant, deux émotions ont été détectées lors d'une reconnaissance multimodale utilisant les expressions faciales et les signaux physiologique. De la même façon que XML, l'émotion globale peut être annotée de deux façons différentes. La première consiste à représenter les deux émotions séparément avec des coefficients différents comme le montre l'exemple suivant :

```

<emotionml xmlns="http://www.w3.org/2009/10/emotionml"
plutchik-dimension-set="http://www.w3.org/TR/emotion-voc/xml#intensity-
dimension">
  <emotion expressed-through="face">
    <plutchik-dimension name="joie" value="0.8" />
    <plutchik-dimension name="tristesse" value="0" />
    <plutchik-dimension name="acceptation" value="0" />
    <plutchik-dimension name="dégoût" value="0" />
    <plutchik-dimension name="peur" value="0" />
    <plutchik-dimension name="colère" value="0" />
    <plutchik-dimension name="surprise" value="0" />
    <plutchik-dimension name="anticipation" value="0" />
  </emotion>
  <emotion expressed-through="physiology">
    <plutchik-dimension name="joie" value="0" />
    <plutchik-dimension name="tristesse" value="0.2" />
    <plutchik-dimension name="acceptation" value="0" />
    <plutchik-dimension name="dégoût" value="0" />
    <plutchik-dimension name="peur" value="0" />
    <plutchik-dimension name="colère" value="0.7" />
    <plutchik-dimension name="surprise" value="0" />
    <plutchik-dimension name="anticipation" value="0" />
  </emotion>
</emotionml>

```

La deuxième représentation consiste à annoter l'émotion globale par le vecteur résultant de la sommation des deux vecteurs représentant les deux émotions détectées lors de la reconnaissance multimodale :

```

<emotionml xmlns="http://www.w3.org/2009/10/emotionml"
plutchik-dimension-set="http://www.w3.org/TR/emotion-voc/xml#intensity-
dimension">
  <emotion expressed-through="face physiology">
    <plutchik-dimension name="joie" value="0.8" />

```

```

<plutchik-dimension name="tristesse" value="0.2" />
<plutchik-dimension name="acceptation" value="0" />
<plutchik-dimension name="dégoût" value="0" />
<plutchik-dimension name="peur" value="0" />
<plutchik-dimension name="colère" value="0.7" />
<plutchik-dimension name="surprise" value="0" />
<plutchik-dimension name="anticipation" value="0" />
</emotion>
</emotionml>

```

2.5 Discussion

Notre modèle formel multidimensionnel correspond à l'approche de Plutchik qui définit la couche psychologique. Le modèle de Plutchik est basé sur huit émotions de base faites de 4 paires opposées : joie-tristesse, acceptation-dégoût, peur-colère, surprise et anticipation. Dans notre travail, nous avons choisi de représenter les émotions dans un espace à huit dimensions. Une autre alternative est de définir une base à quatre dimensions où la valeur de chaque axe varie entre -1 et 1. Cette représentation est plus proche du modèle de Plutchik, car ce dernier regroupe les huit émotions de base en 4 paires opposées. La base serait alors $\tilde{B} = (Joie-Tristesse, Acceptation-Dégoût, Peur-Colère, Surprise-Anticipation)$, et chaque émotion serait représenté par le vecteur

$$\tilde{E} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{pmatrix}_{\tilde{B}}, \alpha_i \in [-1, 1]$$

Cette représentation est certes plus proche de l'approche de Plutchik, mais elle est limitée et ne permet pas la description de toutes les émotions telles que les émotions masqués. Plutchik ne prend pas en compte ce type d'émotions et il considère la combinaison entre opposites comme combinaison conflictuelle interdite. Cependant, les émotions auxquelles on fait face tous les jours sont souvent mélangées ou masquées. C'est pour cette raison que notre modèle est composé de huit axes. Notre modèle est plus général que le modèle à quatre axes. En effet chaque vecteur décrit dans la base $\tilde{B}$ a un équivalent dans la base B et l'inverse n'est pas vrai. Par exemple le vecteur équivalent au vecteur $V1$ exprimé dans la base $\tilde{B}$

$$\tilde{V}1 = \begin{pmatrix} 0.5 \\ -0.7 \\ 0.2 \\ -0.4 \end{pmatrix}_{\tilde{B}}$$

est donné par le vecteur $W1$ suivant qui est exprimé dans la base B

$$\tilde{W}1 = \begin{pmatrix} 0.5 \\ 0 \\ 0 \\ 0.7 \\ 0.2 \\ 0 \\ 0 \\ 0.4 \end{pmatrix}_{\tilde{B}}$$

Cependant, le vecteur $W2$ exprimé dans la base B n'a pas d'équivalent dans la base $\tilde{B}$

$$\tilde{W}2 = \begin{pmatrix} 0.5 \\ 0.7 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}_{\tilde{B}}$$

Or ce vecteur représente une émotion masquée, qui peut être ou bien de la joie masquée par des expressions faciales reflétant la tristesse ou bien l'inverse.

2.6 Conclusion

Dans ce chapitre, nous avons proposé un modèle émotionnel générique qui permet de faciliter la communication et l'échange émotionnel entre utilisateurs et qui assure la communication entre applications hétérogènes multimodales. Notre modèle est basé sur des travaux en psychologie et il est composé de trois couches distinctes : la couche psychologique, la couche formelle de calcul et enfin la couche langage. La première couche (la couche psychologique) représente le modèle psychologique que nous avons adopté pour décrire les émotions. Il s'agit de la théorie de Plutchik, définie par le circumplex qui est un modèle multidimensionnel basé sur des émotions primaires de base dont vont dériver des émotions secondaires. La seconde couche est un modèle formel multidimensionnel qui correspond à l'approche psychologique de la première couche : chaque émotion est représentée par un vecteur dans un espace 8-dimensions où chaque axe représente une émotion de base tels que défini par Plutchik. Cette approche multidimensionnelle se base sur la représentation algébrique des émotions dans un espace vectoriel. L'idée de base consiste à représenter une émotion par un vecteur dans un espace approprié, puis à lui appliquer toute une gamme de traitements algébriques. Cela nous a permis de représenter une infinité d'émotions et de proposer des outils mathématiques très performants pour l'analyse et le traitement de ces émotions comme l'addition, la projection la décomposition, etc. Enfin, la dernière couche qui est la couche langage, basée sur les descriptions

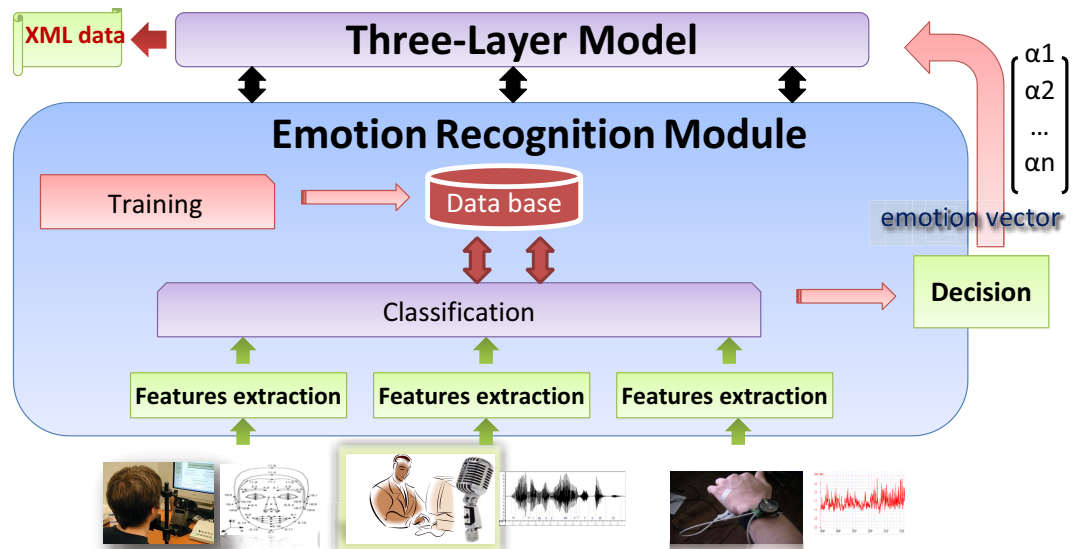


FIGURE 2.7 – Schéma global de notre modèle de représentation des émotions.

XML des données émotionnelle finales ce qui permet de faciliter le transfert entre les applications. Notre proposition facilite l'échange inter-systèmes. Elle permet le transfert efficace de l'information émotionnelle entre applications hétérogènes, indépendamment des modalités et des capteurs utilisés pour la détection.

Afin de valider notre modèle, il est nécessaire d'effectuer une simulation. Nous proposons donc d'enrichir notre modèle en ajoutant un module de reconnaissance des émotions. La FIGURE 2.7 montre le schéma global de notre modèle en ajoutant le module de reconnaissance des émotions. Le chapitre suivant porte sur ce module. Nous nous intéressons à la reconnaissance des émotions tout en suivant la modélisation présentée dans ce chapitre, afin d'en tester la réelle utilité et montrer comment calculer les valeurs des composantes α_i . Ainsi, le chapitre suivant présente en détail une approche de reconnaissance automatique des émotions fondée sur le traitement des signaux physiologiques.

Reconnaissance des émotions basée sur les signaux physiologiques

Sommaire

3.1	Introduction	59
3.2	Acquisition des données	61
3.3	Schéma général de la méthode	62
3.4	Apprentissage	64
3.4.1	Découpage	64
3.4.2	Lissage	65
3.4.3	Détection des pics	69
3.4.4	Extractions des vecteurs caractéristiques	70
3.4.5	Création de la base d'apprentissage	72
3.5	Détection	72
3.6	Expérimentation et résultats	74
3.6.1	Choix de la taille de la fenêtre de lissage	75
3.6.2	Choix de la valeur k de l'algorithme KNN	77
3.6.3	Résultats généraux et comparaison	81
3.7	Conclusion	85

3.1 Introduction

Une modélisation ne trouve tout son intérêt que dans son utilité applicative. Nous nous intéressons, dans ce chapitre, à la reconnaissance des émotions en utilisant le modèle formel pour la représentation et l'analyse des émotions que nous avons proposé dans le chapitre précédent. La reconnaissance des émotions est un domaine très actif, marqué par un intérêt croissant depuis sa création. Ce domaine a pour but de doter les machines des capacités de compréhension des comportements humains afin d'améliorer l'interaction Homme-machine. En effet, les émotions ont une grande influence aussi bien sur le comportement de l'individu que sur son raisonnement. Une émotion génère en particulier des réactions motrices et neurophysiologiques qui peuvent se traduire à travers plusieurs canaux de communication émotionnelle tels que :

- les expressions faciales ;
- les intonations de la voix (la prosodie) ;
- les réactions du Système Nerveux Autonome (les signaux physiologiques) ;
- les positions et mouvements du corps.

L'acquisition de données émotionnelles peut se faire à l'aide de plusieurs instruments de captation des émotions : microphones, caméras, capteur de réponse épidermique, moniteur de rythme cardiaque, capteur de pression, etc. La reconnaissance des émotions, a pour but de déceler l'émotion exprimée voire ressentie chez un utilisateur humain afin de prendre en compte cet état émotionnel. Picard propose dès 1997 une série d'étapes à suivre pour effectuer une reconnaissance affective [Picard 1997b] :

Acquisition du signal d'entrée : cette étape consiste tout d'abord à mettre en place des dispositifs d'acquisition ou de capture des données du monde réel. Par exemple, des microphones pour capter le signal sonore de la voix, ou une caméra pour le visage. Ensuite, à enregistrer les changements physiologiques de l'organisme en réponse à un stimulus externe.

Identifier des formes dans le signal : c'est-à-dire détecter et extraire des formes caractéristiques ou des variations typiques relatives à une émotion ou un état affectif. Cette forme caractéristique sera utilisée comme référence pour l'étape suivante.

identifier : c'est-à-dire être capable, d'après les caractéristiques et formes reconnues, d'inférer l'émotion plus probablement exprimée par l'utilisateur.

Apprentissage : c'est-à-dire entraîner la machine à reconnaître et classer une émotion.

Évaluation : évaluer automatiquement le signal et délivrer l'émotion finalement identifiée.

Dans ce chapitre, nous présentons une nouvelle approche de reconnaissance des émotions fondée sur le traitement des signaux physiologiques. La méthode de détection d'émotions que nous proposons, suit ces différentes étapes, tout en se basant sur l'application des techniques de traitement de signal et en utilisant le modèle multidimensionnel décrit dans le chapitre précédent pour la représentation des émotions.

Le traitement du signal est la discipline qui développe et étudie les techniques de traitement, d'analyse et d'interprétation des signaux. Bien que cette discipline trouve son origine dans les Sciences de l'Ingénieur (particulièrement l'électronique et l'automatique), elle fait aujourd'hui largement appel à de nombreux domaines des mathématiques, comme la théorie du signal, les processus stochastiques, les espaces vectoriels et l'algèbre linéaire et des mathématiques appliquées, notamment la théorie de l'information, l'optimisation ou encore l'analyse numérique. Parmi les types d'opérations possibles sur ces signaux, on peut dénoter le filtrage, le débruitage, la déconvolution, la prédiction, l'identification, la classification, etc. Mais définissons tout d'abord ce qu'est un signal. On appelle signal toute grandeur physique tensorielle qui varie soit continûment (signaux analogiques) soit discrètement (signaux numériques) au cours du temps. L'évolution dans le temps de la grandeur considérée est régie par la dynamique spécifique du signal. Quelque-fois la loi temporelle

régissant le phénomène est bien connu (signaux déterministes) et d'autre fois il est difficile, voir impossible de le décrire (signaux aléatoires). Les signaux que nous allons analyser sont des signaux numériques issus de quatre capteurs physiologiques qui sont :

- La résistance de la peau (GSR) ;
- le volume sanguin pulsionnel (BVP) ;
- L'électromyogramme (EMG) ;
- La respiration

Le processus d'acquisition de ces signaux sera détaillé dans la section suivante. Dans ce chapitre nous présentons une approche de reconnaissance automatique des émotions à partir des signaux physiologiques. Nous commençons, tout d'abord, par décrire les différents signaux physiologiques qu'on va utiliser dans notre travail. Ensuite, nous détaillons les différentes étapes du traitement effectuées pour la reconnaissance des émotions à partir des signaux physiologiques. Puis, nous discutons les résultats obtenus. Enfin, et pour conclure ce chapitre, nous faisons une brève synthèse des idées traitées.

3.2 Acquisition des données

Dans notre travail, nous avons utilisé les signaux physiologiques fournis par MIT (Massachusetts Institute of Technology) utilisés par J.Healey dans sa thèse [Healey 2000]. Les données de MIT ont été recueillies par 4 capteurs :

- la résistance de la peau (GSR) ;
- la respiration ;
- la pression sanguine (BVP) ;
- et le signal électromyographie (EMG), auxquels s'ajoute le rythme cardiaque déduit des variations de pression sanguine.

Ils ont utilisé la méthode d'un seul participant pendant plusieurs jours pour la collecte des données. Le participant était un acteur qui a exprimé huit émotions différentes à l'aide d'un système développé par Clynes [Clynes 1982]. L'acteur n'exprime pas seulement chaque émotion de l'extérieur, mais il doit vivre la situation de chaque émotion de l'interne.

Les données collectées représentent 25 minutes d'enregistrement par jour, sur une période de 20 jours. Chaque jour contient 4 signaux correspondant aux 4 modalités citées ci-dessus, représentant les 8 émotions basiques à identifier (aucune émotion, colère, haine, peine, amour platonique, amour romantique, joie et vénération). La numérisation des signaux obtenus a été effectuée avec un échantillonnage de 20 échantillons par seconde. Chaque signal est donc formé par 30.000 échantillons.

La procédure de Clynes consiste à enregistrer les tensions ressenties par le candidat au fur et à mesure qu'il écoute une musique ou regarde des photos favorisant l'émergence d'une expression émotionnelle spécifique (Les cycles sensibles). Le participant doit appuyer sur un bouton (sentograph) avec une force comparable à la tension ressentie. Plus il ressent de tension, plus il appuie avec force, et vice-versa.

TABLE 3.1 – Ensemble des images utilisées pour provoquer les huit émotions ainsi qu’une description de ces états émotionnels

Émotion	Images	Description	Intensité	Valence
aucune émotion	papier blanc, machine à écrire	Ennui, offre d’emploi	bas	neutre
colère	des gens en colère	désir de se battre	Très haut	Très négative
haine	Injustice cruauté	colère passive	Bas	négative
peine	perte de mère enfant souffrant	perte, tristesse	haut	négative
amour Platonique	famille, été	joie, paix	bas	positive
amour romantique	rencontre romantique	excitation, désir	très haut	positive
joie	de la musique	beaucoup, de joie	moyennement haut	positive
vénération	église prière	calme, paix	très bas	neutre

La procédure de Clynes incite le sujet à exprimer la même émotion pendant 3 minutes. Pendant la procédure de génération des émotions, l’acteur doit donner son interprétation de chaque émotion et choisir des photos en liaison avec chaque émotion. Le tableau suivant montre l’ensemble des images utilisées pour provoquer les huit émotions ainsi qu’une description de ces états émotionnels.

la FIGURE 3.1 montre l’exemple d’une session de données pour un utilisateur recueillies dans une journée en utilisant les 4 modalités citées précédemment. Elle montre les 4 signaux correspondant aux 4 modalités. Pour chaque modalité, on a un signal d’environ 25 minutes d’enregistrement, représentant les huit émotions (aucune émotion, colère, haine, peine, amour platonique, amour romantique, joie et vénération) exprimées par l’acteur.

3.3 Schéma général de la méthode

La méthode de détection d’émotions que nous proposons est composée, comme le montre la FIGURE 3.2, de deux étapes principales. La première étape, appelée "Apprentissage", est une étape "offline" qui prend en entrée pour chacune des différentes modalités : le signal physiologique et l’information émotionnelle et donne en sortie la base émotionnelle relative à cette modalité. Elle permet, à partir d’un signal physiologique dont les émotions correspondantes sont connues, de trouver et d’extraire les formes caractéristiques typiques pour chacune des émotions. Nous utilisons ces formes pour générer des vecteurs caractéristiques que nous regroupons

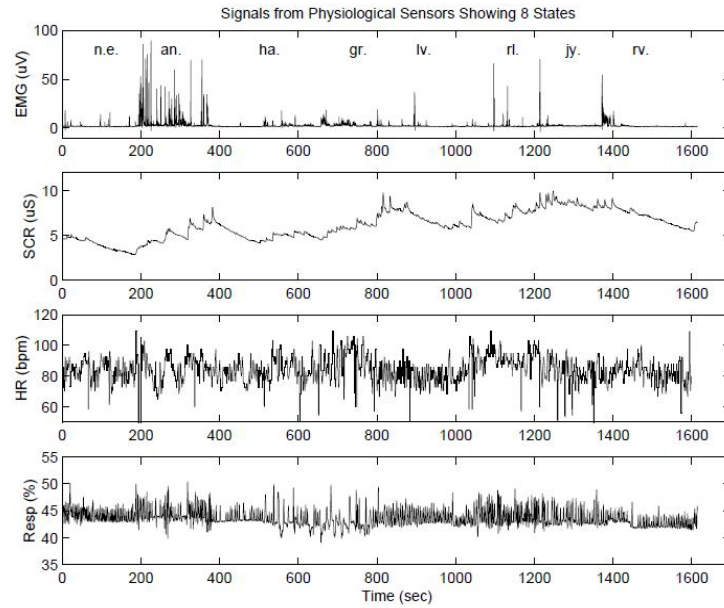


FIGURE 3.1 – Exemple d'une session de données recueillies dans une journée par les 4 capteurs : EMG, GSR, BVP et respiration

ensemble pour former une base émotionnelle d'apprentissage.

La deuxième étape, appelée détection, est une étape "online". Elle prend en entrée la base émotionnelle d'apprentissage créée à l'étape précédente et un signal physiologique dont les émotions sont inconnues. Elle analyse ensuite en temps réel ces signaux physiologiques pour détecter d'éventuelles activités émotionnelles. Lorsque une émotion est détectée, nous isolons la partie du signal correspondante et nous calculons son vecteur caractéristique. Nous cherchons ensuite dans la base d'apprentissage, l'émotion la plus proche en appliquant l'algorithme de calcul de distance euclidienne (cf. section 2.3.2.5).

Dans notre travail, nous avons utilisé la moitié des signaux, fournis par MIT et

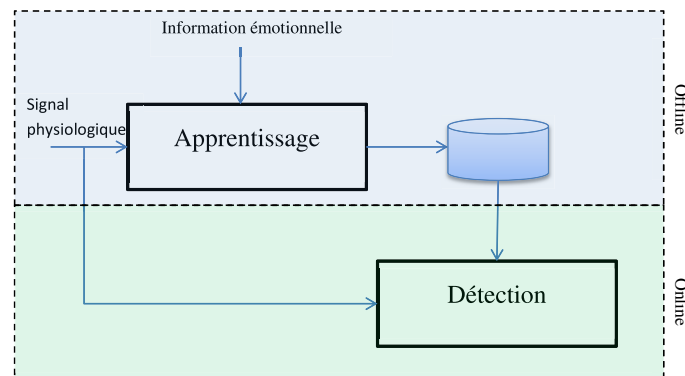


FIGURE 3.2 – Schéma général de la méthode de détection monomodale d'émotions

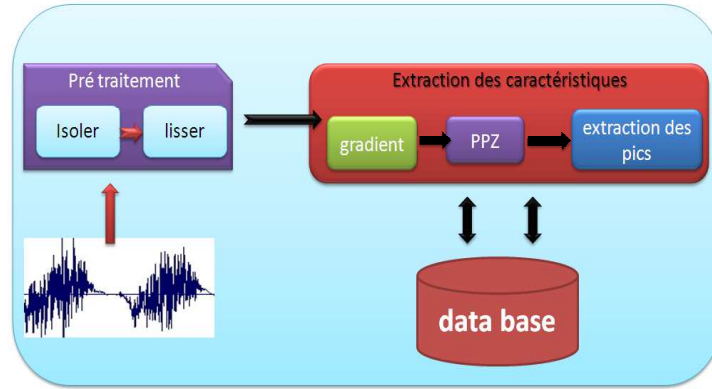


FIGURE 3.3 – Schéma général de l'étape d'apprentissage

recueillis par les 4 modalités pour la phase d'apprentissage et l'autre moitié pour effectuer la détection et le test. Chacune des étapes d'apprentissage et de détection est l'enchaînement de plusieurs sous-étapes faisant appel à des méthodes et des techniques du domaine du traitement de signal. Nous détaillons par la suite ces étapes en commençant tout d'abord par l'étape d'apprentissage.

3.4 Apprentissage

La première étape de la méthode de détection proposée, est l'étape d'apprentissage. Le but de cette étape est la génération d'une base émotionnelle formée par les vecteurs caractéristiques de chaque émotion. Ces vecteurs sont regroupés par modalité. La génération de ces vecteurs est effectuée en suivant les différentes étapes données par la FIGURE 3.3 qui sont :

Le découpage : Elle permet de découper le signal d'entrée représentant les huit émotions afin d'isoler une émotion donnée.

Le lissage : Cette étape est importante car elle permet de débruiter le signal d'entrée. Ceci rend les traitements qui suivent beaucoup plus efficaces.

Détection des pics : Elle permet de détecter les activités émotionnelles dans le signal et d'isoler les parties correspondantes.

Extraction des caractéristiques : Elle génère un vecteur caractéristique décrivant la partie du signal extrait précédemment.

Nous détaillons dans la suite chacune de ces différentes étapes en commençant tout d'abord par le découpage.

3.4.1 Découpage

La première étape de l'apprentissage est le découpage du signal représentant les huit émotions afin d'isoler chaque émotion. Il s'agit de découper le signal en se basant sur la durée de chaque émotion induite. Nous obtenons ainsi, huit portions de signal

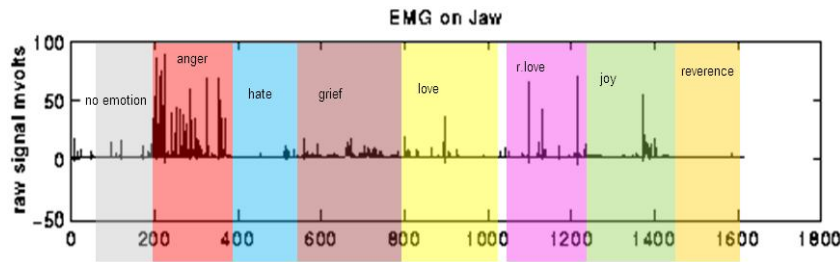


FIGURE 3.4 – Découpage du signal en portions correspondant aux huit émotions

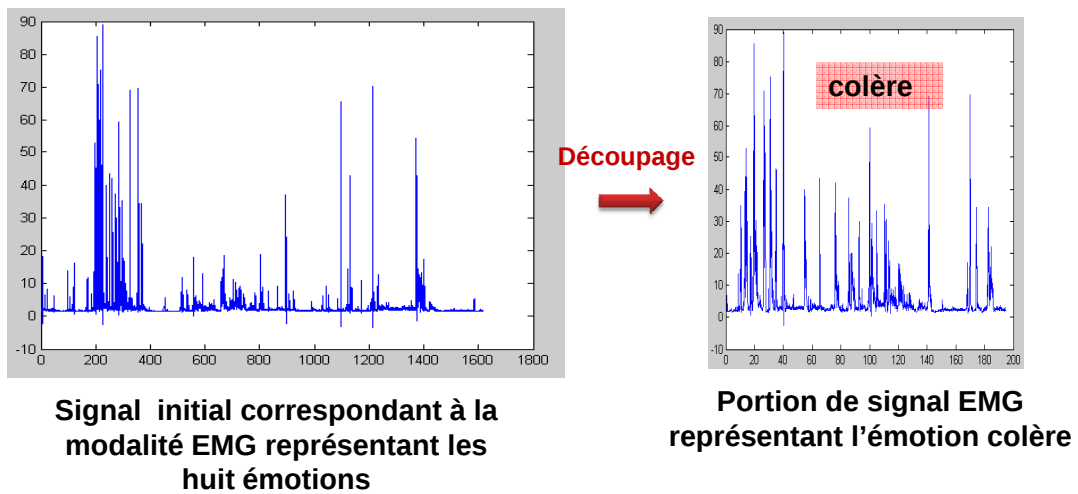


FIGURE 3.5 – Découpage du signal EMG : isolation de l'émotion colère

correspondant aux huit émotions (cf. FIGURE 3.4). Par exemple, pour isoler une émotion donnée (eg colère), nous prenons la portion de signal qui correspond à la période pendant laquelle l'utilisateur exprime cette émotion. La FIGURE 3.5 présente la courbe que nous avons obtenue lors du découpage du signal qui correspond à la modalité EMG afin d'isoler l'émotion colère .

La FIGURE 3.6 montre des portions de signaux physiologiques mesurés pour un utilisateur pendant qu'il exprimait la colère (à gauche) et la peine (à droite). Nous remarquons que les segments montrés ici sont visiblement différents pour les deux émotions.

3.4.2 Lissage

Ce traitement vise, en premier lieu, à supprimer tout ou du moins une partie du bruit afin de récupérer un signal dont l'information et les caractéristiques sont plus faciles à extraire. En effet, dans l'idéal, on obtiendrait grâce à l'étape d'acquisition (cf. section 3.2), une mesure correspondant directement aux quantités pertinentes que l'on cherche à étudier. Mais, dans la réalité, lorsqu'on réalise ces mesures, on observe l'apparition de signaux parasites, appelés bruit de mesure, qui viennent se

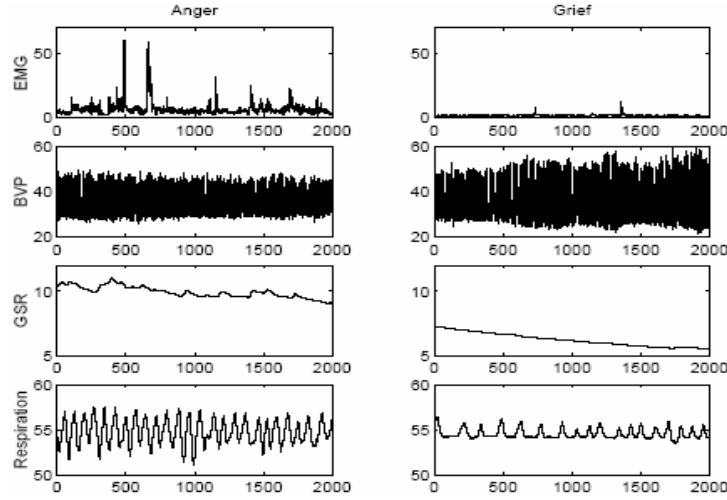


FIGURE 3.6 – Des portions des signaux physiologiques correspondant aux émotions : la colère (à gauche) et la peine (à droite). Les signaux de haut en bas : electromyogram, pression de volume de sang , conductivité de peau et respiration.

superposer au signal dit utile (i.e l'information que l'on souhaite récupérer). Ces signaux sont une gêne pour la compréhension et l'étude de l'information que le signal transporte. Il permet ensuite de regrouper les variations du signal relatives à une émotion dans une forme caractéristique comme nous allons le voir plus tard.

Il existe principalement deux manières pour débruiter un signal :

- la première consiste à effectuer tout d'abord un changement de domaine en passant du domaine spatial au domaine fréquentiel en appliquant une transformée de Fourier définie par l'équation suivante :

$$F(k) = \sum_{j=1}^N s(k) e^{-2i\pi k \frac{j}{N}} \quad 1 \leq k \leq N \quad (3.1)$$

s représente le signal d'entrée formé par N échantillons.

Ensuite un filtre passe bas est appliqué pour couper les hautes fréquences qui correspondent aux bruits. Enfin, la transformée de Fourier inverse est appliquée pour revenir au domaine spatial.

- La deuxième méthode, qui est équivalente, consiste à convoluer le signal dans le domaine spatial avec les différents filtres. Ce qui est moins coûteux en termes de calculs. La convolution du signal $s(t)$ par un filtre $h(t)$ de longueur $l + 1$, noté $s(t) * h(t)$, consiste à appliquer la formule suivante :

$$S(k) = \sum_{j=-\frac{l}{2}}^{j=\frac{l}{2}} s(k)h(k-j) \quad (3.2)$$

Dans notre méthode nous utilisons la deuxième approche c'est-à-dire la convolution. Il existe plusieurs filtres qui assurent le débruitage comme le filtre moyennneur,

médian, gaussien, haning... Nous allons comparer dans la suite les trois filtres : moyennneur, gaussien et haning.

Le filtre moyennneur consiste à calculer pour chaque élément du signal, la moyenne des l éléments voisins. La FIGURE 3.7 montre la forme d'un filtre moyenne de longueur $l = 5$. Les effets du filtre moyennneur varient avec la taille du filtre : plus les dimensions du noyau sont importantes, plus le bruit sera éliminé ; mais en contrepartie, les détails fins seront eux-aussi effacés.

1/5	1/5	1/5	1/5	1/5
-----	-----	-----	-----	-----

FIGURE 3.7 – Forme d'un filtre moyennneur de longueur égale à $l = 5$

Le filtre gaussien, comme son nom l'indique est un filtre dont la réponse impulsionnelle est une fonction gaussienne. Contrairement au filtre moyennneur les différents coefficients n'ont pas le même poids. Les pondérations les plus importants sont données aux éléments les plus proches de l'élément central. La FIGURE 3.8 montre les coefficients d'un filtre gaussien de longueur $l = 5$.

0.06	0.24	0.4	0.24	0.06
------	------	-----	------	------

FIGURE 3.8 – Coefficients d'un filtre gaussien de longueur égale à $l = 5$

Le filtre ou fenêtre de Hanning est un filtre similaire au filtre de Hamming. Il apporte des pondérations en forme de cloche mais annule les points des extrémités. Il permet de détecter les pics tout en améliorant la forme du spectre.

0.0833	0.25	0.3333	0.25	0.0833
--------	------	--------	------	--------

FIGURE 3.9 – Coefficients d'un filtre de Hanning de longueur égale à 5

Dans la FIGURE 3.10 nous présentons les résultats de lissage en utilisant ces 3 derniers filtres et un autre filtre moyennneur de longueur $l = 100$. Un appel récursif $n = 10$ fois est effectué pour les filtres moyennneur et gaussien de longueur $l = 5$ pour augmenter l'effet de lissage. Le filtre de Hanning utilisé est de longueur $l = 100$. Nous remarquons que le meilleur résultat est obtenu pour le filtre de Hanning. En effet, l'effet de lissage est plus prononcé. Le filtre moyennneur de longueur $l = 100$ donne un résultat proche mais en effectuant un zoom sur les courbes, on remarque que le filtre de Hanning est meilleur comme le montre la FIGURE 3.11. De plus, le filtre de Hanning permet de regrouper plusieurs variations du signal en un seul pic ce qui permet une interprétation émotionnelle plus efficace.

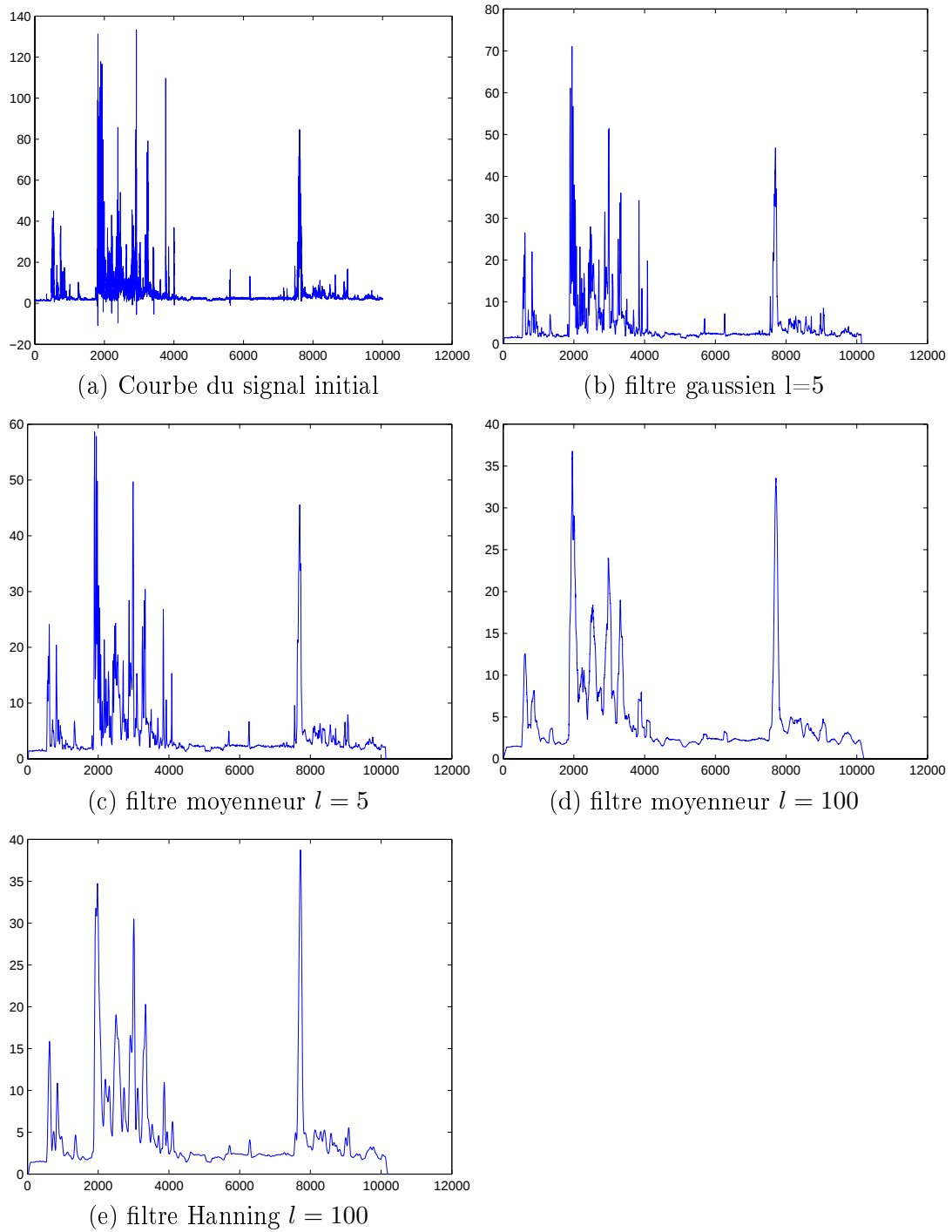


FIGURE 3.10 – Lissage du signal EMG avec les différents filtres : filtre moyenneur de longueur 5, filtre gaussien de longueur 5, filtre moyenneur de longueur 100 et filtre de Hanning de longueur 100

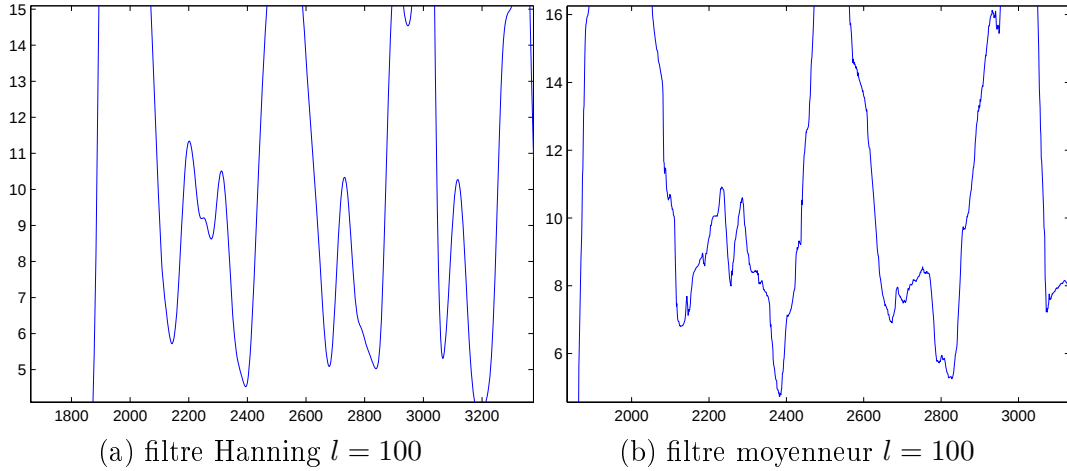


FIGURE 3.11 – Comparaison de la qualité de lissage avec le filtre de Hanning $l = 100$ et le filtre moyeneur $l = 100$. On remarque que la courbe (a) est bien lissée par-contre la deuxième courbe (b) présente des variations qui reflète l'existence du bruit dans le signal

3.4.3 Détection des pics

Une activité émotionnelle se traduit par des variations plus ou moins importantes sur les signaux physiologiques. Ces variations sont généralement limitées dans le temps et se caractérisent par des pics dans le signal. Elles traduisent généralement le passage d'un état stable à un état non stable. La fin de l'émotion est caractérisé par un retour à l'état d'origine ou un passage à une autre émotion. La détection d'une activité émotionnelle consiste donc à détecter, dans les signaux physiologiques, l'apparition des pics. Lorsque un pic est détecté nous isolons la partie du signal correspondante pour les traitements ultérieurs.

L'approche que nous proposons pour la détection des pics se base sur le calcul du gradient et ensuite sur la recherche des passages par zero du gradient. Soit F une fonction numérique dérivable sur un intervalle ouvert contenant au moins un pic (cf. FIGURE 3.12(a)). F admet un maximum local (resp. minumum local) en a , si $\text{grad } F(x) = 0$. Nous détectons ainsi comme le montre la FIGURE 3.12(b) 3 points vérifiant cette condition : deux minima (colorés en bleu) et un maximum (coloré en rouge). Pour détecter et isoler un pic il nous faut donc trouver dans l'ordre, un minima suivi d'un maxima suivi d'un minima. Nous isolons alors la partie du signal se trouvant entre les deux minima. En pratique et parce que nous manipulons des signaux discrets, il est rare de trouver des points où le gradient est exactement égal à 0. Nous cherchons alors des inversions du signe du gradient. Les inversions de signe passant d'un gradient positif à un gradient négatif nous donnent les maxima. Les changements passant d'un gradient négatif à un gradient positif nous donnent les minima. La FIGURE 3.13 montre le résultat de cette étape sur un signal EMG lissé. Les parties du signal colorées en rouge représentent les pics détectés et valides.

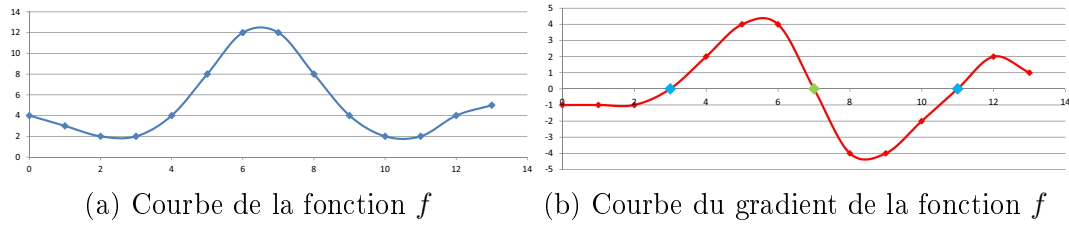


FIGURE 3.12 – Détection des maxima et minima locaux dans un signal donné

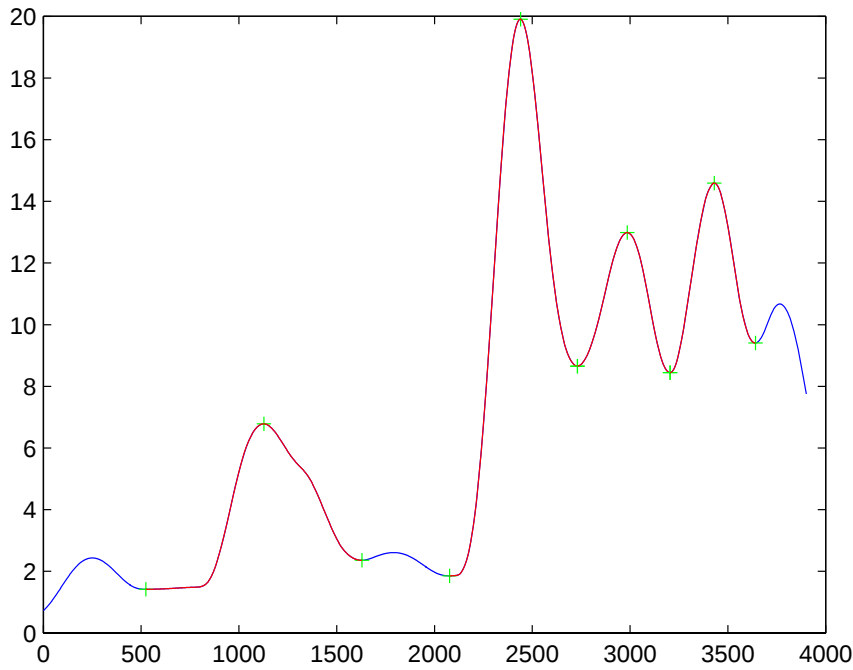


FIGURE 3.13 – Détection des pics dans un signal

3.4.4 Extractions des vecteurs caractéristiques

Après l'isolation des pics trouvés dans le signal nous générons pour chacun des pics un vecteur caractéristique le décrivant. En effet les pics des différentes émotions n'ont pas les mêmes formes et les mêmes caractéristiques. La FIGURE 3.14 montre les pics, extrait du signal EMG, caractéristiques de trois émotions différentes qui sont la colère, la joie et la haine. Nous remarquons par exemple que l'intensité du pic relatif à la colère est plus importante que les deux autres. La forme du pic peut donc être considérée comme un élément caractéristique qui peut définir les émotions ressenties. D'autres éléments caractéristiques peuvent être considérés comme la moyenne du pic, sa variance, la moyenne du gradient, etc. Pour chaque signal physiologique S , nous avons calculé 7 paramètres caractéristiques définis comme suit :

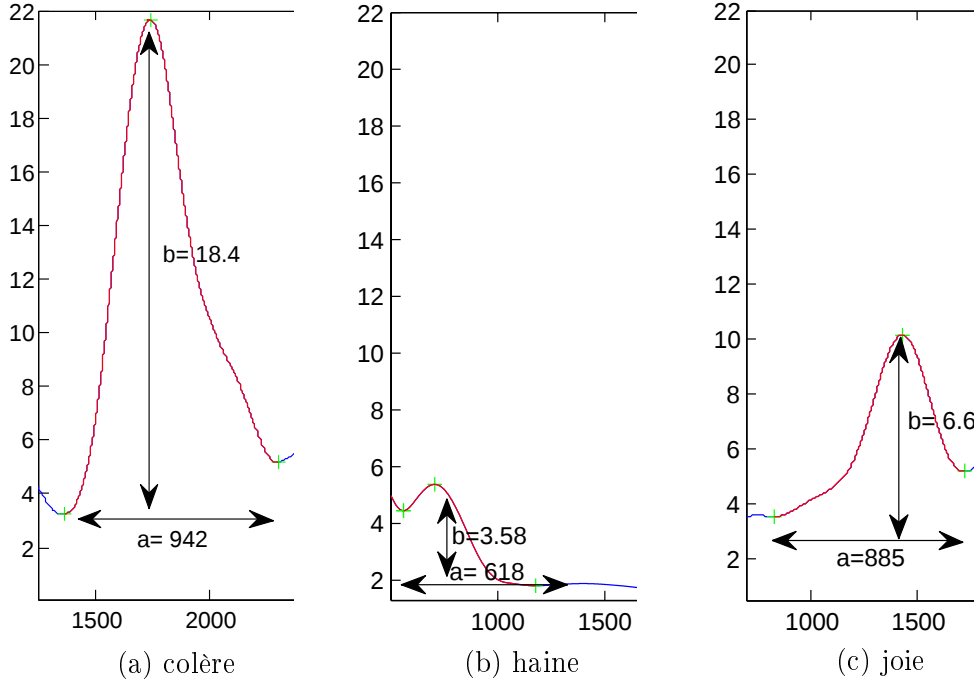


FIGURE 3.14 – Formes des pics relatives aux trois émotions : haine, colère et joie

La moyenne du signal ν_S :

$$\mu_S = \frac{1}{N} \sum_{n=1}^N S(t) \quad (3.3)$$

La moyenne du signal lissé $\widetilde{\mu}_S$:

$$\widetilde{\mu}_S = \frac{1}{N} \sum_{n=1}^N \widetilde{S}(t) \quad (3.4)$$

L'écart Type du signal σ_S :

$$\sigma_S = \sqrt{\frac{1}{T} \sum_{t=1}^T (S(t) - \mu_S)^2} \quad (3.5)$$

L'écart Type du signal lissé $\widetilde{\sigma}_S$:

$$\widetilde{\sigma}_S = \sqrt{\frac{1}{T} \sum_{t=1}^T (\widetilde{S}(t) - \widetilde{\mu}_S)^2} \quad (3.6)$$

La moyenne du gradient du signal lissé δ_x :

$$\delta_x = \frac{1}{T-1} \sum_{t=1}^{T-1} |\widetilde{S}(t+1) - \widetilde{S}(t)| = \frac{1}{T-1} (s_T - s_1) \quad (3.7)$$

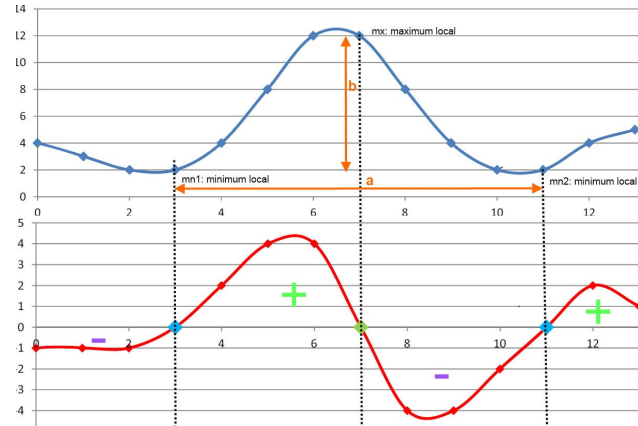


FIGURE 3.15 – Calcul de la largeur et de la longueur d'un pic

La largeur du pic a :

$$a = |t_{mn2} - t_{mn1}| \quad (3.8)$$

La largeur du pic b :

$$b = |S(t_{mx}) - \min(S(t_{mn1}), S(t_{mn2}))| \quad (3.9)$$

avec $mn1$, $mn2$ des minima locaux et mx un maximum local (cf. FIGURE 3.15).

3.4.5 Création de la base d'apprentissage

Après l'extraction des vecteurs caractéristiques, nous créons les bases d'apprentissage qui seront utilisées lors de l'étape de détection. Nous créons une base pour chaque modalité. Nous obtenons donc à la fin quatre bases différentes. Chacune de ces bases est formée par 240 vecteurs. En effet, nous générons 30 vecteurs pour chaque émotion ce qui donne 240 vecteurs au total. La création de la base a été effectuée en utilisant 40% des signaux disponibles pour chaque modalité. Les 60% restants seront utilisés lors de l'étape de test.

3.5 Détection

La deuxième étape de la méthode proposée, est l'étape de détection. Cette étape est "online", elle peut se faire directement en temps réel lors de l'acquisition des données ou bien à la fin de l'acquisition. Le module détection est composé de deux grandes parties qui sont :

- l'extraction des vecteurs caractéristiques
- la classification

Afin d'extraire les vecteurs caractéristiques, nous effectuons les mêmes étapes qu'on a appliquées lors de l'apprentissage, à savoir : lissage, détection des pics et extraction des vecteurs caractéristiques. A contrario, nous ne disposons pas ici de l'information à priori sur l'émotion.

La deuxième partie de l'étape de détection est la classification automatique des émotions. Elle se fait en utilisant un classifieur appris sur des données d'apprentissage. Le but de cette étape est de prédire la classe ou l'étiquette émotionnelle du vecteur caractéristique calculé à l'étape précédente en utilisant notre base d'apprentissage, c'est-à-dire l'ensemble des vecteurs caractéristiques regroupés par classe émotionnelle.

Dans notre travail, nous avons opté pour l'algorithme de classification des k -plus proches voisins (KNN). Nous avons choisi cette classification en raison de sa simplicité et de son efficacité de calcul [Beyer 1999]. L'algorithme classique de KNN consiste à faire un calcul de distance entre tous les individus et la nouvelle observation x à classer. La classe ou l'étiquette de x est déterminée en fonction de la classe majoritaire parmi ses k plus proches voisins (cf. section 1.4.5.1).

Nous proposons d'utiliser KNN en appliquant une classification statistique. C'est-à-dire, nous prenons en considération le degré d'appartenance d'une nouvelle observation de données de test x , l'algorithme de k -plus proches voisins donne les k voisins les plus proches dans la base d'apprentissage. Soit $N_k(x)$, l'ensemble des k plus proches voisins de x , et $c(z)$, l'étiquette de la classe émotionnelle z . La cardinalité de $N_k(x)$ est égale à k . On définit, alors le sous-ensemble $N_k^e(x)$, incluant les plus proches voisins appartenant à une classe donnée (e) avec :

$$(e) \in \{aucune\ motion, col\grave{e}re, haine, peine, amour\ platonique, amour\ romantique, joie, v\acute{e}n\acute{e}ration\}$$

comme étant :

$$N_k^e(x) = \{z \in N_k(x), c(z) = e\} \quad (3.10)$$

En divisant chaque valeur de $N_k^e(x)$ par k , Nous obtenons les probabilités d'appartenance à chaque classe d'émotion. Cette nouvelle valeur sera notée par $n_k^e(x) \in [0, 1]$, et le résultat de classification sera défini comme une combinaison linéaire des classes émotionnelles :

$$e^* = \sum \langle n_k^e(x), e \rangle e \quad (3.11)$$

donc

$$e^* = n_k^{aucune\ \acute{e}motion}(x) aucune\ \acute{e}motion + n_k^{col\grave{e}re}(x) col\grave{e}re + \dots + n_k^{joie}(x) joie + n_k^{v\acute{e}n\acute{e}ration}(x) v\acute{e}n\acute{e}ration. \quad (3.12)$$

Ainsi, nous créons un modèle de probabilité pour chaque classe émotionnelle, où $n_k^e(x)$ représente la probabilité de la classe émotionnelle correspondante. Par exemple : si $k = 10$, et parmi les 10 plus proches voisins, nous avons 8 éléments appartenant à la classe émotionnelle *colère* et 2 éléments appartenant à la classe émotionnelle *peine*, nous obtenons une probabilité égale à 0,8 pour la classe émotionnelle *colère* ($n_{10}^{anger}(x) = 0,8$) et 0,2 pour la classe émotionnelle *peine* ($n_{10}^{peine}(x) = 0,2$). Ainsi la classification finale est définie par $(e^*) = 0.8anger + 0.2grief$. Ainsi la méthode proposée pour la reconnaissance des émotions permet de fournir une information

complète et détaillée sur l'émotion détectée, en donnant la probabilité d'appartenance à chaque classe émotionnelle. Cette représentation peut être transformée par la suite en un vecteur exprimé dans la base de notre modèle vectoriel multidimensionnel défini dans la section 2.3 en utilisant une matrice de passage [Paugam 2007] et les règles de combinaisons des émotions de Plutchik. Ces données émotionnelles peuvent être ensuite annotées pour effectuer des échanges entre différents utilisateurs.

3.6 Expérimentation et résultats

Dans cette partie nous allons évaluer notre méthode de reconnaissance des émotions et la comparer par rapport aux autres méthodes de l'état de l'art. Pour cela nous allons affecter à une nouvelle observation x la classe majoritaire parmi les k plus proches voisins de l'observation. Autrement dit, nous allons garder du vecteur de probabilité, que nous obtenons en appliquant knn, l'émotion qui possède la probabilité la plus élevée.

Nous allons commencer cette partie par valider tout d'abord le choix des paramètres de la méthode de détection proposée. En effet, nous allons tester la méthode proposée en modifiant les deux principaux paramètres :

- la taille de la fenêtre de lissage
- la valeur de k , qui représente le nombre des voisins à prendre en compte dans l'algorithme de classification KNN

Le choix des valeurs de ces paramètres se fait à travers plusieurs essais et évaluation pour atteindre les performances optimales. Nous calculons à chaque fois la précision (le taux de reconnaissance) et nous retenons à la fin les valeurs qui la maximisent. La précision représente le rapport entre le nombre de données correctement classées et le nombre total des données testées. L'équation suivante donne la formule utilisée pour calculer la précision :

$$P = \frac{1}{N} \sum_{i=1}^N L(y_i, f(x_i)) \quad (3.13)$$

avec :

$$L = \begin{cases} 1 & \text{si } y_i = f(x_i) \\ 0 & \text{sinon} \end{cases}$$

Généralement, la précision est donnée sous forme de pourcentage ce qui nécessite de multiplier l'équation précédente par 100.

Les tests ont été effectués en utilisant la base de données fournis par MIT. Cette base de données a été divisée en deux parties. La première partie, représentant 40% de la base, a permis la création de notre base d'apprentissage. La deuxième partie, représentant les 60% restants, est utilisée pour les tests. Toutes les fonctions de notre méthode ont été programmées en utilisant Matlab.

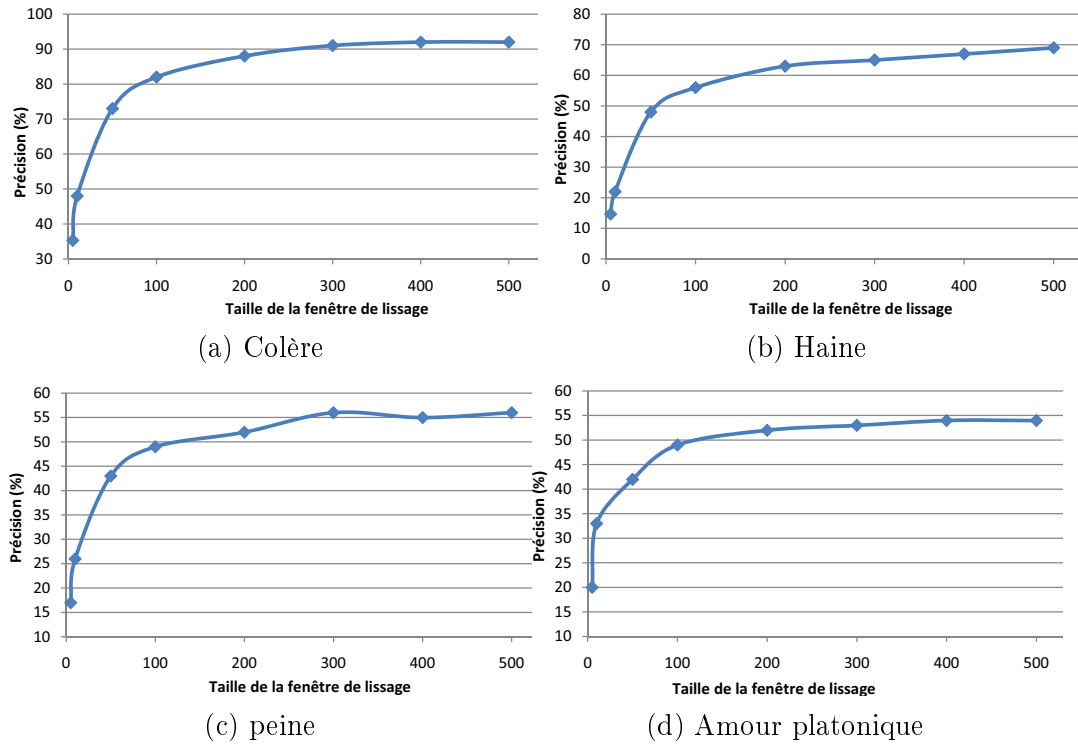


FIGURE 3.16 – Courbes de précision relatives aux émotions : amour platonique, colère, haine et peine, en utilisant les signaux physiologique EMG en fonction de la taille de la fenêtre de lissage

3.6.1 Choix de la taille de la fenêtre de lissage

Le choix de la taille de la fenêtre de lissage est très délicat. Il influe fortement, comme nous allons le voir, sur l'efficacité de la détection. Afin de choisir au mieux la taille de cette fenêtre, nous optons pour une méthode empirique, qui consiste à tester diverses fenêtres de lissage et à examiner le résultat pour trouver la valeur optimale. Pour nos tests, nous fixons la valeur de k à 10 (k représente le nombre des voisins à prendre en compte dans l'algorithme de classification KNN). Nous varions ensuite la valeur de la fenêtre de lissage n . Les valeurs que nous allons tester sont : $n = 10$, $n = 50$, $n = 100$, $n = 200$, $n = 300$, $n = 400$ et $n = 500$. Nous calculons à chaque fois les précisions correspondantes aux huit émotions. Enfin, nous allons calculer la précision globale de classification de notre méthode.

La FIGURE 3.16 montre l'évolution de la précision en fonction de la taille de la fenêtre de lissage utilisée pour les émotions : amour platonique, colère, haine, peine. Les signaux physiologiques utilisés sont relatifs à la modalité EMG. Nous remarquons que pour ces quatre émotions l'allure des courbes est la même. En effet, la précision au début s'accroît rapidement en augmentant la taille de la fenêtre de lissage et elle se stabilise à partir d'une fenêtre de dimension $n = 300$. A partir de cette valeur, le gain en précision devient faible. Par exemple, en passant de $n = 300$

TABLE 3.2 – Précisions obtenues en faisant varier la taille de la fenêtre de Hanning

	$n = 10$	$n = 50$	$n = 100$	$n = 200$	$n = 300$	$n = 400$	$n = 500$
aucune émotion	19%	60%	70%	79%	82%	83%	83%
colère	48%	73%	82%	88%	91%	92%	92%
haine	22%	48%	56%	63%	65%	67%	69%
peine	26%	43%	49%	52%	56%	55%	56%
amour platonique	33%	42%	49%	52%	53%	54%	54%
amour romantique	26%	39%	47%	51%	54%	53%	53%
joie	29%	41%	48%	52%	55%	53%	54%
vénération	33%	49%	63%	76%	79%	83%	82%

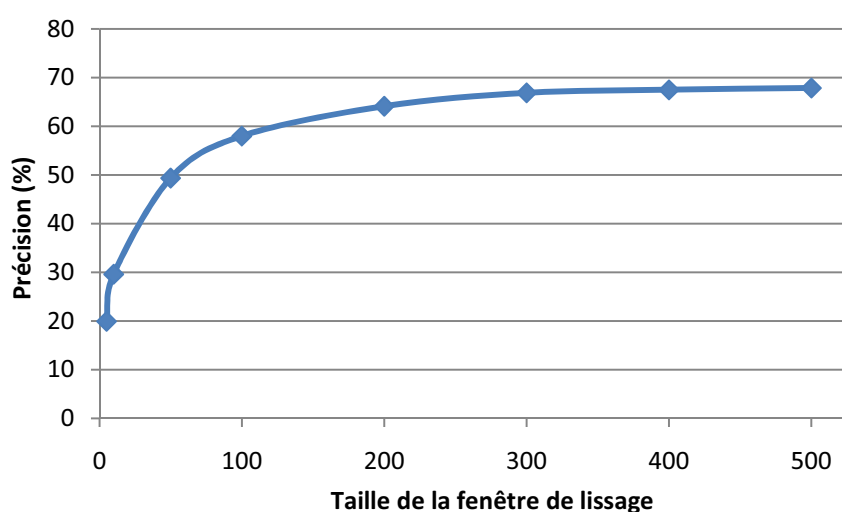


FIGURE 3.17 – précision globale de classification en fonction de la taille de la fenêtre de lissage

à $n = 500$ la précision de la détection de l'émotion colère n'augmente que de 1%.

Le tableau 3.6.1, donne l'évolution de la précision pour les huit émotions détectées. Nous remarquons que l'augmentation de la précision varie d'une émotion à une autre. En effet elle est égale par exemple à 47% pour la haine en passant de 22% à 69% et de 30% pour la peine en passant de 26% à 56%. Cette différence n'est pas liée à la taille de la fenêtre de lissage mais plutôt à la méthode de détection.

La FIGURE 3.17 montre la précision globale de classification en fonction de la taille de la fenêtre de lissage. Cette courbe a la même forme que les courbes précédentes, que nous avons tracées pour chacune des émotions, avec une augmentation rapide de la précision au début et une stabilisation de cette valeur à partir d'une fenêtre de lissage de taille $n = 300$. A partir de cette courbe, nous pouvons choisir facilement la valeur optimale de n pour la modalité EMG. En effet, une valeur entre 400 et 500 donnera une précision maximale pour cette modalité.

Traçons maintenant les mêmes courbes pour la modalité RESPIRATION pour

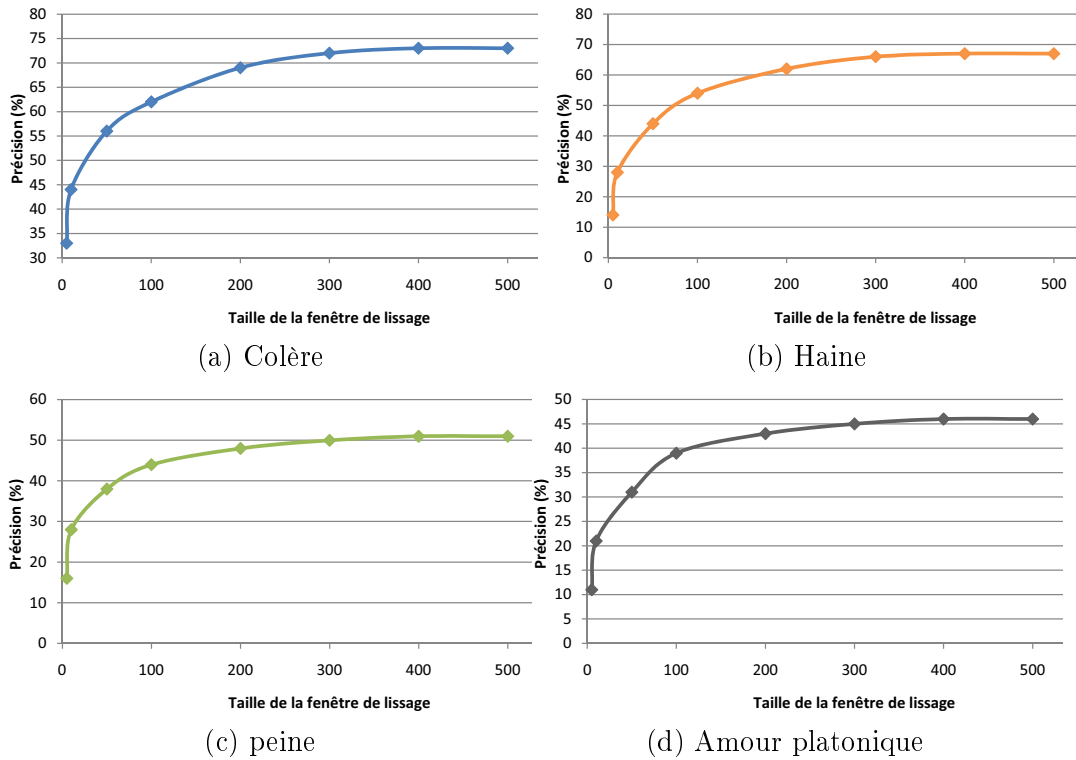


FIGURE 3.18 – Courbes de précision relatives aux émotions : amour platonique, colère, haine et peine, en utilisant les signaux physiologiques EMG en fonction de la taille de la fenêtre de lissage

vérifier si ce choix est aussi valide pour les autres modalités ou non. La FIGURE 3.18 montre l'évolution de la précision en fonction de la taille de la fenêtre de lissage utilisée pour les émotions : amour platonique, colère, haine, peine pour cette modalité. Nous remarquons aussi que pour ces quatre émotions l'allure des courbes est la même.

Nous remarquons aussi la même chose pour la FIGURE 3.19 qui montre la précision globale de classification en fonction de la taille de la fenêtre de lissage. À partir de cette courbe nous pouvons dire que la valeur optimale de n pour la modalité RESPIRATION peut être aussi une valeur comprise entre 400 et 500. Dans la suite des tests nous fixons la valeur de n à 500.

3.6.2 Choix de la valeur k de l'algorithme KNN

Pour le choix de la valeur de la variable k de l'algorithme KNN, nous allons effectuer le même type de test. Nous allons faire varier k en calculant à chaque fois la précision de classification. La FIGURE 3.20 montre, pour la modalité EMG, l'évolution de la précision en faisant varier k pour les trois émotions : colère, haine et peine. Nous remarquons que pour ces trois émotions, la précision ne varie sensi-

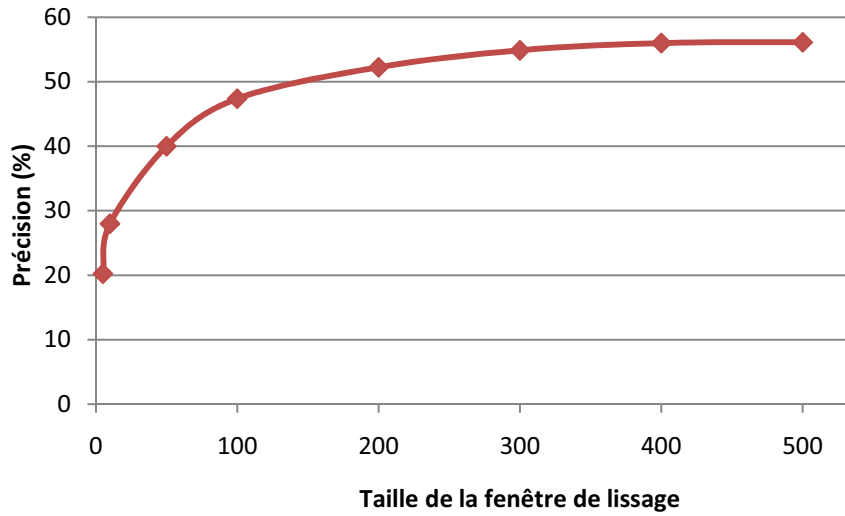


FIGURE 3.19 – précision globale de classification en fonction de la taille de la fenêtre de lissage

TABLE 3.3 – Précisions obtenues en faisant varier la valeur k de l'algorithme KNN

	$k = 4$	$k = 8$	$k = 10$	$k = 20$
aucune émotion	81%	82%	83%	83%
colère	89%	90%	92%	92%
haine	68%	68%	69%	69%
peine	55%	55%	56%	56%
amour platonique	54%	54%	54%	54%
amour romantique	53%	53%	53%	53%
joie	54%	54%	54%	54%
vénération	81%	81%	82%	82%

blement qu'en fonction de k .

Le tableau 3.6.2 montre, pour la modalité EMG, les précisions de classification obtenues pour les différentes émotions en choisissant $k = 4, 8, 10$ et 20 . Nous remarquons que la variation de la précision de classification est très faible pour les différentes valeurs de k et ne varie plus à partir de $k = 10$. Cette faible amélioration de la précision en augmentant la valeur de k est due à la réduction de l'effet des « outlier » (Donnée aberrante). En effet, plus la valeur de k est grande plus l'effet des « outlier » sur le résultat final est faible.

La FIGURE 3.21 montre l'évolution de la précision générale de classification en fonction de k pour les modalités RESPIRATION, GSR et BVP. Nous remarquons que, comme pour la modalité EMG, la précision pour ces 3 modalités ne varie sensiblement qu'en fonction de k . Cette variation est de l'ordre de un pour cent.

Le choix de k n'a pas une grande influence ici sur la précision globale de classification car, comme nous l'avons mentionné au début de cette partie de tests, nous

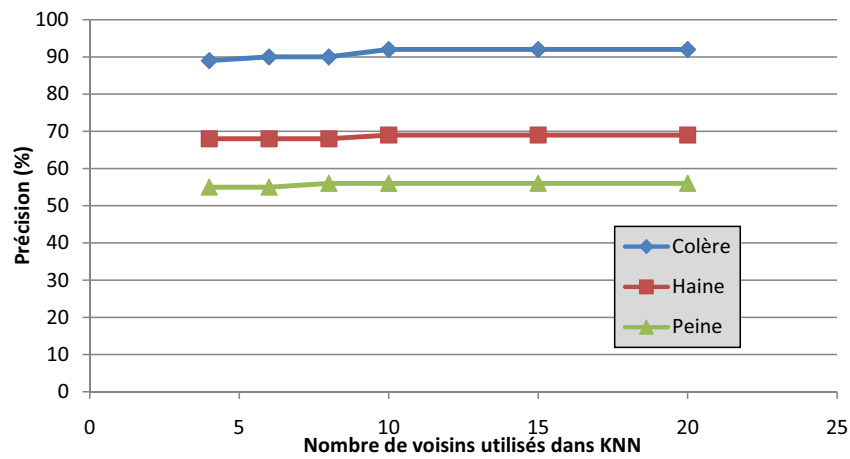


FIGURE 3.20 – Précision globale de classification pour la modalité EMG en fonction de la valeur de k

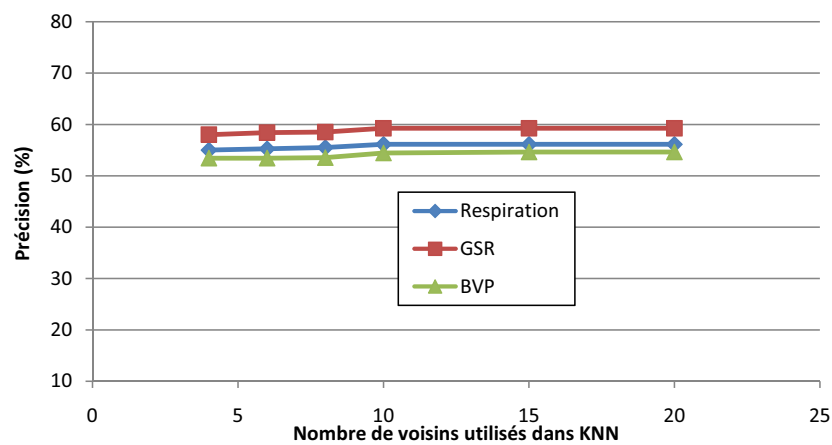


FIGURE 3.21 – Précision globale de classification pour les modalités RESPIRATION, GSR et BVP en fonction de la valeur de k

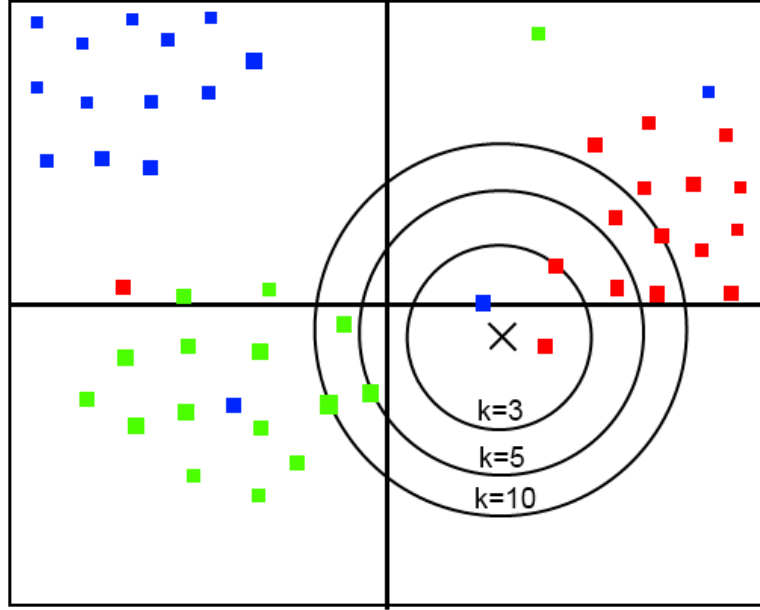


FIGURE 3.22 – Exemple de recherche utilisant KNN avec $k = 3$, $k = 5$, $k = 10$. Les points rouges représentent la classe colère, les points bleus la classe joie et les points verts la classe haine.

allons garder du vecteur émotionnel l'émotion avec la probabilité la plus élevée. Mais si nous gardons la représentation vectorielle, le choix de la bonne valeur de k devient nécessaire. En effet, la FIGURE 3.22 montre un exemple de recherche en utilisant KNN avec $k = 3$, $k = 5$, $k = 10$. Les points rouges représente la classe colère, les points bleus la classe joie et les points verts la classe haine. On obtient alors les 3 vecteurs suivants notés V_3 , V_5 et V_{10} :

$$V_3 = \begin{pmatrix} 0 \\ 0.66 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0.33 \end{pmatrix}, V_5 = \begin{pmatrix} 0 \\ 0.6 \\ 0.2 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0.2 \end{pmatrix}, V_{10} = \begin{pmatrix} 0 \\ 0.6 \\ 0.3 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0.1 \end{pmatrix}$$

Si pour l'interprétation des ces 3 vecteurs, on ne gardait que l'émotion avec la probabilité la plus élevée, le résultat serait le même pour les 3 vecteurs c'est-à-dire l'émotion colère. Par contre si on adopte une analyse vectorielle le résultat sera différent. En effet, pour $k = 3$, nous allons interpréter le vecteur comme émotion complexe combinant avec différents pourcentages la colère et la joie. Ce qui est faux car le représentant de la joie, ici choisi, est un « outlier ». Pour $k = 5$, nous allons interpréter le vecteur comme étant l'émotion colère. Enfin, pour $k = 10$, nous

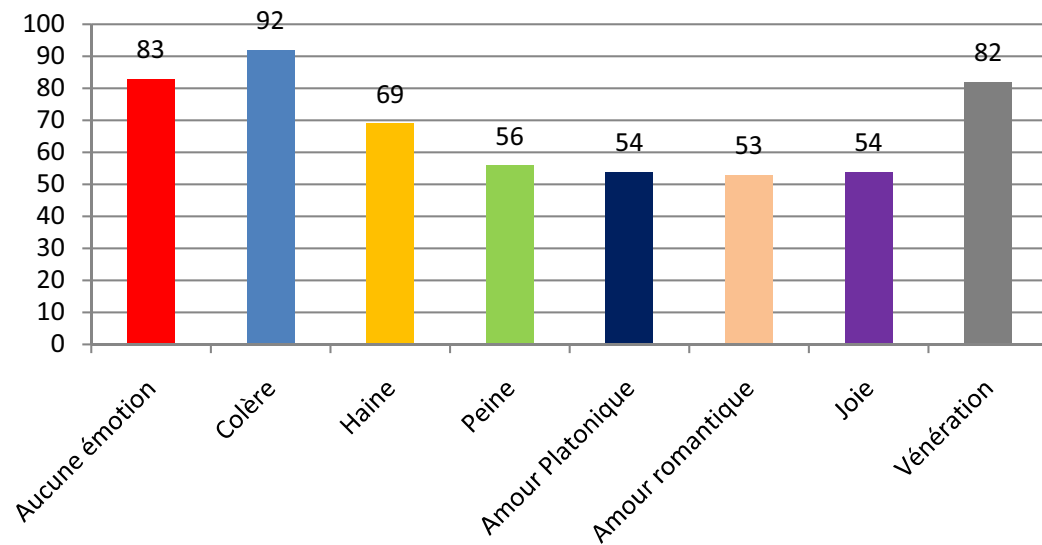


FIGURE 3.23 – Taux de reconnaissance finaux de la méthode de détection proposée pour la modalité EMG.

allons interpréter le vecteur comme émotion complexe combinant avec différents pourcentages la colère et la haine ce qui est plus proche de la vérité. Une valeur plus grande de k par exemple 50 mènera à une fausse interprétation car elle donnera les 3 classes : colère, haine et joie avec le même pourcentage ce qui est faux.

En général, les grandes valeurs de k réduisent l'effet des « outlier » sur la classification et donc le risque d'erreur, mais rendent les frontières entre classes moins distinctes. Le meilleur choix de k pour notre méthode de détection serait une valeur comprise entre 8 et 12.

3.6.3 Résultats généraux et comparaison

Les FIGURES 3.23, 3.24, 3.25, 3.26 montrent, pour les quatre modalités EMG, RESPIRATION, GSR et BVP, les résultats finaux de la méthode de reconnaissance d'émotions que nous proposons. Nous avons fixé la taille de la fenêtre de lissage à $n = 500$ et le nombre de voisins k de l'algorithme KNN à 10. Nous remarquons que les taux de reconnaissance varient d'une modalité à une autre. Les meilleurs résultats ont été obtenus pour la modalité EMG avec un taux de reconnaissance moyen de 68% contre 56% pour la RESPIRATION, 55% pour BVP et 59% pour GSR.

Chacune des quatre modalités permet de détecter, avec des taux de reconnaissance élevés, des émotions différentes. En effet, notre méthode avec la modalité EMG permet par exemple de détecter facilement les deux émotions colère et vénération avec des taux respectifs de 92% et 83%. Tandis que pour la modalité BVP, elle permet de détecter l'amour platonique avec un taux de 69%. Nous pouvons remarquer aussi qu'en utilisant la modalité GSR, notre méthode permet de reconnaître avec un taux de 83% l'émotion haine. Cette propriété est très importante car elle

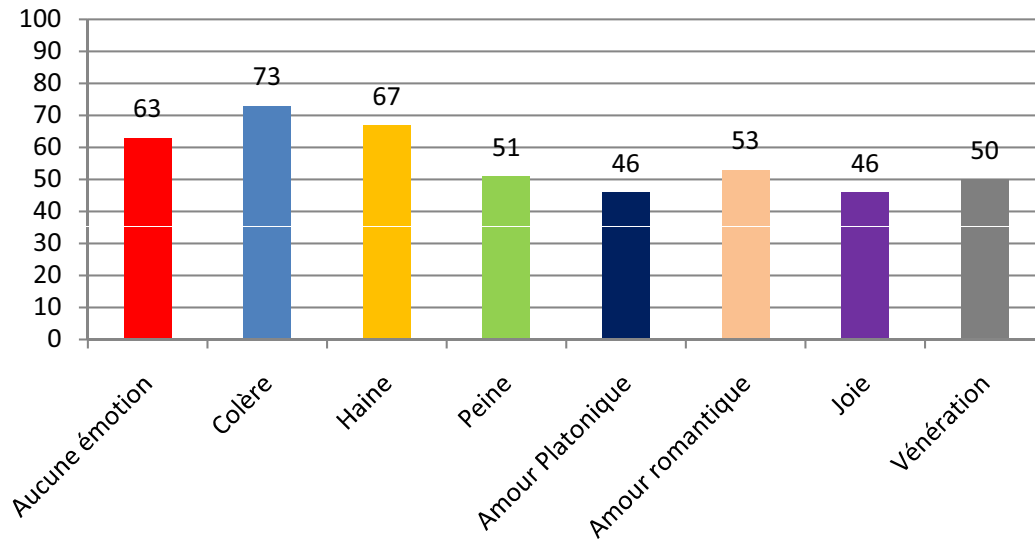


FIGURE 3.24 – Taux de reconnaissance finaux de la méthode de détection proposée pour la modalité RESPIRATION.

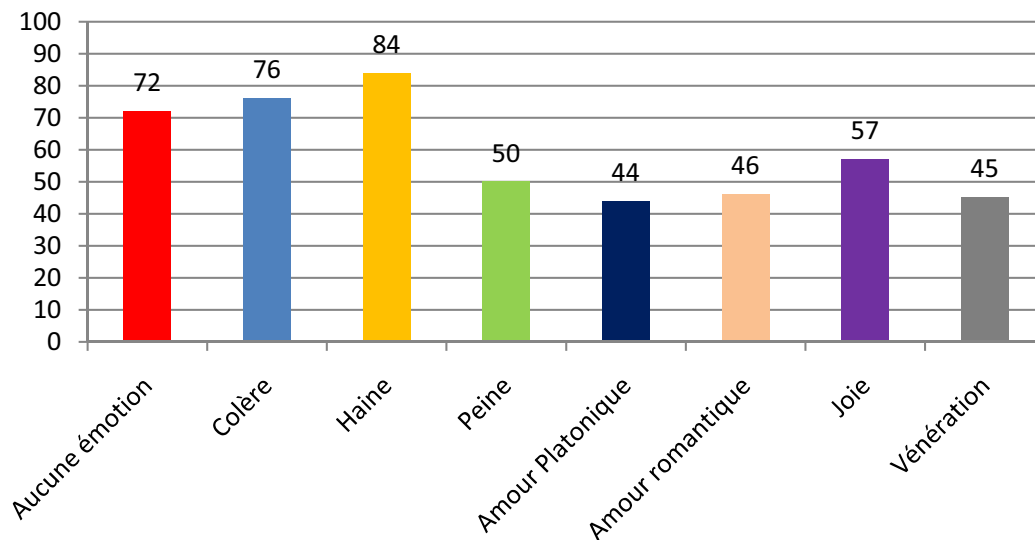


FIGURE 3.25 – Taux de reconnaissance finaux de la méthode de détection proposée pour la modalité GSR.

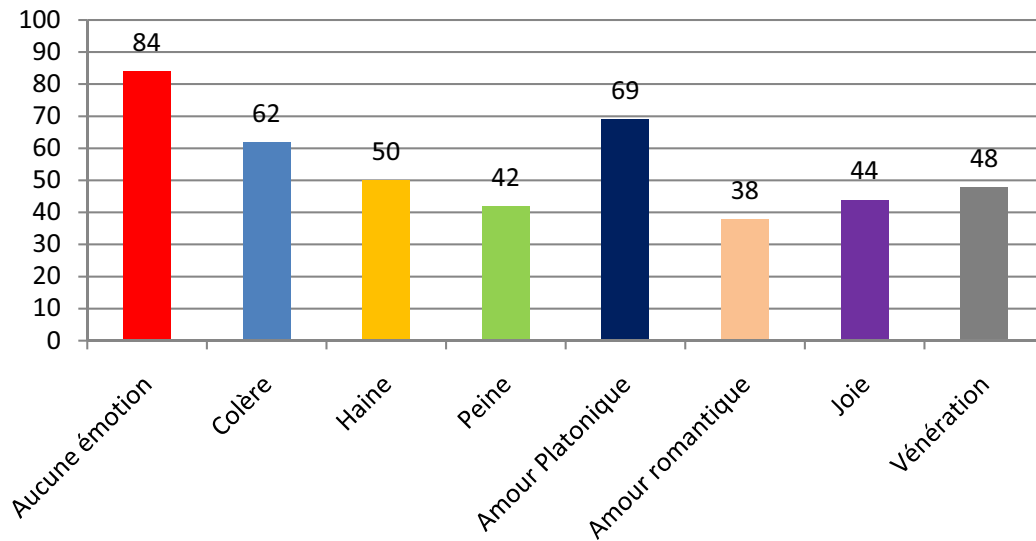


FIGURE 3.26 – Taux de reconnaissance finaux de la méthode de détection proposée pour la modalité BVP.

permettra comme nous allons le voir dans le chapitre suivant d'améliorer les taux de reconnaissance des différentes émotions en adoptant une approche multimodale combinant les résultats de détection de chacune de ces quatre modalités.

Le tableau 3.6.3 compare les taux de classification de la méthode proposée avec les différentes méthodes de l'état de l'art pour les quatre modalités EMG, RESPIRATION, GSR et BVP.

Nous remarquons que pour la modalité GSR, la méthode que nous proposons donne le taux de reconnaissance le plus élevé avec 59.25% contre 30% pour la méthode Fusion based HHT features [Zong 2009]. C'est à dire que notre méthode donne un résultat 2 fois plus précis que la méthode citée précédemment. Pour la modalité EMG, nous obtenons aussi le taux de reconnaissance le plus élevé avec une précision de 67.8%. Ce résultat dépasse les autres méthodes qui sont la méthode Baseline [Muthusamy 2012], la méthode fission based HHT (HilbertHuang transform) features [Zong 2009] et la méthode Fusion based HHT features) en améliorant le taux de reconnaissance de plus de 15%. Enfin pour la modalité RESPIRATION, nous obtenons un taux de reconnaissance de 56.12% qui est proche de la valeur de la méthode Fission based HHT features qui donne le meilleur taux de reconnaissance de 60%. Mais nous remarquons que notre méthode dépasse en performance (la moyenne du taux de reconnaissance) les autres méthodes avec un taux de reconnaissance de 59.44% contre 44.33% pour la méthode Fusion based HHT features, 50.33% pour la méthode Fission based HHT features et 45% pour la méthode Baseline. La FIGURE 3.27 montre la moyenne globale de taux de reconnaissance de toutes les modalités. Nous obtenons une valeur de 75.75% pour la colère, 76.5% pour la haine et 56.25% pour la vénération.

TABLE 3.4 – Tableau comparatif du taux de classification des différents méthodes

Différentes méthodes	GSR	EMG	RESPIRATION	BVP	La moyenne du taux de reconnaissance
Baseline	42%	39%	42%	-	45%
Fission based HHT features	39%	52%	60%	-	50.33%
Fusion based HHT features	30%	44%	59%	-	44.33%
Méthode proposée	59.25%	67.8%	56.12%	54.62%	59.44%

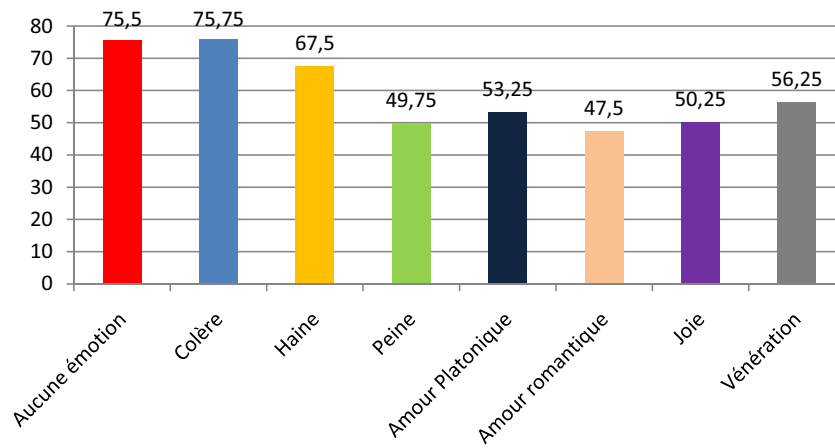


FIGURE 3.27 – La moyenne globale du taux de reconnaissance des modalités pour toutes les émotions

3.7 Conclusion

Dans ce chapitre nous avons présenté une approche de reconnaissance automatique des émotions en utilisant les signaux physiologiques. Notre approche se base sur l'application des techniques de traitement de signal et utilise le modèle multidimensionnel pour la représentation des émotions décrit dans le chapitre précédent . La démarche de notre approche consiste à extraire les informations caractéristiques des signaux pour chaque classe émotionnelle, suivi par une étape de classification basée sur la méthode des K plus proches voisins KNN. Les résultats expérimentaux montrent que nous avons atteint une moyenne du taux de reconnaissance de 59.44% (tableau 3.6.3) pour les huit émotions. Afin d'améliorer ce taux de reconnaissance, nous allons procéder à une reconnaissance multimodale. Le chapitre suivant fait l'objet d'une approche multimodale pour la reconnaissance des émotions.

Reconnaissance multimodale

Sommaire

4.1	Introduction	87
4.2	Fusion de données	88
4.2.1	Fusion de descripteurs	88
4.2.2	Fusion de décisions	88
4.3	Approche proposée pour la reconnaissance multimodale des émotions	90
4.3.1	Résultats	91
4.4	Exemple d'application : Détection de la dépression	94
4.4.1	Introduction	94
4.4.2	Présentation du projet Gerhome	96
4.4.3	La Dépression	97
4.4.4	Méthode de prévention de la dépression	98
4.5	Conclusion	100

4.1 Introduction

Bien que de nombreuses études ont exploré l'utilisation soit des expressions faciales [Black 1997, Tian 2000], ou de la parole [Clavel 2010] ou des signaux physiologiques [Healey 2000] pour détecter l'état émotionnel de l'homme, relativement moins d'efforts ont été concentrés sur la reconnaissance des émotions en utilisant deux ou plusieurs modalités [Sebe 2005].

L'utilisation d'une approche multimodale peut donner non seulement de meilleures performances mais également plus de robustesse lorsque l'une de ces modalités est acquise dans un environnement bruyé [Pantic 2003]. En effet, la fusion d'informations issues de différentes modalités permet d'améliorer l'estimation et la reconnaissance de l'état émotionnel des utilisateurs.

Dans ce chapitre nous essayons de prendre en compte la nature essentiellement multimodale des émotions. Nous proposons, donc une approche multimodale pour la reconnaissance des émotions capable de fusionner et intégrer des informations multiples issues de plusieurs modalités dans le but d'améliorer le taux de reconnaissance. La fusion d'informations multimodales permet de coupler des indices collectés à différents niveaux afin d'avoir une classification plus fine. Elle peut prendre différentes formes selon le moment où elle est effectuée. On distingue la fusion de descripteurs et

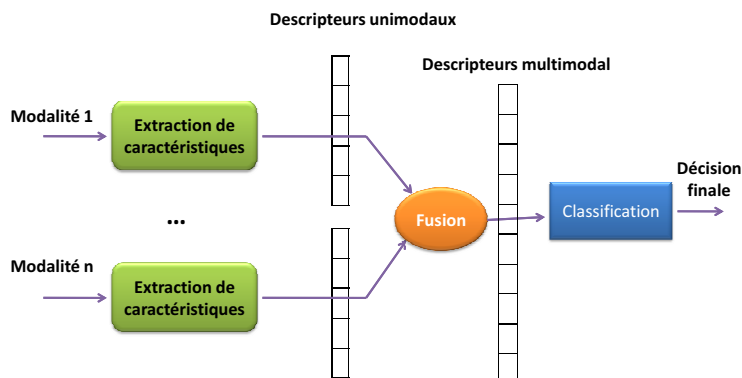


FIGURE 4.1 – Organisation générale d'un processus multimodal basé sur la fusion de descripteurs

la fusion de décisions. Pour mieux appréhender nos travaux nous allons tout d'abord donner un aperçu rapide sur ces deux approches.

4.2 Fusion de données

On distingue généralement deux types de fusion de données :

- la fusion de descripteurs, autrement dit fusion précoce (early fusion)
- la fusion de décisions, ou fusion tardive (late fusion).

dans la partie suivante nous allons présenter ces deux niveaux de fusion

4.2.1 Fusion de descripteurs

La fusion de descripteurs (cf. FIGURE 4.1), également appelée fusion précoce, permet d'obtenir, à partir de vecteurs de descripteurs extraits de chacune des modalités impliquées, un vecteur de caractéristiques multimodal. Elle consiste à synchroniser et à regrouper toutes caractéristiques extraites des différents dispositifs et selon les différents canaux de communication émotionnelle avant de les interpréter.

Plusieurs méthodes ont été proposées dans la littérature pour fusionner les descripteurs pour obtenir un vecteur multimodal. La Cascia et ses collègues [Cascia 1998], par exemple, ont utilisé la méthode la plus simple qui consiste à simplement concaténer les vecteurs unimodaux. Ayache [Ayache 2007], quant à lui, propose avant la concaténation une normalisation des vecteurs unimodaux en fonction de leur taille. D'autres méthodes, comme l'analyse en composantes principales ou l'analyse linéaire discriminante, permettent de sélectionner les caractéristiques les plus pertinentes.

4.2.2 Fusion de décisions

La fusion au niveau décision (cf. FIGURE 4.2) est une fusion des émotions et consiste à interpréter les caractéristiques de chaque canal séparément : on obtient

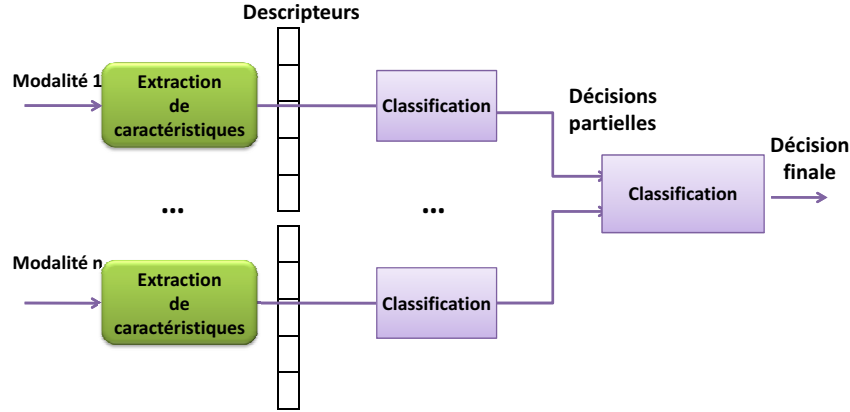


FIGURE 4.2 – Organisation générale d'un processus multimodal basé sur la fusion de décisions

ainsi autant d'émotions que de canaux étudiés [Clay 2009]. Les émotions ainsi obtenues doivent alors être synchronisées et amalgamées pour n'obtenir qu'une émotion finale. Le premier enjeu de la fusion tardive repose sur le choix de la méthode utilisée pour fusionner les résultats de plusieurs systèmes unimodaux. Plusieurs solutions ont été proposées dans la littérature comme la méthode de vote, les approches probabilistes comme les réseaux bayésiens.

4.2.2.1 Méthode de Vote

La méthode de fusion d'informations la plus simple est le principe du vote [Kittler 1998]. Notons $S_j(x) = i$ le fait que la modalité S_j attribue la classe C_i à l'observation x . Nous supposons ici que les classes C_i sont exclusives. A chaque source nous associons la fonction indicatrice :

$$M_i^j = \begin{cases} 1 & \text{si } S_j(x) = i \\ 0 & \text{sinon} \end{cases} \quad (4.1)$$

La combinaison des modalités s'écrit par :

$$M_k^E(x) = \sum_{j=1}^m M_k^j(x) \quad (4.2)$$

La décision finale consiste à la règle du vote majoritaire, c'est à dire choisir la décision par le maximum de modalités. Une autre méthode, appelée le vote majoritaire pondéré, ajoute des hypothèses et utilise une somme pondérée au lieu de combiner les réponses des sources par une somme simple. L'idée est d'ajouter un poids qui représente la fiabilité de chaque modalité pour une décision donnée. Ainsi, la combinaison des modalités s'écrit par :

$$M_k^E(x) = \sum_{j=1}^m \alpha_{jk} M_k^j(x) \quad (4.3)$$

où $\sum_{j=1}^m \sum_{k=1}^n \alpha_{jk} = 1$

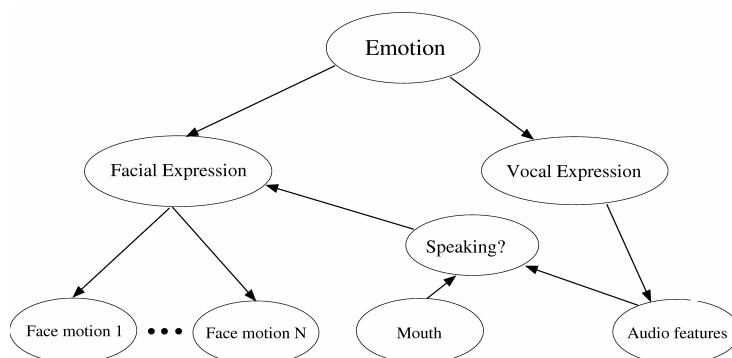


FIGURE 4.3 – La topologie d'un réseau Bayésien pour la reconnaissance bimodale des émotions [Sebe 2005]

4.2.2.2 Méthode probabiliste : les réseaux Bayésiens

Parmi les approches probabilistes, on trouve les réseaux Bayésiens, qui sont définis comme un mariage entre la théorie des probabilités et la théorie des graphes. Les probabilités permettent à ces modèles de prendre en compte l'aspect incertain présent dans les applications réelles. La partie graphique offre un outil intuitif et attractif dans de nombreux domaines d'applications où les utilisateurs ont besoin de "comprendre" le modèle qu'ils utilisent.

Sebe et al [Sebe 2005] ont proposé une topologie du réseau bayésien pour la reconnaissance des émotions à partir d'expression faciale et des informations auditives. La topologie du réseau combine les deux modalités d'une manière probabiliste. Le nœud du haut est la variable de la classe (l'expression émotionnelle reconnue). Elle dépend des expressions du visage reconnues, des expressions vocales reconnues et du contexte. La reconnaissance dépend également d'une autre variable qui indique si la personne parle ou non.

4.3 Approche proposée pour la reconnaissance multimodale des émotions

Dans la vie de tous les jours, l'être humain essaie de camoufler ou bien simuler son émotion. En effet, l'expression d'une émotion ne révèle pas toujours celle ressentie [Ochs 2006]. La plupart du temps, on feint une émotion particulière pour atteindre un but particulier, ou pour être conforme à certaines conventions socioculturelles (définies par Ekman comme des règles d'expression d'émotions [Friesen 2009]). Ainsi, nous distinguons les émotions ressenties appelées émotion déclenchées des émotions exprimées. D'après les théories cognitivistes, une émotion déclenchée est issue de l'évaluation subjective d'un événement significatif [Scherer 2000]. Un individu peut alors essayer de la supprimer, l'intensifier, l'atténuer, la masquer ou la remplacer pour exprimer une émotion consistante avec des règles conventionnelles d'expression d'émotions [Friesen 2009]. Par exemple, nous pouvons nous forcer à sourire

alors qu'on est triste. Nous pouvons également faire semblant d'apprécier une personne alors qu'on la méprise. Pour pouvoir déceler ce type d'émotion (masquée ou simulée) nous proposons une nouvelle approche multimodale basée sur notre modèle multidimensionnel défini dans le chapitre précédent 3.7. En effet, l'utilisation d'une seule modalité pour la reconnaissance des émotions ne donne pas de bons résultats dans le cas des émotions simulées ou masquées. C'est pour cela qu'avec l'approche multimodale proposée nous allons essayer de mieux cerner ce problème et proposer des éléments de solution pour la reconnaissance de ce type d'émotion. Elle permet de combiner les informations issues des différentes modalités afin d'améliorer la reconnaissance automatique des émotions.

L'approche proposée se base sur la combinaison des résultats issus de chaque modalité. Nous prenons ainsi en compte tous les aspects de l'émotion ressentie tout en réduisant l'incertitude sur l'information résultante. La combinaison des différents résultats se base sur une fusion au niveau décision (cf. FIGURE 4.2). Cette fusion permet d'interpréter les caractéristiques de chaque modalité séparément en utilisant les outils de la reconnaissance unimodale proposée dans le chapitre 3. Nous appliquons tout d'abord, comme le montre la FIGURE 4.4 une étape de détection monomodale sur chacune des modalités utilisées en se basant sur les différentes étapes décrites dans le chapitre précédent. Nous obtenons, ainsi, autant de vecteurs émotionnels que de modalités. Nous appliquons ensuite l'étape de fusion. La stratégie de fusion consiste à exploiter le modèle multidimensionnel avec tous les outils que nous avons déjà définis dans le chapitre 3.7. En effet, nous allons appliquer l'opérateur « addition » sur les vecteurs émotionnels obtenus à la fin de l'étape de détection monomodale. Nous obtenons alors directement le vecteur émotionnel résultant de la détection multimodale 2.4. La décision finale est obtenue en appliquant l'algorithme de distance proposé dans le chapitre 3.7. Nous remarquons que l'algorithme de détection multimodale a été simplifié grâce à l'utilisation du modèle formel pour la représentation des émotions proposé dans ce manuscrit. En effet, ce modèle, grâce aux outils mathématiques qu'il propose, permet d'analyser et de manipuler aisément les émotions. Nous pouvons noter aussi que notre méthode multimodale est indépendante du nombre et des types des modalités utilisées. Il supporte tout type de canal de communication humain comme les signaux physiologiques, les expressions faciales, les expressions gestuelles, la voix, etc.

4.3.1 Résultats

Dans cette partie, nous allons évaluer notre méthode de reconnaissance multimodale des émotions et la comparer à l'approche monomodale décrite dans le chapitre précédent. Puis, nous allons la comparer aux méthodes multimodales de l'état de l'art. La FIGURE 4.5 montre les résultats obtenus en effectuant une reconnaissance multimodale utilisant deux, trois et quatre modalités. Nous avons utilisé les modalités physiologiques du chapitre précédent qui sont : EMG, RESPIRATION, GSR et BVP. Nous remarquons que la moyenne du taux de reconnaissance des émotions s'améliore en augmentant le nombre de modalités. La

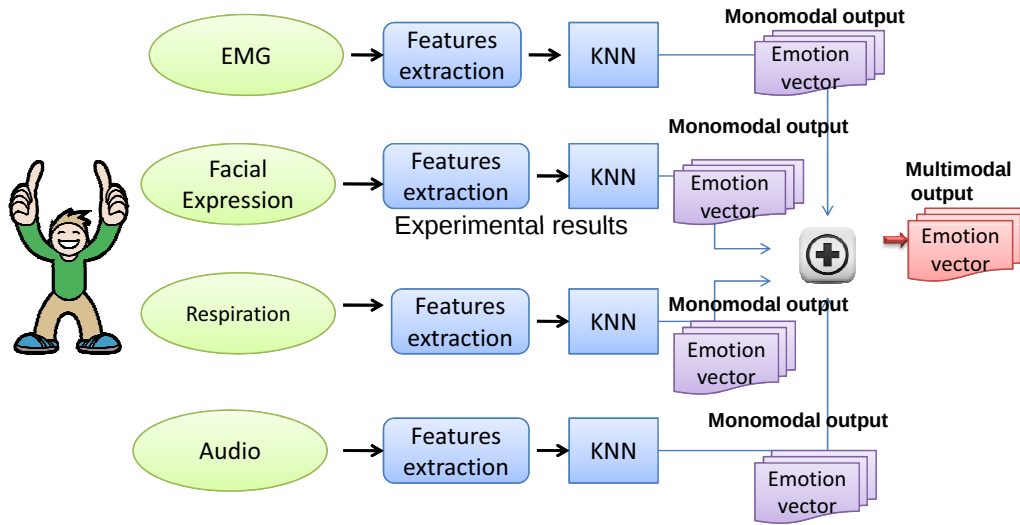


FIGURE 4.4 – Reconnaissance multimodale en utilisant l’addition du modèle multi-dimensionnel

comparaison des résultats conduit aux observations suivantes :

- La moyenne du taux de reconnaissance utilisant une seule modalité (RESPIRATION) est de 56.12%
- La moyenne du taux de reconnaissance utilisant deux modalités (bimodale : RESPIRATION+GSR) est de 61.125% qui est une amélioration de 8% par rapport à la méthode monomodale.
- L’ajout d’une nouvelle modalité RESPIRATION+GSR+BVP améliore la moyenne du taux de reconnaissance qui passe de 61.125% à 65.5%
- Enfin, la meilleure valeur de la moyenne du taux de reconnaissance est obtenue en utilisant quatre modalités (multimodale : RESPIRATION+GSR+BVP+EMG) avec une valeur de 72.125%, ce qui fait une amélioration de 28% par rapport à l’approche monomodale.

La FIGURE 4.6 montre le taux de reconnaissance monomodale en utilisant la modalité BVP et le taux de reconnaissance multimodale. Nous remarquons une amélioration considérable pour toutes les émotions. Par exemple, le taux de reconnaissance passe de 62% à 92% pour la colère et de 50% à 84% pour la haine. Le tableau 4.3.1 montre les taux de bonne classification pour différentes méthodes de l’état de l’art. Nous remarquons que notre méthode multimodale présente le meilleur résultat avec 72.125% de bonne classification en comparaison avec la méthode de Kim [Kim 2002] avec un taux de reconnaissance de 61.2% et la méthode Fusion based HHT features avec un taux de 62%. Par contre, nous pouvons remarquer que la méthode SFFS-FP de MIT présente le meilleur taux de classification avec un taux de 83%. Cependant, la méthode de MIT [Healey 2000], ne permet pas une détection instantanée de l’émotion, mais plutôt la détection d’émotion sur un intervalle de temps (scénario). Le principe de la méthode est de calculer les informations caractéristiques sur toute

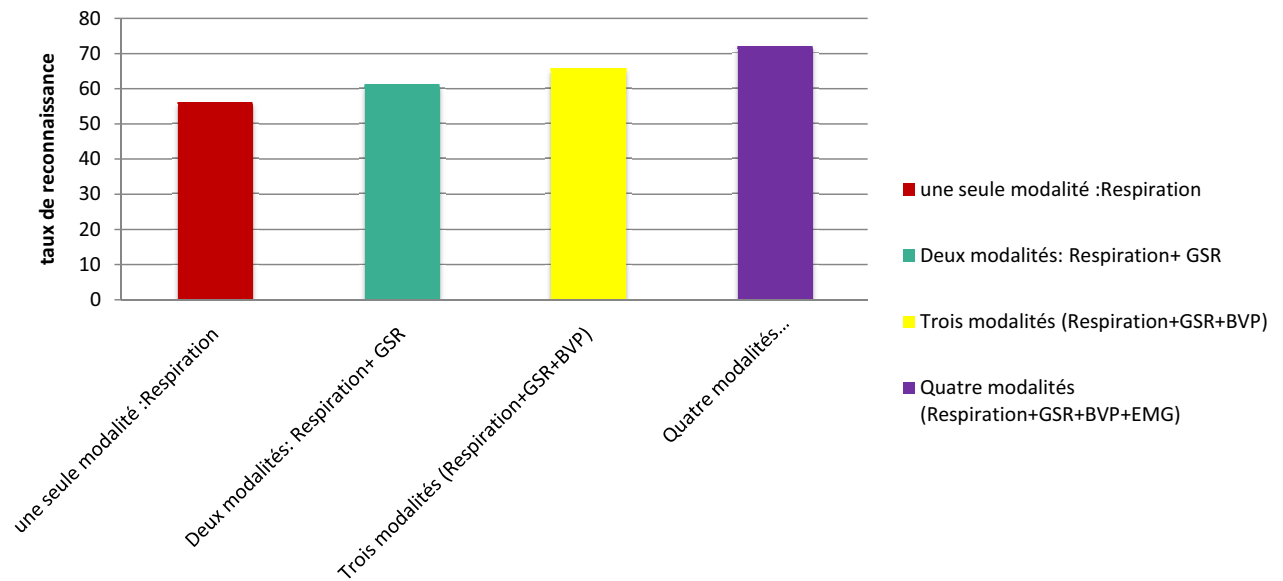


FIGURE 4.5 – Comparaison entre les résultats de l'approche monomodale et de l'approche multimodale

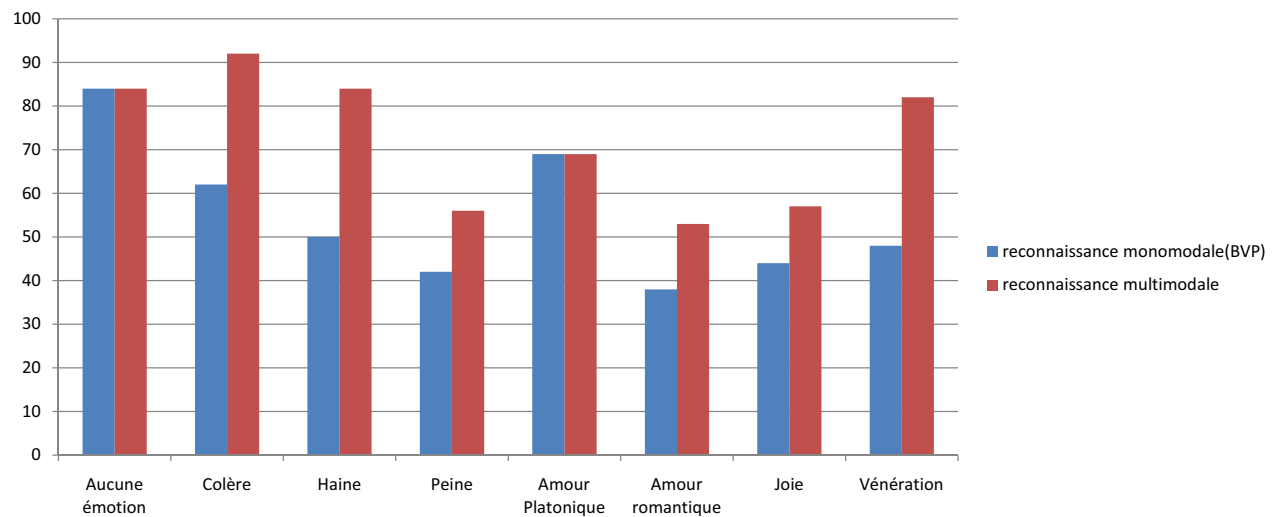


FIGURE 4.6 – Comparaison entre les résultats de l'approche monomodale et de l'approche multimodale

TABLE 4.1 – Tableau comparative des taux de classification des différents méthodes

Différentes méthodes	Taux de bonne classification
Methode de Kim	61.2%
Fusion based HHT features (fusion de 4 signaux physiologiques)	62%
SFFS-FP de MIT	83%
Baseline (fusion de 4 signaux physiologiques)	71%
Méthode proposée	72.125%

la durée du scénario d'une émotion donnée. Ils ont utilisé la méthode SFFS-FP pour réduire le nombre de données caractéristiques. Enfin, Healey et ses collègues [Healey 2000] ont utilisé 11 caractéristiques calculées sur une durée bien déterminée pour identifier les émotions (cf. tableau 4.3.1). En adoptant cette approche globale, ils peuvent détecter une émotion exprimé sur une longue période et non pas une reconnaissance de l'émotion en temps réel.

La FIGURE 4.7 montre un exemple d'un signal physiologique EMG formé par les huit émotions (aucune émotion, colère, haine, peine, amour platonique, amour romantique, joie et vénération) de durée égale (3 minutes chacune). L'évaluation se fait sur la durée totale d'une émotion donnée. Si on essaie d'analyser une partie à cheval entre l'émotion colère et l'émotion haine, la méthode de MIT, donnera un résultat erroné. En effet, en calculant les valeurs caractéristique du signal EMG entre la seconde 300 et 480, nous obtenons une valeur à-peu-près égale à 1.7 qui correspond à l'émotion joie.

4.4 Exemple d'application : Détection de la dépression

4.4.1 Introduction

Le modèle que nous avons proposé permet la représentation, la détection et l'analyse des émotions. Il peut être utilisé dans plusieurs domaines comme la robotique, les jeux vidéo interactifs ou même la surveillance des personnes dans les habitats intelligents. Le concept d'Habitat Intelligent est un domaine récent qui vise à redonner une vie autonome, dans leur domicile, à des personnes qui devraient normalement être placées dans des établissements spécialisés : patients souffrants de certaines maladies chroniques, handicapés, mais aussi personnes âgées dépendantes. Dans ce contexte, et dans le cadre du projet GERHOME [CSTB 2009] du CSTB de Sophia Antipolis, nous avons proposé un système de prévention de la dépression chez les personnes âgées. En effet, la dépression des personnes âgées est fréquente mais non diagnostiquée et donc non traitée dans 60 à 70% des cas . La dépression chez les personnes âgées passe souvent inaperçue parce que certains signes de dépression

TABLE 4.2 – les 11 valeurs caractéristiques extraites pour les 8 émotions du travail de MIT[Healey 2000]. μ_ϵ la moyenne du signal EMG, $\tilde{\mu}_g$ la moyenne du signal GSR normalisée, $\tilde{\delta}_g$ l'écart type du signal GSR normalisé, μ_B la moyenne de BVP, δ_B l'écart type de BVP, $\tilde{\mu}_R$ la moyenne du signal RESPIRATION normalisée, $\tilde{\sigma}_R^2$ l'écart type de la RESPIRATION, p_1 , p_2 et p_3 représentent la densité spectrale de puissance du signal RESPIRATION

caractéristiques	μ_ϵ	$\tilde{\mu}_g$	$\tilde{\delta}_g$	μ_B	δ_B	$\tilde{\mu}_R$	$\tilde{\sigma}_R^2$	p_1	p_2	p_3	p_4
aucune émotion	1.0	.56	-.36	78	.002	.10	.02	.32	.38	.07	.03
colère	3.2	.68	.76	86	.0001	.14	.02	.26	.36	.12	.04
haine	1.5	.60	-.82	80	.0004	.02	.02	.34	.27	.07	.03
peine	1.9	.54	.05	81	.0000	-.12	.08	.33	.18	.07	.03
amour platonique	1.2	.54	-.01	80	-.003	-.01	.02	.43	.23	.04	.02
amour romantique	1.5	.56	.01	83	.0001	.01	.02	.24	.29	.11	.04
joie	1.7	.6	.20	80	-.0001	.00	.06	.35	.24	.06	.02
vénération	0.9	.57	-.44	78	-.0007	-0.2	.01	.33	.36	.06	.03

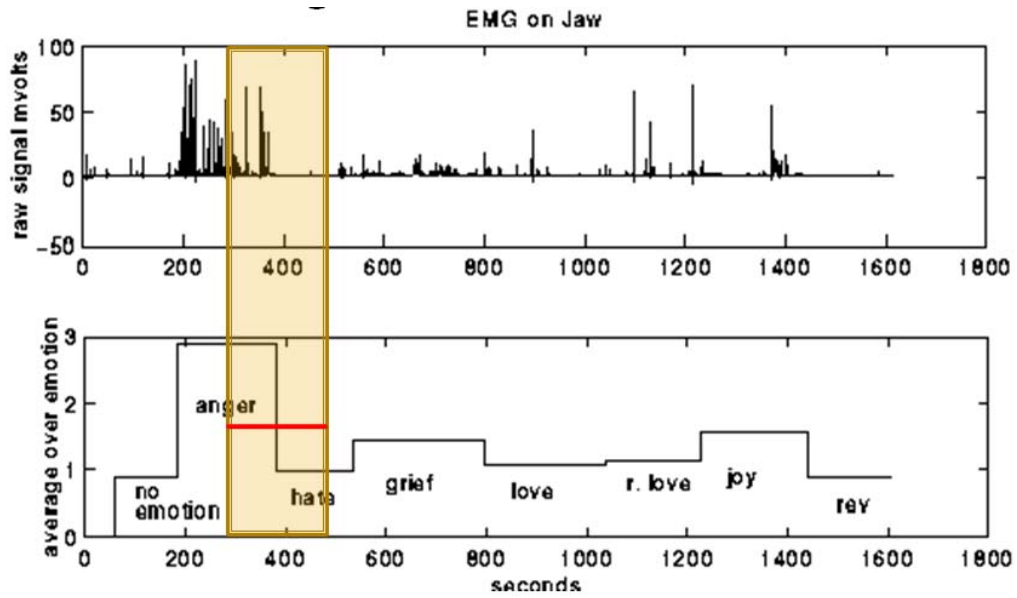


FIGURE 4.7 – Extraction de la caractéristique d'un signal EMG. La moyenne est calculée par scénario (intervalle de temps de chaque émotion)



FIGURE 4.8 – Appratement Gerhome

peuvent être confondus avec des signes de vieillissement. Si la dépression n'est pas traitée à temps, elle peut aggraver la situation. Plus la dépression est diagnostiquée tôt, plus les chances de guérison sont élevées.

4.4.2 Présentation du projet Gerhome

Gerhome (cf. FIGURE 4.8) est un laboratoire expérimental qui reconstitue un appartement type, meublé et instrumenté pour évaluer des solutions facilitant le maintien à domicile des personnes âgées. Il a été mis en place sur le site du CSTB à Sophia Antipolis depuis novembre 2005. L'objectif du projet Gerhome (Gerontology at Home) est de concevoir des solutions techniques et des services d'aide au maintien des personnes âgées dans leur domicile, en utilisant des technologies domotiques pour assurer l'autonomie, le confort de vie et la sécurité. Cet habitat est équipé de dispositifs technologiques permettant d'assurer à distance le suivi de la personne. Des capteurs de présence, des capteurs environnementaux et des capteurs physiologiques sont destinés à recueillir des données sur les activités de la personne dans l'appartement. Les données permettent de décrire précisément l'activité quotidienne de la personne âgée dans son logement et à partir de là, de détecter certains comportements inhabituels, alertes d'une pathologie éventuelle telles qu'une infection urinaire, une dépression ou encore une insuffisance cardiaque.

La FIGURE 4.9 montre le système complet permettant le suivi à distance d'une personne âgée. Ce système s'appuie sur les éléments suivants :

1. Un ensemble de capteurs de différents types (physiologique, environnemental, activité) installés dans l'habitat ou portés par la personne, reliés en réseau pour la collecte en temps réel des données ;
2. Une unité de traitement centrale responsable du stockage et du traitement des signaux reçus des capteurs, de la gestion d'une base de connaissances relative à la personne logée et de l'émission de messages et d'alarmes ;

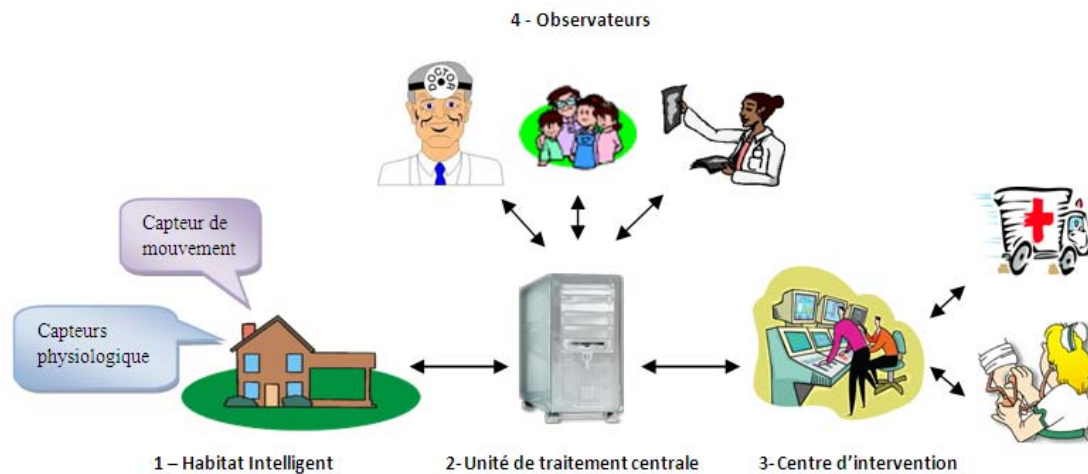


FIGURE 4.9 – Architecture du système de suivi à distance d'une personne à domicile

3. Un centre d'intervention pour le traitement des messages et alarmes reçus des différents habitats
4. Un ensemble d'observateurs (personnel médical, membres de la famille, la personne logée) qui peuvent accéder à tout moment aux données de l'unité de traitement centrale, après authentification et selon leurs privilèges, pour consulter l'évolution de l'état de l'occupant.

Notre travail se situe dans la partie « unité de traitement centrale ». Dans ce contexte, nous proposons une méthode de traitement des données provenant des différents capteurs installés dans l'habitat, afin de prévenir une éventuelle dépression en générant des messages d'alertes.

4.4.3 La Dépression

La dépression est un trouble d'humeur qui se manifeste sur une longue durée par des symptômes de tristesse et de perte de goût de vivre. C'est une véritable maladie qui se caractérise notamment par une grande tristesse, un sentiment de désespoir, une perte de motivation et l'impression de ne pas avoir de valeur en tant qu'individu. Les patients qui connaissent un épisode dépressif majeur accordent une attention particulière aux émotions négatives et les ressentent intensément, tandis qu'ils ne s'attachent pas ou peu aux émotions positives [Quentin Debray 2011]. Tout le monde peut éprouver des sentiments de tristesse et d'angoisse à un moment donné suite à un problème ou un deuil. Mais, si ces émotions négatives s'intensifient et persistent pendant plus de deux semaines et commencent à nuire aux activités quotidiennes, elles signalent peut-être une dépression [Johnstone 2007].

Dans notre travail, nous proposons une méthode de prévention de la dépression en analysant les émotions de la personne âgée. Nous nous intéressons à détecter les



FIGURE 4.10 – Chaussure avec capteur de conductivité de la peau



FIGURE 4.11 – Boucle d'oreille captant la pression sanguine

émotions négatives qui se manifestent de façon permanente pour prédire l'apparition d'une éventuelle dépression. On se place donc, dans le contexte de l'étude de l'évolution à long terme de l'état affectif de la personne âgée logée.

4.4.4 Méthode de prévention de la dépression

Dans cette partie, nous proposons un système de prévention de la dépression. Le système consiste à utiliser les signaux physiologiques afin d'estimer l'état affectif de la personne concernée dans le but d'identifier une éventuelle dépression. Les signaux physiologiques sont recueillis à partir de différents capteurs installés dans l'habitat. Il existe plusieurs formes de capteurs physiologiques non intrusifs que nous pouvons utiliser. Parmi eux nous pouvons citer par exemple les vêtements intelligents comme le gilet Visuresp conçu par l'équipe PRETA [GIARD 004], les chaussures de MIT pour la détection de la conductivité de la peau [Healey 2000], la boucle d'oreille pour mesurer la pression sanguine volumique et le bracelet montre AphyCare, qui permet d'acquérir et de traiter à la fois des paramètres physiologiques (fréquence respiratoire, fréquence cardiaque, température cutanée) et des paramètres d'actimétrie.

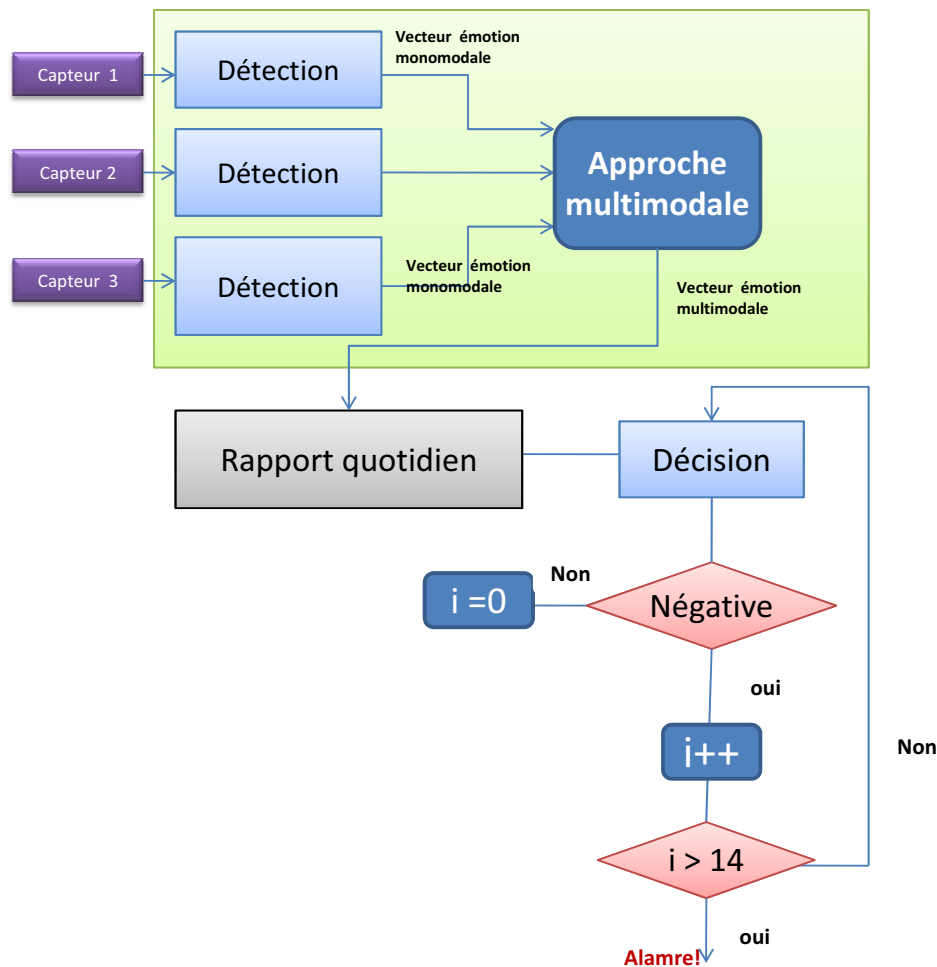


FIGURE 4.12 – Boucle d'oreille captant la pression sanguine

Notre système de détection de la dépression est constitué de trois parties (cf. FIGURE 4.12). Dans la première partie, nous analysons les données émotionnelles collectées des différents capteurs physiologiques. Ensuite nous déterminons l'émotion la plus proche pour chaque modalité (capteur physiologique) en appliquant l'algorithme monomodal de détection des émotions proposé dans le chapitre 3. Nous obtenons ainsi autant de vecteurs émotionnels que de capteurs physiologiques différents. La deuxième étape prend en entrée les vecteurs émotionnels générés à l'étape précédente. Nous appliquons ensuite l'approche multimodale décrite dans la section 4.3 en combinant les informations issues des différentes modalités afin de déterminer l'ensemble des états émotionnels éprouvés par la personne dans l'habitat. Grâce à notre approche multimodale, nous pouvons obtenir par exemple l'émotion « culpabilité » (guilt) en combinant la « peur » et la « joie » et l'émotion « désespoir » en combinant la « peur » et la « tristesse » qui sont deux émotions importantes pour la détection de la dépression.

Dans la troisième partie, nous nous intéressons aux émotions négatives. Nous

détections toutes les émotions éprouvées par la personne âgée le long d'une journée. Nous classifions donc les émotions en deux grands groupes : les émotions positives comme la joie et l'amour ; et les émotions négatives comme la tristesse, l'anxiété et le dégoût. Nous analysons et classifions ainsi ces émotions pour pouvoir donner une évaluation finale de la journée. Finalement la méthode de détection de la dépression consiste à détecter des émotions négatives qui persistent sur une durée de deux semaines. En effet, selon les recherches de [Johnstone 2007] [Men 1998], une personne qui présente un trouble d'humeur (une tristesse intense et une anxiété) qui se manifeste de façon quasi permanente pendant une période supérieure à 2 semaines peut être le sujet d'une dépression majeure. Ainsi, notre algorithme (algorithme 1) calcule le nombre de jours négatifs successifs pour déterminer la présence d'une éventuelle dépression. Si le nombre de jours négatifs successifs est supérieur à 14 (deux semaines), notre système conclut que la personne présente des symptômes de dépression et envoie un message d'alerte aux personnels médicaux concernés.

Algorithme 1 Détection de la dépression

```

1: int negative_day = 0
2: tant que negative_day < 14 faire
3:   tant que True faire
4:     Wait_for_new_day();
5:     Ed = Emotion_Detection(); //gives all the emotions felt throughout the
      day
6:     boolean day_is_negative = positive_or_negative(Ed); //gives true if the
      new day is a negative day
7:     si day_is_negative alors
8:       negative_day ++;
9:     sinon
10:      negative_day = 0;
11: send_alert();
  
```

La FIGURE 4.13 présente deux exemples de rapport d'évaluation émotionnelle, sur une période de vingt cinq jours, générés par notre algorithme de détection de la dépression. Dans le premier scénario, notre algorithme calcule le nombre de jours négatifs successifs et conclut que la personne peut être sujet à une éventuelle dépression et envoie un message d'alerte au personnels médicaux. Le deuxième scénario montre un bilan avec des alternances entre jours positifs et jours négatifs. Dans ce cas notre algorithme conclut que la personne ne présente pas de symptômes d'une dépression.

4.5 Conclusion

Dans ce chapitre, nous avons présenté dans un premier temps, une approche multimodale pour la reconnaissance des émotions qui combine des informations issues de plusieurs sources. Notre approche prend en compte tous les aspects de l'émotion

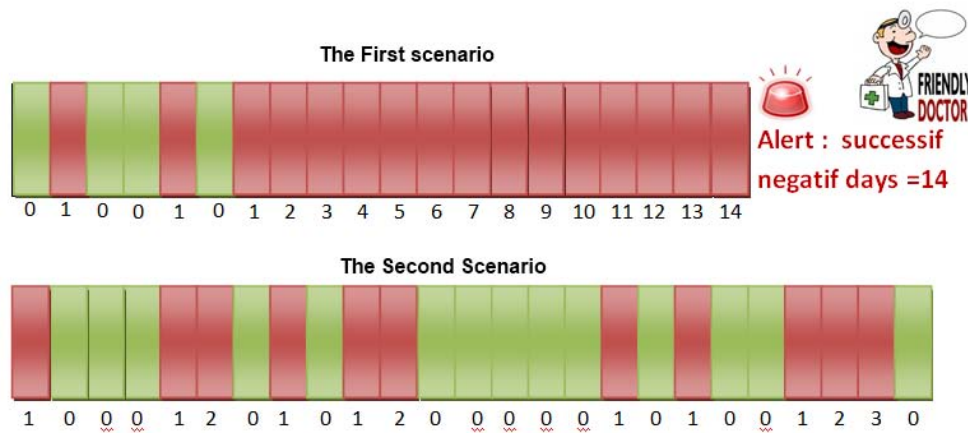


FIGURE 4.13 – Prévention de la dépression

ressentie pour améliorer la reconnaissance des émotions. Elle est basée sur le modèle multidimensionnel défini dans le chapitre 3.7 pour analyser les différents vecteurs émotionnels issus des différentes modalités. Les résultats expérimentaux montrent une amélioration significative des taux de reconnaissance des huit émotions par rapport aux résultats obtenus avec l'approche monomodale.

Dans un second temps, nous avons décrit un cas applicatif de nos travaux en reconnaissance d'émotions. Ce cas applicatif est une méthode de détection de la dépression en utilisant les signaux physiologiques. Cette méthode repose sur toutes les approches et les modèles que nous avons proposés dans ce document. Cependant, par manque de temps, nous n'avons pas testé notre méthode de détection de la dépression sur des cas réels. En effet, Pour évaluer notre méthode, il est nécessaire de générer notre propre base émotionnelle relative à la personne télésurveillée pour pouvoir effectuer le test.

Conclusion générale et perspectives

Conclusion générale

Dans cette thèse nous nous sommes focalisés sur la modélisation, la reconnaissance et l'annotation des émotions. L'objectif de notre travail était d'étudier aussi bien les émotions simples que les émotions les plus complexes comme les émotions mélangées, masquées, simulées, etc pour contribuer à long terme à la conception des futurs systèmes affectifs interactifs.

Nous avons commencé ce manuscrit en présentant les plus importants travaux de l'état de l'art traitant les émotions. Nous avons ensuite défini le phénomène émotionnel et présenté quelques notions en psychologie concernant les émotions. Puis, nous avons décrit quelques travaux en intelligence artificielle qui ont traité les émotions, en mettant l'accent sur les théories et les techniques les plus utilisées dans la reconnaissance automatique des émotions. Enfin nous avons présenté des travaux de représentation et d'annotation des émotions tout en décrivant les différents types de théories utilisées pour les représenter (étiquettes, dimensions, etc.). L'étude bibliographique de ces travaux, nous a permis de montrer que le problème d'échange d'émotions entre applications n'est pas suffisamment traité et qu'il est limité à des langages de modélisation comme EARL et EmotionML qui permettent d'effectuer l'échange émotionnel entre applications en utilisant le dictionnaire de langues naturelles pour la définition des émotions. Cependant, ces langages ne permettent pas d'identifier les qualités sémantiques de données émotionnelles et par la suite ne permettent pas l'échange du concept émotionnel entre applications.

Dans le deuxième chapitre nous avons présenté notre première contribution majeure qui consiste en un modèle à trois couches qui permet de modéliser, annoter et échanger des informations émotionnelles entre applications indépendamment de la modalité utilisée. Il s'agit d'un modèle de représentation hiérarchique composé de trois couches : la couche psychologique, la couche formelle de représentation et de traitement des émotions et la couche langage. La première couche présente le modèle psychologique de Plutchik [Plutchik 1980] que nous avons adopté pour décrire les émotions. Ce modèle (le circumplex de Plutchik) est basé sur huit émotions primaires de base à partir desquelles vont dériver des émotions secondaires. La deuxième couche représente le modèle multidimensionnel que nous avons proposé. Ce modèle se base sur la représentation algébrique des émotions dans un espace vectoriel. Chaque émotion correspond à un vecteur exprimé dans une base à 8 dimensions dont chacune représente une émotion de base définie par Plutchik. Ainsi, nous pouvons décrire dans cette base une infinité d'émotions y compris les émotions masquées et simulées. Nous avons abordé aussi la notion d'intensité de l'émotion ainsi que les techniques de composition et décomposition d'états émotionnels com-

plexes. Nous avons également proposé des opérateurs vectoriels pour analyser les émotions dans l'espace vectoriel proposé, comme l'addition de deux émotions et la distance entre deux vecteurs émotionnels. La troisième couche, qui est la couche langage, permet de coder et annoter les informations émotionnelles afin d'effectuer l'échange de données émotionnelles entre les utilisateurs lors des différentes interactions. Pour cela nous avons proposé d'enrichir le langage de description emotionML proposé par le W3C, avec une nouvelle balise « *plutchik-dimension* » pour pouvoir représenter les émotions dans notre modèle vectoriel.

La deuxième contribution majeure est une approche monomodale de reconnaissance des émotions fondée sur l'analyse des signaux physiologiques. Cette approche se base sur l'application des techniques de traitement du signal et utilise le modèle multidimensionnel que nous avons proposé pour la représentation des émotions. Nous avons calculé, tout d'abord, un ensemble de paramètres statistiques pour extraire les informations caractéristiques de chaque classe émotionnelle. Puis nous avons utilisé le classifieur KNN pour établir une correspondance entre les paramètres calculés et les émotions. Les résultats expérimentaux sont très satisfaisants et permettent d'atteindre 59.44% de classification correcte pour un nombre de classes d'émotions égal à 8. Nous avons montré que notre méthode dépasse en performance les autres méthodes de l'état de l'art comme la méthode Fusion based HHT features dont le taux de classification est égale à 44.33% et la méthode Baseline avec un taux de 45% de bonne classification.

Nous avons ensuite exploité les outils mathématiques de notre modèle vectoriel multidimensionnel afin de définir une approche multimodale. Nous avons utilisé une fusion au niveau de la décision qui consiste à interpréter les caractéristiques de chaque modalité séparément. Les résultats expérimentaux montrent une amélioration significative des taux de reconnaissance des huit émotions par rapport aux résultats obtenus avec l'approche monomodale.

Enfin nous avons proposé une méthode de détection de la dépression des personnes âgées dans un habitat intelligent basée sur notre approche multimodale de reconnaissance des émotions. Nous avons défini cette méthode dans le cadre du projet GERHOME, qui a pour but de concevoir des solutions techniques et des services d'aide au maintien des personnes âgées dans leur domicile. Notre méthode permet d'assurer à distance le suivi à long terme de l'évolution de l'état affectif d'une personne pour identifier l'installation d'une éventuelle dépression et générer un message d'alerte aux personnels médicaux.

Perspectives

Ce travail présente plusieurs perspectives prometteuses. Une première perspective vise l'extension de notre modèle de représentation des émotions pour qu'il tienne compte du contexte d'émergence de l'émotion tels que le contexte culturel (origine, éthique, religion, etc), le contexte personnel (sexe, âge) ou encore le contexte situationnel (lieu, situations et événements déclencheurs de l'émotion). Nous pouvons

ainsi enrichir notre modèle de représentation des émotions en tenant compte du modèle Big-Five de McCrae [McCrae 2002] et en ajoutant des fonctions d'affaiblissement des émotions comme dans les travaux de [Sehaba 2007] pour pouvoir prévoir l'état émotionnel à l'instant suivant.

De façon générale, ce champ de recherche est en pleine évolution. Nos travaux de modélisation ont permis de prendre en compte la nature essentiellement multimodale des émotions, et nous avons apporté des premières pistes pour les représenter. De nombreuses recherches sont cependant encore nécessaires pour doter les systèmes de demain d'émotions complexes, que ce soit en génération ou en détection.

La deuxième perspective de ce travail concerne la partie classification de la méthode de reconnaissance des émotions proposée dans le chapitre 3. Nous envisageons de mener de nouvelles expérimentations avec d'autres méthodes de classification comme la SVM et les réseaux bayésiens Naïfs et comparer les résultats de classification en terme de performance. Ainsi, plusieurs améliorations peuvent encore être apportées à notre approche de reconnaissance des émotions.

Enfin, nous proposons de tester notre méthode de détection de dépression sur des cas réels. Pour cela nous proposons de générer notre propre base émotionnelle relative à la personne télésurveillée pour pouvoir effectuer l'évaluation. Nous proposons également d'étendre notre méthode pour qu'elle tienne compte d'autres indicateurs de la dépression. Nous avons déjà commencé quelques travaux de modélisation de la dépression en utilisant un modèle logistique (voir annexe B) dont la validation se fera plus tard.

Publications

Revues

En cours de publication

Imen Tayari Meftah, Nhan Le Thanh and Chokri Ben Amar, *Sharing emotional states Using an Algebraic model of Emotion*, Elsevier International Journal of Human-Computer Studies 2013.

Imen Tayari Meftah, Nhan Le Thanh and Chokri Ben Amar, *Emotion recognition from physiological signals using unimodal and multimodal approach*, International Journal of Applied Evolutionary Computation (IJAE) 2013.

Conférences internationales

Imen Tayari Meftah, Nhan Le Thanh and Chokri Ben Amar, *Multimodal Approach for Emotion Recognition Using an Algebraic Representation of Emotional States*, The 8th International Conference on Signal Image Technology et Internet Based Systems (SITIS 2012). November 25-29,2012, Sorrento-Naples, Italy.

Imen Tayari Meftah, Nhan Le Thanh and Chokri Ben Amar, *Emotion Recognition using KNN Classification for user modeling and sharing of affect states*, ICONIP 2012, The 19th International Conference on Neural Information Processing,Doha, Qatar November 2012.

Imen Tayari Meftah, Nhan Le Thanh and Chokri Ben Amar, *Detecting depression using a multidimensional model of emotional states*, GLOBAL HEALTH 2012, The First International Conference on Global Health Challenges, Venice, Italy October 2012.

Imen Tayari Meftah, Nhan Le Thanh and Chokri Ben Amar, *Multimodal recognition of emotions using a formal computational model*, ICCS'12 International conference on complex systems, Agadir Morocco November 2012.

Imen Tayari Meftah, Nhan Le Thanh and Chokri Ben Amar, *Detecting Depression Using Multimodal Approach of Emotion Recognition*, ICCS'12 International conference on complex systems, Agadir Morocco November 2012

Imen Tayari Meftah, Nhan Le Thanh and Chokri Ben Amar, *Sharing Emotional Information Using A Three Layer Model (long paper)*, in Proceedings of the The Sixth International Conference on Internet and Web Applications and Services (ICIW 2011), IARIA, Xpert Publishing Services, Maarten, Netherlands Antilles, 20-25 march 2011, pp 130-135 (Acceptation Rate : 28 %)

Imen Tayari Meftah, Nhan Le Thanh and Chokri Ben Amar, *Towards an algebraic modeling of emotional states*, in Proceedings of the The Fifth International Conference on Internet and Web Applications and Services (ICIW 2010), IARIA : Int. Academy, Research and Industry Association, pages 513-518, CPS-IEEE CSDL, Barcelona, Spain, 9-15 may 2010 (Acceptation Rate : 31 %).

Imen Tayari Meftah, Nhan Le Thanh and Chokri Ben Amar, *An algebraic modeling of emotional states*, First International workshop, Rencontres scientifiques UNS-UD (RUNSUD'2010), March 23-25, 2010, Nice Sophia Antipolis, France, pp 411-415

Imen Tayari Meftah, Nhan Le Thanh and Chokri Ben Amar, *A Three-Layer Model for Emotion Exchange (poster)*, The 11th International Multisensory Research Forum (IMRF), June 16 - 19, 2010 Liverpool, United Kingdom

Rapports techniques

Imen Tayari, Nhan Le Thanh et Chokri Ben Amar, *Modélisation des états émotionnels par un espace vectoriel multidimensionnel*, Rapport de recherche 2009.

Imen Tayari, Nhan Le Thanh et Chokri Ben Amar, *Un modèle logistique pour la modélisation de la dépression dans les Habitat Intelligent pour la Santé*, Rapport de recherche 2011.

Glossaire

Humeur : état affectif qui, contrairement à l'émotion, a des origines diffuses (i.e. elle n'est pas déclenchée par un événement en particulier). L'humeur est de plus longue durée et de plus faible intensité [Brave 2003].

Affect : Terme regroupant plusieurs phénomènes affectifs incluant les émotions et l'humeur.

Agent émotionnel : agent virtuel dans lequel est modélisé au minimum un des trois sous-processus de l'émotion (déclenchement de l'émotion, expression des émotions ou influence des émotions sur le processus cognitif).

Émotion déclenchée : Émotion générée suite à l'évaluation subjective d'un événement par un individu. Elle représente la signification émotionnelle d'un événement pour un individu. Une émotion déclenchée modifie les émotions ressenties d'un individu (i.e. son état émotionnel).

Émotion feinte : une émotion feinte est une émotion exprimée par un individu mais non ressentie.

L'intelligence émotionnelle : capacité d'un individu à comprendre les émotions, leurs origines, leurs influences, leurs conséquences et à savoir les contrôler.

Intensité d'une émotion : valeur numérique ou littérale représentant l'impact émotionnel d'un événement pour un individu.

Émotion primaire : (appelée aussi émotion réflexe) émotion générée suite à un traitement court et rapide mais imprécis d'un stimulus. Il n'y a pas de traitement cognitif de l'information. Ce système permet à un individu de répondre rapidement à un stimulus avant même de savoir de quoi il s'agit exactement [Damasio 1994].

Émotion secondaire : émotion déclenchée par un traitement cognitif d'une information [Damasio 1994].

Personnalité : la personnalité (ou tempérament) est ce qui détermine l'individualité de chacun. Elle est liée à l'émotion dans le sens où elle peut déterminer les dispositions d'un individu à ressentir certaines émotions [Brave 2003].

Big Five : Modèle de la personnalité en cinq facteurs, composé de 5 traits de personnalité qui constituent un repère pour la description et l'étude théorique de la personnalité.

- Extraversion : extériorisation des sentiments et des réactions
- Névrosisme : sensation des émotions négatives
- Extraversion : recherche de l'harmonie sociale
- Ouverture à l'expérience : curiosité intellectuelle, ouverture à l'expérience.
- Agréabilité : organisation et persistance

Un modèle logistique pour la modélisation de la dépression

Nous nous sommes basés sur les échelles de dépistage de dépression comme l'échelle HAD (Hospital Anxiety and Depression Scale) et l'échelle Hamilton) pour déterminer les indicateurs de la dépression. On a essayé de prendre presque tous les faits qui expliquent mieux la dépression chez la personne âgée. Ces facteurs qui sont appelés aussi facteurs pronostiques sont : l'agitation, trouble de sommeil, perte de poids, solitude, ralentissement (problème de concentration et de mémoire), émotions négatives.

Donc notre but est de mettre en évidence l'existence d'une liaison fonctionnelle sous-jacente (underlying concept) de la forme $Y = f(X, \alpha)$ entre ces variables. La fonction $f()$ est le modèle de prédiction, on parle aussi de classifieur ; α est le vecteur des paramètres de la fonction, on doit en estimer les valeurs à partir des données disponibles.

Comme dans toute démarche de modélisation, plusieurs questions se posent immédiatement :

- Choisir la forme de la fonction
- Estimer les paramètres du modèle à partir d'un échantillon Ω
- Identifier les descripteurs pertinents dans la prédiction de Y , évacuer celles qui ne sont pas significatives et/ou celles qui sont redondantes.

Donc pour notre démarche de modélisation nous avons choisi la régression logistique pour calculer le risque de survenue ou la probabilité d'apparition de la dépression lorsque les valeurs des variables X_i sont connues. Plusieurs raisons motivent ce choix :

- la régression logistique est relativement simple à comprendre et à appliquer
- elle permet de répondre précisément à chacune des questions posées ci-dessus.
- elle permet de modéliser directement une probabilité
- Les coefficients estimés par le modèle sont en effet liés mathématiquement à l'odds-ratio (ou rapport des cotes) qui représente la force de l'association entre un facteur et la dépression

Nous supposons que la dépression est dichotomique : les valeurs possibles sont 0 (absence de dépression) ou 1 (présence de la dépression). Nous cherchons à prédire les valeurs de la dépression Y , mais nous pouvons également vouloir quantifier la probabilité d'un individu à être positif.

B.1 Modélisation avec la régression Logistique- Principe

Nous cherchons à prédire une variable catégorielle Y (la dépression) à partir d'une collection d'indicateurs $(X_1, X_2, \dots, X_J)$.

- Y caractérise la dépression (Présence/Absence),
- Les X_i caractérisent les i facteurs de risque (ou facteurs pronostiques) de la dépression

Dans ce cas, on estime que Y suit une loi de Bernoulli de paramètre p et nous cherchons à prédire correctement les valeurs de Y et quantifier la probabilité d'un individu à être positif (ou négatif).

$P[Y(\omega) = +/X(\omega)] = \Pi(\omega)$ représente la probabilité a posteriori qu'un individu ω soit positif sachant les valeurs prises par les descripteurs. Ici également, lorsqu'il ne peut y avoir de confusions, nous écrirons Π . Ce dernier terme est très important. En effet, c'est la probabilité que l'on cherche à modéliser en apprentissage supervisé.

Le LOGIT d'un individu ω s'écrit :

$$\ln \left[\frac{\Pi(\omega)}{1 - \Pi(\omega)} \right] = a_0 + a_1 X_1(\omega) + \dots + a_J X_J(\omega) \quad (\text{B.1})$$

$a_0, a_1, \dots, a_J$ sont les paramètres que l'on souhaite estimer à partir des données.

Le quotient

$$\frac{\Pi(\omega)}{1 - \Pi(\omega)} = \frac{P(Y = +/X)}{P(Y = -/X)} \quad (\text{B.2})$$

exprime un odds c'est à dire un rapport de chances pour un individu ω d'obtenir la réponse $Y = 1$ (présence de la dépression). Par exemple, si un individu présente un odds de 2, cela veut dire qu'il a 2 fois plus de chances d'être positif que d'être négatif.

B.1.0.1 Estimation des paramètres par la maximisation de la vraisemblance

Pour estimer les paramètres de la régression logistique par la méthode du maximum de vraisemblance, nous devons tout d'abord déterminer la loi de distribution de $P(Y/X)$.

D'après ce qui est dit précédemment, Y est une variable binaire définie dans $\{+, -\}$, (ou $\{1, 0\}$ pour simplifier les écritures). Pour un individu ω , on modélise la probabilité à l'aide de la loi de Bernoulli de paramètre Π alors :

$$P[Y(w)/X(w)] = \Pi(w)^{y(w)} (1 - \Pi(w))^{(1-y(w))} \quad (\text{B.3})$$

Cette modélisation est cohérente avec ce qui a été dit précédemment, en effet :

- Si $y(w) = 1$ alors $P[Y(w) = 1/X(w)] = \Pi$
- Si $y(w) = 0$ alors $P[Y(w) = 0/X(w)] = 1 - \Pi$

Donc la vraisemblance d'un échantillon Ω s'écrit alors

$$\prod_w \Pi(w)^{y(w)} (1 - \Pi(w))^{(1-y(w))} \quad (\text{B.4})$$

Les paramètres $a = (a_0, a_1, \dots, a_j)$ qui maximisent cette quantité sont les estimateurs du maximum de vraisemblance de la régression logistique. Contrairement à la régression linéaire, une solution analytique exacte n'existe pas. Plusieurs techniques d'optimisation existent. La méthode la plus connue est la méthode Newton-Raphson qui est une méthode itérative du gradient.

Dans ce qui suit, nous notons a le vecteur des paramètres à estimer. La méthode de Newton-Raphson démarre avec une initialisation quelconque du vecteur de paramètre a , pour passer de l'étape (i) à l'étape $(i + 1)$, elle se rapproche de la solution finale en utilisant la formule suivante :

$$a^{i+1} = a^i - \left(\frac{\partial^2 L}{\partial a \partial a'} \right)^{-1} \frac{\partial L}{\partial a} \quad (\text{B.5})$$

$\frac{\partial L}{\partial a}$ est le vecteur des dérivées partielles premières de la vraisemblance.
 $\frac{\partial^2 L}{\partial a \partial a'}$ la matrice des dérivées partielles secondes de la vraisemblance.

Les itérations sont interrompues lorsque la différence entre deux vecteurs de solutions successifs est négligeable. Pour cela, on fixe souvent une valeur seuil epsilon, on arrête le processus si l'écart d'une étape à l'autre est plus petit que le seuil.

Algorithme Newton Raphson l'algorithme newton raphson permet de trouver la valeur $\hat{a}$ de a qui maximise la fonction de vraisemblance L .

Algorithme 2 NewtonRaphsonAlgo(a_0, ε)

```

1: int  $i = 0$ 
2: répéter
3:    $a^{i+1} \leftarrow a^i - \left( \frac{\partial^2 L(i)}{\partial a \partial a'} \right)^{-1} \frac{\partial L(i)}{\partial a}$ 
4:    $i \leftarrow i + 1$ 
5: jusqu'à  $\|a^{i+1} - a^i\| < \varepsilon$ 
6:  $i \leftarrow i + 1$ 
7:  $\hat{a} \leftarrow a^{i+1}$ 
8: return( $\hat{a}$ );
```

B.2 Jeu de données pour la régression logistique

Dans notre travail, On cherche à prédire la probabilité pour qu'un occupant soit sujet d'une dépression. Donc nous allons effectuer n observations avec les 7 variables prédictives pour illustrer la régression logistique binaire. L'objectif est de prédire la présence ou l'absence d'une dépression (dépression « Y » ; avec « *prsence* = + » et « *absence* = - ») à partir de l'agitation (binaire - X_1), trouble de sommeil (binaire - X_2), perte de poids (binaire - X_3), Solitude (binaire - X_4), Ralentissement de mémoire (binaire - X_5), émotions négatives (binaire - X_6). Le but est d'estimer les paramètres de la régression logistique pour notre problème de prédiction de la

Annexe B. Un modèle logistique pour la modélisation de la dépression

dépression. pour cela il faut effectuer les mesures des différents variables relevées sur un échantillon de n personnes.

L'intérêt majeur de cette technique est de quantifier la force de l'association entre chaque variable indépendante et la variable dépendante, en tenant compte de l'effet des autres variables intégrées dans le modèle « mesure ajustée ». Cette méthode est relativement simple à comprendre et à appliquer, ses résultats peuvent être aisément interprétés car liés directement à l'épidémiologie. Les coefficients estimés par le modèle sont en effet liés mathématiquement à l'odds-ratio (ou rapport des cotes) qui représente la force de l'association entre un facteur et une maladie, bien qu'il ne soit qu'une approximation du risque relatif.

Bibliographie

- [Abboud 2004] Bouchra Abboud, Franck Davoine et Mo Dang. *Facial expression recognition and synthesis based on an appearance model*, 2004. (Cité en page 19.)
- [ABDAT 2010] Faiza ABDAT. *Reconnaissance automatique des émotions par données multimodales : expressions faciales et signaux physiologiques*. PhD thesis, Université Paul Verlaine de Metz, juin 2010. (Cité en pages 22 et 23.)
- [Arnold 1960] Magda B. Arnold. *Emotion and personality : Psychological aspects*. 1960. (Cité en pages 9 et 11.)
- [Ayache 2007] Stéphane Ayache, Georges Quénot et Jérôme Gensel. *Classifier Fusion for SVM-Based Multimedia Semantic Indexing*. In Giambattista Amati, Claudio Carpineto et Giovanni Romano, éditeurs, *Advances in Information Retrieval*, volume 4425 of *Lecture Notes in Computer Science*, pages 494–504. Springer Berlin / Heidelberg, Berlin, Heidelberg, 2007. (Cité en page 88.)
- [Baggia 2010] Paolo Baggia, Felix Burkhardt, Alessandro Oltramari, Catherine Pelachaud, Christian Peter et Enrico Zovato. *Emotion Markup Language EmotionML 1.0*. <http://www.w3.org/TR/2010/WD-emotionml-20100729/>, July 2010. (Cité en pages 30 et 31.)
- [Ben Ammar 2007] M. Ben Ammar, Adel M. Alimi, Mahmoud Neji et Guy Gouardères. *Agent-based collaborative affective e-learning system*. In *Proceedings of the First International Conference on Immersive Telecommunications, ImmersCom '07. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering)*, 2007. (Cité en page 16.)
- [Beyer 1999] Kevin Beyer, Jonathan Goldstein, Raghu Ramakrishnan et Uri Shaft. *When Is "Nearest Neighbor" Meaningful?* In *Int. Conf. on Database Theory*, pages 217–235, 1999. (Cité en page 73.)
- [Black 1997] Michael J. Black et Yaser Yacoob. *Recognizing Facial Expressions in Image Sequences Using Local Parameterized Models of Image Motion*. *Int. J. Comput. Vision*, pages 23–48, 1997. (Cité en page 87.)
- [Brave 2003] Scott Brave et Clifford Nass. *The human-computer interaction handbook*. chapitre *Emotion in human-computer interaction*, pages 81–96. L. Erlbaum Associates Inc., Hillsdale, NJ, USA, 2003. (Cité en page 109.)
- [Bruyer 2007] R. Bruyer, P. Granato et J.-P. Van Gansberghe. *Un score individuel de reconnaissance d'une série de stimuli intermédiaires entre deux sources : la perception catégorielle des expressions faciales*. *Revue Européenne de Psychologie Appliquée/European Review of Applied Psychology*, vol. 57, no. 1, pages 37 – 49, 2007. (Cité en page 15.)
- [Cacioppo 1997] John T. Cacioppo, Gary G. Bernston, David J. Klein et Kirsten M Poehlmann. *Annual review of gerontology and geriatrics*, chapitre *psycho-*

- physiology of emotion across the lifespan, pages 27–74. Springer Publishing Company, 1997. (Cité en page 7.)
- [Cascia 1998] Marco La Cascia, Saratendu Sethi et Stan Sclaroff. *Combining Textual and Visual Cues for Content-based Image Retrieval on the World Wide Web*. In In IEEE Workshop on Content-based Access of Image and Video Libraries, pages 24–28, 1998. (Cité en page 88.)
- [Chanel 2007] Guillaume Chanel, Karim Ansari-Asl et Thierry Pun. *Valence-arousal evaluation using physiological signals in an emotion recall paradigm*. In SMC’07, pages 2662–2667, 2007. (Cité en page 26.)
- [Clavel 2007] Chloé Clavel. *Analyse et reconnaissance des manifestations acoustiques des émotions de type peur en situations anormales*. PhD thesis, Ecole Nationale Supérieure des Télécommunications, mars 2007. (Cité en pages 18 et 21.)
- [Clavel 2010] Chloé Clavel et Gaël Richard. Système d’interaction émotionnelle, chapitre Reconnaissance acoustique des émotions. C. Pelachaud, Hermès, 2010. (Cité en pages 21 et 87.)
- [Clay 2009] Alexis Clay. *La branche émotion, un modèle conceptuel pour l’intégration de la reconnaissance multimodale d’émotions dans des applications interactives : application au mouvement et à la danse augmentée*. PhD thesis, Université de Bordeaux I, décembre 2009. (Cité en pages 47 et 89.)
- [Clynes 1982] Manfred Clynes. Neurobiologic functions of rhythm, time, and pulse in music. plenum publishing corporation, manfred clynes édition, 1982. (Cité en page 61.)
- [Collet 1997] Christian Collet, Evelyne Vernet-Maury, Georges Delhomme et André Dittmar. *Autonomic nervous system response patterns specificity to basic emotions*. Journal of the Autonomic Nervous System, pages 45–57, 1997. (Cité en page 22.)
- [Conati 2003] Cristina Conati, Romain Chabbal et Heather Maclaren. *A Study on Using Biometric Sensors for Monitoring User Emotions in Educational Games*, 2003. (Cité en page 26.)
- [CSTB 2009] CSTB. *GERHOME project*. <http://gerhome.cstb.fr/en/home/introduction.html>, September 2009. (Cité en page 94.)
- [Damasio 1994] A. Damasio. *Emotion Reason and the Human Brain*. 1994. (Cité en page 109.)
- [daniel kahneman 1973] daniel kahneman. Attention of effort, chapitre Arousal and attention. 1973. (Cité en page 23.)
- [Darwin 1872] Charles Darwin. The expression of the emotions in man and animals. Oxford University Press Inc, 3 sub édition, 1872. (Cité en pages 6 et 10.)
- [Dave 1997] Morris Dave. La clé des gestes. 1997. (Cité en page 47.)

- [de Bonis 1996] Monique de Bonis. Connaitre les émotions humaines, volume 212. Psychologie et sciences humaines, 1996. (Cité en pages 6, 12 et 41.)
- [Dellaert 1996] Frank Dellaert et al. *Recognizing Emotion in Speech*. In Proceedings of the ICSLP '96, October 1996. (Cité en page 21.)
- [Devillers 2005] Laurence Devillers, Laurence Vidrascu et Lori Lamel. *Challenges in real-life emotion annotation and machine learning based detection*. Neural Netw., vol. 18, no. 4, pages 407–422, Mai 2005. (Cité en page 13.)
- [Devillers 2006] L. Devillers et L. Vidrascu. *Real-life emotions detection on Human-Human spoken dialogs*. In Workshop on corpora for research on emotion and affect, Gênes, 2006. (Cité en page 18.)
- [Ekman 1978] P. Ekman et W. Friesen. Facial Action Coding System : A Technique for the Measurement of Facial Movement. Consulting Psychologists Press, Palo Alto, 1978. (Cité en page 19.)
- [Ekman 1982] Paul Ekman. Emotion in the human face. Cambridge University Press, New York, 1982. (Cité en page 6.)
- [Ekman 1994] Paul Ekman et Richard J. Davidson. The nature of emotion : Fundamental questions. Oxford University Press, New York, 1994. (Cité en pages 6 et 41.)
- [Ekman 1999] P Ekman. Basic emotions. Sussex, U.K. : John Wiley & Sons, Ltd., 1999. (Cité en page 11.)
- [Ekman 2006] P. Ekman. Commentaries : Duchenne and facial expression of emotion, pages 270–284. Studies in Emotion and Social Interaction. Cambridge University Press, 2006. (Cité en page 24.)
- [Elliott 1999] Clark Elliott, Jeff Rickel et James Lester. *Lifelike Pedagogical Agents and Affective Computing : An Exploratory Synthesis*, 1999. (Cité en page 16.)
- [Friesen 2009] Wallace V. Friesen et Paul Ekman. Unmasking the face : A guide to recognizing emotions from facial clues. Malor Books, 2009. (Cité en page 90.)
- [Frijda 1986] Nico H. Frijda. The emotions. Cambridge University Press, 1986. (Cité en pages 6 et 54.)
- [Gao 2002] Yongsheng Gao et Maylor K.H. Leung. *Face Recognition Using Line Edge Map*. IEEE Transactions on Pattern Analysis and Machine Intelligence, pages 764–779, 2002. (Cité en page 20.)
- [Gao 2003] Yongsheng Gao, Maylor K. H. Leung, Siu Cheung Hui et M. W. Tananda. *Facial expression recognition from line-based caricatures*. IEEE Transactions on Systems, Man, and Cybernetics, Part A, pages 407–412, 2003. (Cité en pages ix et 20.)
- [GIARD 004] Jean GIARD. *L'innovation technologique au service du maintien à domicile des personnes âgées*. Rapport de la mission personnes âgées, Commandité par le Conseil Général de l'Isère et la Ville de Grenoble, 2004,. (Cité en page 98.)

- [Gouaillier 2008] David Gouaillier, Vincent Hugel, Pierre Blazevec, Chris Kilner, Jérôme Monceaux, Pascal Lafourcade 0002, Brice Marnier, Julien Serre et Bruno Maisonnier. *The NAO humanoid : a combination of performance and affordability*. CoRR, 2008. (Cité en page 1.)
- [Gratch 2004] Jonathan Gratch et Stacy Marsella. *A domain-independent framework for modeling emotion*. Cognitive Systems Research, vol. 5, no. 4, pages 269–306, Décembre 2004. (Cité en page 9.)
- [Gross 2011] J. J. Gross et L. F. Barret. *Emotion Generation and Emotion Regulation : One or Two Depends on Your Point of View*. 2011. (Cité en page 40.)
- [Hammal 2007] Z. Hammal, L. Couvereur, A. Caplier et M. Rombaut. *Facial expression classification : An approach based on the fusion of facial deformations using the transferable belief model*. Int. J. Approx. Reasoning, vol. 46, no. 3, pages 542–567, Décembre 2007. (Cité en page 19.)
- [Hanson 2005] David Hanson. *Expanding the Aesthetics Possibilities for Humanlike Robots*. In Proceedings of the IEEE "Humanoid Robotics" Conference, special session on the Uncanny Valley, Tskuba, Japan, December 2005. (Cité en page 1.)
- [Healey 2000] Jennifer Healey. *Wearable and Automotive Systems for the Recognition of Affect from Physiology*. PhD thesis, Massachusetts Institute of Technology, 2000. (Cité en pages xiii, 25, 27, 61, 87, 92, 94, 95 et 98.)
- [Istrate 2003] Dan Mircea Istrate. *Détection et reconnaissance des sons pour la surveillance médicale*. PhD thesis, Institut National Polytechnique de Grenoble, 2003. (Cité en page 21.)
- [Izard 1971] Carroll Ellis Izard. *The face of emotion*. Appleton-Century-Crofts, New York, 1971. (Cité en page 10.)
- [Izard 1977] Carroll Ellis Izard. *Human emotions*. Plenum Press, New York, 1977. (Cité en page 6.)
- [Johnson 2004] W. Lewis Johnson, Carole Beal, Anna Fowles-Winkler, Ursula Lauer, Stacy Marsella, Shrikanth Narayanan, Dimitra Papachristou et Hannes Vilhjálmsson. *Tactical Language Training System : An Interim Report*. 2004. (Cité en page 16.)
- [Johnstone 2007] Tom Johnstone, Carina Marije Van Reekum, Heather L. Urry, Ned H. Kalin et Richard J. Davidson. *Failure to regulate : counterproductive recruitment of top-down prefrontal-subcortical circuitry in major depression*. Journal of Neuroscience, vol. 27, no. 33, pages 8877–8884, August 2007. (Cité en pages 97 et 100.)
- [Jones 2008] Christian Jones et Jamie Sutherland. *Affect and emotion in human-computer interaction : From theory to applications*, volume 4868, chapitre Acoustic Emotion Recognition for Affective Computer Gaming, pages 209–219. Springer Verlag, Berlin, 2008. (Cité en page 17.)

- [Karpouzis 2003] K. Karpouzis, A. Raouzaïou et St. Kollias. Moving avatars : Emotion synthesis in virtual worlds. *Human Computer Interaction :Theory and Practice*, 2003. (Cité en page 15.)
- [Kim 2002] Kyung Hwan Kim, Seok Won Bang et Sang Ryong Kim. *Development of person-independent emotion recognition system based on multiple physiological signals*. In *Proceedings of the Second Joint Conference and the Annual Fall Meeting of the Biomedical Engineering Society EMBS/BMES Conference*, volume 1, pages 50– 51, 2002. (Cité en pages 26 et 92.)
- [Kittler 1998] Josef Kittler, Mohamad Hatef, Robert P. W. Duin et Jiri Matas. *On combining classifiers*. *IEEE transactions on pattern analysis and machine intelligence*, vol. 20, pages 226–239, 1998. (Cité en page 89.)
- [Kleinginna 1981] Paul R. Kleinginna et Anne M. Kleinginna. *A categorized list of emotion definitions, with suggestions for a consensual definition*. *Motivation and Emotion*, vol. 5, pages 345–379, 1981. (Cité en page 7.)
- [Lang 1995] P.J. Lang. *The emotion probe : Studies of motivation and attention*. *American psychologist*, pages 372–372, 1995. (Cité en page 24.)
- [Lauter 1985] H. Lauter. *Hand, D. J. : Discrimination and Classification*. *John Wiley and Sons*. *Biometrical Journal*, 1985. (Cité en page 20.)
- [Lazarus 1991] Richard S. Lazarus. *Emotion and adaptation*. 1991. (Cité en page 9.)
- [Lee 2003] Chul Min Lee. *Recognizing emotions from spoken dialogs : A signal processing approach*. 2003. (Cité en page 21.)
- [Levenson 1992] R. W. Levenson. *Autonomic nervous system differences among emotions*. *Psychological Science*, 1992. (Cité en page 23.)
- [Lib 2001] *a library for support vector machines*. Rapport technique 65, Department of computer science and information engineering, Taiwan University, 2001. (Cité en page 30.)
- [Maclean 1993] P. D. Maclean. *Cerebral evolution of emotion*. Guilford Press, New-York, *handbook of emotions édition*, 1993. (Cité en page 6.)
- [Mahdhaoui 2010] Ammar Mahdhaoui. *Analyse de signaux sociaux pour la modélisation de l'interaction face à face*. Thèse de doctorat, Université Pierre et Marie Curie- Paris 6, Décembre 2010. (Cité en page 29.)
- [Marrin 1998] Teresa Marrin et Rosalind Picard. *The Conductor's Jacket : A Device For Recording Expressive Musical Gestures*. In *Proceedings of the International Computer Music Conference*, 1998. (Cité en page 24.)
- [Mathieu-Dupas 2010] Eve Mathieu-Dupas. *Algorithme des K plus proches voisins pondérés (WKNN)*. 42èmes Journées de Statistique, 2010. (Cité en page 27.)
- [McCrae 2002] Robert R. McCrae. *The five-factor model of personality across cultures*. *illustrée édition*, 2002. (Cité en page 105.)
- [Men 1998] *Mental health A REPORT FOCUSING ON DEPRESSION*. *National health priority areas report*, Commonwealth Department of Health and

- Aged Care and Australian Institute of Health and Welfare, 1998. (Cité en page 100.)
- [Mercier 2010] Ludovic Mercier. *Les machines à vecteurs support pour la classification en imagerie hyperspectrale : implémentation et mise en oeuvre*, février 2010. (Cité en pages ix et 29.)
- [Muthusamy 2012] Revathi Priya Muthusamy. *Seminar Paper : Emotion Recognition from Physiological signals using Bio-sensors*. pages 334– 339, 2012. (Cité en page 83.)
- [Nasoz 2003] Fatma Nasoz, Kaye Alvarez, Christine Laetitia Lisetti et Neal Finkelstein. *Emotion recognition from physiological signals for user modeling of affect*. In UM'2003, 9th International Conference on User Model, June 22-26, 2003, Pittsburg, USA, Pittsburg, Å ?TATS-UNIS, 2003. (Cité en page 25.)
- [Neviarouskaya 2007] Alena Neviarouskaya, Helmut Prendinger et Mitsuru Ishizuka. *An Expressive Avatar for Instant Messaging Endowed with Emotional Intelligence*, 2007. (Cité en page 15.)
- [Ochs 2006] Magalie Ochs, Radoslaw Niewiadomski, Catherine Pelachaud et David Sadek. *Expressions intelligentes des émotions*. Revue d’Intelligence Artificielle, vol. 20, no. 4-5, pages 607–620, 2006. (Cité en page 90.)
- [Ochs 2008] Magalie Ochs, Nicolas Sabouret et Vincent Corruble. *Un modèle de la dynamique des relations sociales pour des agents virtuels*. In 16th Journées Francophones sur les Systèmes Multi-Agents, pages 87–96, 2008. (Cité en page 17.)
- [Ortony 1988] Andrew Ortony, Gerald L. Clore et Allan Collins. The cognitive structure of emotions. 1988. (Cité en page 9.)
- [Ortony 1990] A. Ortony et T. J. Turner. *What’s basic about basic emotions ?* Psychol Rev, pages 315–331, 1990. (Cité en pages 13 et 47.)
- [Panksepp 1982] J. Panksepp. *Toward a general psychobiological theory of emotions*. 1982. (Cité en pages 11 et 13.)
- [Pantic 2003] Maja Pantic et Leon J. M. Rothkrantz. *Toward an Affect-Sensitive Multimodal Human-Computer Interaction*. In Proceedings of the IEEE, pages 1370–1390, 2003. (Cité en page 87.)
- [Paugam 2007] Annette Paugam. Agrégation de mathématiques, questions délicates en algèbre et géométrie. Collection sciences sup - 288 pages édition, 2007. (Cité en page 74.)
- [Petropoulou 2006] Hélène Petropoulou. *Profil émotionnel et cognitif au début de la sclérose en plaques : Effets différentiels des émotions sur les performances cognitives*. PhD thesis, Université Paris 8, 2006. (Cité en page 6.)
- [Picard 1997a] Rosalind W. Picard. Affective computing. MIT Press, Cambridge, MA, USA, 1997. (Cité en pages 2, 15 et 16.)

- [Picard 1997b] Rosalind W. Picard. *Affective computing*. MIT Press, 1997. (Cité en pages 20 et 60.)
- [Picard 2000] Rosaline. W. Picard. *Affective Computing*. Technical Report 321, MIT Media Laboratory ;, 20 Ames St., Cambridge, MA 02139, 2000. (Cité en page 16.)
- [Picard 2001] R.W. Picard, E. Vyzas et J. Healey. *Toward Machine Emotional Intelligence : Analysis of Affective Physiological State*. IEEE Transactions on Pattern Analysis and Machine Intelligence, pages 1175–1191, 2001. (Cité en pages 22 et 27.)
- [Plutchik 1980] Robert Plutchik. *Emotion, a psychoevolutionary synthesis*. New York, 1980. (Cité en pages ix, 6, 11, 13, 14, 15, 41, 45, 50 et 103.)
- [Posner 2005] Jonathan Posner, James A. Russell et Bradley S. Peterson. *The circumplex model of affect : An integrative approach to affective neuroscience, cognitive development, and psychopathology*. Development and Psychopathology, vol. 17, no. 03, pages 715–734, 2005. (Cité en pages ix et 11.)
- [Prendinger 2004] Helmut Prendinger, Hiroshi Dohi, Hua Wang, Sonja Mayer et Mitsuru Ishizuka. *Empathic Embodied Interfaces : Addressing Users' Affective State : Embodied Interfaces That Address Users' Physiological State*. In ADS'04, pages 53–64, 2004. (Cité en page 26.)
- [Quentin Debray 2011] Franck Azaïs Quentin Debray Bernard Granger. *Psychopathologie de l'adulte*. 2011. (Cité en page 97.)
- [R. 1996] Sinha R. et Parsons O.A. *Multivariate Response Patterning of Fear and Anger*. Cognition and Emotion, vol. 10, no. 2, pages 173–198, 1996. (Cité en page 22.)
- [Rouillard 2008] José Rouillard. *Adaptation en contexte : contribution aux interfaces multimodales et multicanal*. Habilitation à diriger des recherches, Laboratoire d'Informatique Fondamentale de Lille, 2008. (Cité en pages ix, 35, 36 et 37.)
- [Russell 1980] J.A. Russell. *A circumplex model of affect*. Journal of Personality and Social Psychology, no. 39, pages 1161–1178, 1980. (Cité en page 10.)
- [Russell 1991] James A. Russell. *Culture and the Categorization of Emotions*. Psychological Bulletin, vol. 110, no. 3, pages 426–450, 1991. (Cité en page 13.)
- [Russell 1997] James A. Russell. chapitre How shall an emotion be called ? American Psychological Association, Washington, DC, US, 1997. (Cité en page 10.)
- [Saint-Aime 2007] Sébastien Saint-Aime, Brigitte Le Pévédic et Dominique Duhaut. *Building emotions with 6 degrees of freedom*. In SMC'07, pages 942–947, 2007. (Cité en page 17.)
- [Sander 2010] David Sander. *Cours : Psychologie de l'émotion*, 2010. (Cité en page 46.)
- [Scheirer 2002] Jocelyn Scheirer, Raul Fernandez, Jonathan Klein et Rosalind W Picard. *Frustrating the user on purpose : a step toward building an affective computer*. Interacting with Computers, pages 93 – 118, 2002. (Cité en pages ix, 24 et 25.)

- [Scherer 2000] K. Scherer, Miles Hewstone et Wolfgang Stroebe. *Emotion*, pages 151–191. Oxford Blackwell Publishers, 2000. (Cité en pages 9 et 90.)
- [Scherer 2001a] K. R. Scherer. Appraisal considered as a process of multi-level sequential checking, pages 92–120. Oxford University Press, New York and Oxford, k. r. scherer, a. schorr, & t. johnstone (eds.) édition, 2001. (Cité en page 40.)
- [Scherer 2001b] K. R. Scherer et M. Peper. Psychological theories of emotion and neuropsychological research, pages 17–48. Elsevier, Amsterdam, in f. boller & j. grafman (eds.) édition, 2001. (Cité en page 41.)
- [Schlosberg 1954] Harold Schlosberg. *Three dimensions of emotion*. Psychological Review, vol. 61, no. 2, pages 81–88, 1954. (Cité en page 10.)
- [Sebe 2005] Nicu Sebe, Erwin Bakker, Ira Cohen, Theo Gevers et Thomas Huang. *Bimodal Emotion Recognition*, 2005. (Cité en pages xi, 87 et 90.)
- [Sehaba 2007] Karim Sehaba et Nicoas Sabouretand Vincent Corruble. *An emotional model for synthetic characters with personality*. In International Conference on Affective Computing and Intelligent Interaction (ACII). Lecture Notes in Computer Science (LNCS), 2007. (Cité en page 105.)
- [Smith 1997] Craig A. Smith et Heather S. Scott. The psychology of facial expression, chapitre A Componential Approach to the meaning of facial expressions, pages 229–254. Cambridge University Press, 1997. (Cité en page 19.)
- [Snoek 2003] Cees G.M. Snoek et Marcel Worring. *Multimodal Video Indexing : A Review of the State-of-the-art*. Multimedia Tools and Applications, vol. 25, pages 5–35, 2003. (Cité en page 19.)
- [Stemmler 2001] G. Stemmler, M. Heldmann, C. A. Pauls et T. Scherer. *Constraints for emotion specificity in fear and anger : the context counts*. Psychophysiology, pages 275–291, 2001. (Cité en page 22.)
- [Tato 1999] R. Tato. *Emotion Recognition in Speech Signal*. Thèse de doctorat, University of Manchester, 1999. (Cité en pages xiii et 12.)
- [The HUMAINE Association 2006] The HUMAINE Association. *HUMAINE Emotion Annotation and Representation Language (EARL) : Proposal*. <http://emotion-research.net/projects/humaine/earl/proposal>, juin 2006. (Cité en page 33.)
- [Tian 2000] Ying-Li Tian, Takeo Kanade et Jeffrey Cohn. *Recognizing Lower Face Action Units for Facial Expression Analysis*. In Proceedings of the 4th IEEE International Conference on Automatic Face and Gesture Recognition, pages 484 – 490, March 2000. (Cité en page 87.)
- [Tomkins 1962] Silvan S. Tomkins. Affect imagery consciousness. 1962. (Cité en page 10.)
- [Tomkins 1980] S. Tomkins. *Affect as amplification : some modifications in theory*. Theories of emotions, vol. 1, New York, Academic Press., pages 141–165, 1980. (Cité en page 6.)

- [Veloso 2006] Manuela M. Veloso, Paul E. Rybski, Scott Lenser, Sonia Chernova et Douglas L. Vail. *CMRoboBits : Creating an Intelligent AIBO Robot*. AI Magazine, vol. 27, pages 67–82, 2006. (Cité en page 1.)
- [Villon 2006] Olivier Villon et Christine Laetitia Lisetti. *A user-modeling approach to build user's psycho-physiological maps of emotions using bio-sensors*. In 15th IEEE International Symposium on Robot and Human Interactive Communication, Session Emotional Cues in Human-Robot Interaction, Hatfield, United Kingdom, September 2006. (Cité en page 24.)
- [Vrana 1986] Scott R. Vrana, Bruce N. Cuthbert et Peter J. Lang. *Fear Imagery and Text Processing*. Psychophysiology, 1986. (Cité en page 24.)
- [W3C 2009] W3C. *Extensible MultiModal Annotation markup language*. W3C Recommendation 10, 2009. (Cité en page 34.)
- [Wada 2006] Kazuyoshi Wada et Takanori Shibata. *Robot Therapy in a Care House - its Sociopsychological and Physiological Effects on the Residents*. In Proceedings of the 2006 IEEE International Conference on Robotics and Automation, ICRA 2006, May 15-19, 2006, Orlando, Florida, USA, pages 3966–3971. IEEE, 2006. (Cité en page 17.)
- [Wada 2008] Kazuyoshi Wada, Takanori Shibata, Toshimitsu Musha et Shin Kimura. *Robot therapy for elders affected by dementia*. In Engineering in Medicine and Biology Magazine, IEEE, pages 53– 60. IEEE, 2008. (Cité en page 17.)
- [Wagner 2005] Johannes Wagner, Jonghwa Kim et Elisabeth André. *From Physiological Signals to Emotions : Implementing and Comparing Selected Methods for Feature Extraction and Classification*. In ICME'05, pages 940–943, 2005. (Cité en page 24.)
- [Watson 1992] David Watson et Lee A. Clark. *On Traits and Temperament : General and Specific Factors of Emotional Experience and Their Relation to the Five-Factor Model*. Journal of Personality, vol. 60, pages 441–476, 1992. (Cité en page 7.)
- [Westen 2000] Drew Westen. Psychologie : Pensée, cerveau et culture. 2000. (Cité en page 8.)
- [William 1903] James William. La théorie de l'émotion. Alcan, Paris, 1903. (Cité en pages 6, 7 et 8.)
- [Williams 2004] Andreas Haag; Silke Goronzy; Peter Schaich; Jason Williams. *Emotion Recognition Using Bio-Sensors : First Steps Towards an Automatic System*, 2004. (Cité en pages ix, 23, 24 et 25.)
- [Wundt 1896] Wilhelm Wundt. Grundriss der Psychologie. Leipzig, nd~2004 édition, 1896. (Cité en page 10.)
- [Yazid 2008] Attabi Yazid. *Reconnaissance automatique des émotions à partir du signal acoustique*. PhD thesis, Ecole de technologie supérieure, Université du Québec, février 2008. (Cité en page 21.)

- [Zhang 2005] Yongmian Zhang et Qiang Ji. *Active and Dynamic Information Fusion for Facial Expression Understanding from Image Sequences*. IEEE Trans. Pattern Anal. Mach. Intell., 2005. (Cité en pages ix et 20.)
- [Zong 2009] Cong Zong et Mohamed Chetouani. *Hilbert-Huang transform based physiological signals analysis for emotion recognition*. In Signal Processing and Information Technology (ISSPIT), 2009 IEEE International Symposium on, pages 334– 339, 2009. (Cité en page 83.)

Résumé : Notre travail s'inscrit dans le domaine de l'affective computing et plus précisément la modélisation, détection et annotation des émotions. L'objectif est d'étudier, d'identifier et de modéliser les émotions afin d'assurer l'échanges entre applications multimodales.

Notre contribution s'axe donc sur trois points. En premier lieu, nous présentons une nouvelle vision de la modélisation des états émotionnels basée sur un modèle générique pour la représentation et l'échange des émotions entre applications multimodales. Il s'agit d'un modèle de représentation hiérarchique composé de trois couches distinctes : la couche psychologique, la couche de calcul formel et la couche langage. Ce modèle permet la représentation d'une infinité d'émotions et la modélisation aussi bien des émotions de base comme la colère, la tristesse et la peur que les émotions complexes comme les émotions simulées et masquées.

Le second point de notre contribution est axé sur une approche monomodale de reconnaissance des émotions fondée sur l'analyse des signaux physiologiques. L'algorithme de reconnaissance des émotions s'appuie à la fois sur l'application des techniques de traitement du signal, sur une classification par plus proche voisins et également sur notre modèle multidimensionnel de représentation des émotions.

Notre troisième contribution porte sur une approche multimodale de reconnaissance des émotions. Cette approche de traitement des données conduit à une génération d'information de meilleure qualité et plus fiable que celle obtenue à partir d'une seule modalité. Les résultats expérimentaux montrent une amélioration significative des taux de reconnaissance des huit émotions par rapport aux résultats obtenus avec l'approche monomodale.

Enfin nous avons intégré notre travail dans une application de détection de la dépression des personnes âgées dans un habitat intelligent. Nous avons utiliser les signaux physiologiques recueillis à partir de différents capteurs installés dans l'habitat pour estimer l'état affectif de la personne concernée.

Mots clés : modélisation des émotions, annotation des émotions, modèle multidimensionnel, reconnaissance des émotions, approche monomodale, approche multimodale.

Modeling , detection and annotation of emotional states using an algebraic multidimensional vector space

Abstract : This study focuses on affective computing in both fields of modeling and detecting emotions. Our contributions concern three points. First, we present a generic solution of emotional data exchange between heterogeneous multi-modal applications. This proposal is based on a new algebraic representation of emotions and is composed of three distinct layers : the psychological layer, the formal computational layer and the language layer. The first layer represents the psychological theory adopted in our approach which is the Plutchik's theory. The second layer is based on a formal multidimensional model. It matches the psychological approach of the previous layer. The final layer uses XML to generate the final emotional data to be transferred through the network. In this study we demonstrate the effectiveness of our model to represent an infinity of emotions and to model not only the basic emotions (e.g., anger, sadness, fear) but also complex emotions like simulated and masked emotions. Moreover, our proposal provides powerful mathematical tools for the analysis and the processing of these emotions and it enables the exchange of the emotional states regardless of the modalities and sensors used in the detection step.

The second contribution consists on a new monomodal method of recognizing emotional states from physiological signals. The proposed method uses signal processing techniques to analyze physiological signals. It consists of two main steps : the training step and the detection step. In the First step, our algorithm extracts the features of emotion from the data to generate an emotion training data base. Then in the second step, we apply the k-nearest-neighbor classifier to assign the predefined classes to instances in the test set. The final result is defined as an eight components vector representing the felt emotion in multidimensional space.

The third contribution is focused on multimodal approach for the emotion recognition that integrates information coming from different cues and modalities. It is based on our proposed formal multidimensional model. Experimental results show how the proposed approach increases the recognition rates in comparison with the unimodal approach.

Finally, we integrated our study on an automatic tool for prevention and early detection of depression using physiological sensors. It consists of two main steps : the capture of physiological features and analysis of emotional information. The first step permits to detect emotions felt throughout the day. The second step consists on analyzing these emotional information to prevent depression

Keywords : modeling emotional states, emotion annotation, multidimensional model, emotion recognition, monomodal approach, multimodal approach.
