UNIVERSITÉ PAUL VERLAINE DE METZ

IAEM Lorraine

THÈSE

pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE METZ

Discipline : Automatique, Traitement du Signal et des Images, Génie Informatique

par

Faiza ABDAT

Reconnaissance automatique des émotions par données multimodales : expressions faciales et signaux physiologiques

Soutenue le 15 - 06 - 2010 devant le Jury :

Mme. Isabelle MAGNIN (directeur de recherche) ..... Présidente
M. François CABESTAING (professeur des universités) ..... Rapporteur
M. Norbert NOURY (professeur des universités) ..... Rapporteur
M. Alain PRUSKI (professeur des universités) ..... Directeur de thèse
Mme. Choubeila MAAOUI (maître de conférences) ..... Co-Directrice de thèse
M. Guy BOURHIS (professeur des universités) ..... Examinateur

Laboratoire d'Automatique humaine et de Sciences Comportementales

Remerciements

Je tiens tout d'abord à témoigner toute ma reconnaissance à Monsieur Alain PRUSKI, Professeur de l'Université de Metz, pour son appui scientifique, sa disponibilité et toutes les suggestions qu'il m'a apporté durant ma thèse.

J'exprime mon profond remerciement à Madame Choubeila Maaoui, Maître de Conférences à l'Université de Paul Verlaine-Metz, pour son suivi permanent, sa disponibilité et ses conseilles qu'elle m'a apporté lors de l'encadrement de ma thèse.

Je remercie l'ensemble des membres du jury qui m'ont fait l'honneur d'accepter de lire et de juger ce mémoire : Messieurs Norbert NOURY (Professeur 63e, Université de Lyon), François CABESTAING (Professeur 61e, UST Lille), Isabelle MAGNIN (Directeur de recherche, CNRS Lyon) et Guy BOUHRIS (Professeur 61e, LASC Metz).

Mes remerciements s'adressent à tous les membres du laboratoire LASC, pour leur soutien, leur sympathie et leur aide. Je remercie aussi toute personne a accepté d'effectuer des expériences lors de ce travail de thèse.

J'adresse particulièrement mes remerciements à Monsieur Olivier HABERT, directeur de l'Institut Supérieur d'Électronique et d'Automatique qui m'a accueilli pendant deux années en tant qu'ATER ainsi que tout le personnel du bâtiment ISEA.

Je remercie également mes parents de m'avoir toujours soutenus pendant toutes mes années d'études et pour m'avoir toujours encouragé à aller le plus loin possible. Mes sincères remerciements vont aussi à mes frères et à ma sœur.

Je ne remercierais jamais assez mon mari Ahcene pour sa patience, son soutien, ses encouragements quotidiens, son écoute et son enthousiasme... Je lui dois une très profonde gratitude.

Table des matières

Table des figures

Liste des tableaux

Introduction générale

Les émotions colorent notre vie, permettent d'exprimer les différentes facettes de la personnalité, et, les vivre pleinement, c'est s'autoriser une existence intense. On a beaucoup cru, au siècle précédent, à la toute puissance de la raison en oubliant l'émotion. Et pour cause ! On la considérait comme un obstacle au travail de la raison. Grâce au neuroscientifique et à l'imagerie cérébrale, on sait désormais que l'être humain n'est pas un décideur rationnel et que l'émotion est un partenaire fondamental de la cognition humaine, de sa créativité et de sa prise de décision [52].

Doter la machine des capacités de reconnaissance d'état émotionnel, tel est le défi scientifique autour duquel se rassemblent différentes communautés (traitement du signal, traitement d'images, intelligence artificielle, robotique, interaction homme-machine, etc.).

L'état émotionnel des humains peut être obtenu à partir d'un large éventail d'indices comportementaux et des signaux qui sont disponibles par le biais d'une expression ou d'une présentation visuelle, auditive et physiologique de l'émotion :

- L'état émotionnel à travers l'expression visuelle est évalué en fonction de la modulation des expressions faciales, gestes, postures et plus généralement le langage corporel. Les données sont capturées par une caméra, permettant des configurations non intrusives. Les systèmes sont généralement très sensibles à la qualité de la vidéo, l'éclairage, la pose et la taille du visage sur la vidéo [156] ;

- L'état émotionnel à travers l'expression auditive peut être estimé comme une modulation du signal vocal [156]. Dans ce cas, les données sont captées par un microphone, ce qui permet des configurations d'un système non intrusif. Les besoins en traitement de données vocales propres (rapport signal / bruit (SNR) inférieur à 10 dB) peut sérieusement réduire la qualité de l'estimation [233]. En outre, le traitement est difficilement géré lors de la présence de plus d'une seule voix dans le flux audio ;

- L'état émotionnel à travers la représentation physiologique est estimé par la modulation de l'activité du système nerveux autonome (SNA). L'estimation peut être très fiable [137, 213]

et est moins sensible à la qualité des émotions que celles extraites des modalités auditives et visuelles [156]. La principale limitation est liée à l'intrusion des dispositifs de détection.

Comme le contenu émotionnel reflète le comportement humain, la reconnaissance automatique des émotions est un sujet qui suscite un intérêt croissant. Ce n'est cependant pas une tâche aisée. Les émotions jouent en effet un rôle implicite dans le processus de communication en comparaison du message explicite véhiculé par le niveau lexical. Le phénomène à reconnaître est complexe et subtil, présentant des manifestations très diversifiées et dépendantes de nombreux facteurs (contexte social, culturel, personnalité du locuteur, etc.).

La mesure des émotions est extrêmement délicate. Il est nécessaire de combiner des techniques classiques de mesure [135] utilisées souvent séparément : reconnaissance des expressions faciales, reconnaissance de la parole et analyse des signaux physiologiques. Dans ce travail, nous nous concentrons sur la combinaison des mesures physiologiques et des expressions faciales pour la reconnaissance des émotions. Plusieurs avantages peuvent être attendus lors de la combinaison des signaux physiologiques et des expressions faciales.

Tout d'abord, une expression faciale est une manifestation visible de l'état émotionnel, de l'activité cognitive, de l'intention, de la personnalité et de la psychopathologie d'une personne. Dans [149], Mehrabian a mis en évidence le fait que 55 % du message émotionnel est communiqué par l'expression faciale alors que 7 % seulement par le canal linguistique et 38% par le paralangage. Ainsi, les expressions faciales jouent un rôle important dans la communication humaine et en interaction homme-machine. Mais, l'utilisateur peut consciemment ou inconsciemment, cacher ses émotions détectées par des canaux extérieurs (visage et voix).

D'un autre coté, les capteurs physiologiques nous permettent de recueillir en permanence des informations sur l'état émotionnel de l'utilisateur alors que l'analyse des émotions du visage doit être détectée lorsque les expressions montrent un changement et que la personne est en face de la caméra. De plus, il est difficile pour les utilisateurs de manipuler librement les capteurs physiologiques par rapport aux expressions faciales ou à la voix.

Enfin, une analyse basée sur les signaux physiologiques et sur les expressions faciales permet de lever les ambiguïtés et de compenser les erreurs.

Au niveau de notre laboratoire, on travaille sur une application de la réalité virtuelle dans le cadre d'une thérapie de la phobie sociale. Cette application nécessite une étape de contrôle basée sur l'état émotionnel du patient. Ce travail de thèse est une étape générique sur l'identification des émotions. Pour cela, nous proposons un système bimodal pour la reconnaissance des émotions à partir des expressions faciales et des signaux physiologiques. Nous analysons l'expression externe (visage) et les facteurs internes (signaux physiologiques) de la réponse humaine afin de déterminer l'émotion correspondante.

Le premier chapitre présente en premier lieu des notions de bases sur les émotions puis décrit plus particulièrement les différentes techniques d'analyse d'expressions faciales qui visent

à reconnaître les émotions. Nous abordons ensuite les méthodes proposées pour la reconnaissance des émotions avec les signaux physiologiques. Nous clôturons le chapitre par une revue des quelques modèles décrits dans la littérature concernant la fusion des données provenant de plusieurs modalités.

Le second chapitre est tout d'abord consacré à l'extraction des caractéristiques faciales, où nous présentons la détection de visage avec les descripteurs de HAAR. Nous proposons ensuite notre modèle anthropométrique pour la détection des points faciaux déduit d'expérimentations sur des bases de données où chaque muscle est représenté par la distance entre deux points. La variation des distances par rapport à l'état neutre est utilisée pour le codage des expressions faciales qui sont classées par la suite avec les séparateurs à vastes marges SVM en une émotion parmi celles décrites par Ekman. Les résultats de cette approche sont validés sur deux bases de données.

Le troisième chapitre présente une approche de reconnaissance des émotions fondée sur l'analyse des signaux physiologiques. Nous décrivons tout d'abord les différents signaux physiologiques utilisés pour la prédiction des émotions, ainsi que leurs relations avec les processus émotionnels. Un ensemble de paramètres statistiques est calculé pour extraire les informations caractéristiques qui permettent la classification des différents états émotionnels.

Le dernier chapitre porte sur les résultats expérimentaux obtenus avec un système bimodal pour la reconnaissance des émotions. Nous présentons tout d'abord les différents niveaux de fusion de données possibles dans notre étude. Nous décrivons une méthode de transformation des données basée sur l'analyse en composantes principales et une autre méthode de sélection des caractéristiques basée sur l'information mutuelle. Ces deux méthodes font parties de la fusion des caractéristiques. Pour la fusion au niveau de la décision, nous présentons une méthode simple basée sur le processus de vote et une autre méthode basée sur les réseaux Bayésiens dynamiques. Ensuite, nous discutons les résultats des tests réalisés sur plusieurs panels de personnes valides au laboratoire.

Chapitre 1

État de l'art sur la reconnaissance des émotions

1.1 Introduction

Dans ce chapitre, nous allons présenter quelques notions concernant les émotions telles que leurs définitions, leurs différentes théories, et leurs composantes. En se basant sur ces dernières, nous présentons un état de l'art sur les expressions faciales, sur l'analyse des signaux physiologiques, et nous terminons le chapitre par les systèmes multimodaux.

1.2 Notions sur les émotions

1.2.1 Définition

Les émotions, de façon générale, sont des états motivationnels. Elles sont constituées d'impulsions, de désirs ou d'aversions ou, plus généralement, elles comportent des changements de motivation. Elles poussent l'individu à modifier sa relation avec un objet, un état du monde, un état de soi, ou à maintenir une relation existante malgré des obstacles ou des interférences [51].

Notons une caractéristique essentielle de ces *motivations* : les émotions sont relationnelles. Elles se jouent entre le sujet et le monde. Les émotions ne sont pas des états subjectifs, intérieurs à une personne, ou du moins pas en première instance [51].

Évidemment, une émotion peut rester intérieure à une personne et rester limitée à son expérience intime. Mais, même dans ce cas, la tendance à l'action est présente, se manifeste dans le ressenti et à travers l'imagination. En colère, on pense à ce qu'on voudrait faire à l'adversaire ou, de façon plus discrète encore, à ce qu'on aimerait qu'il lui arrive. Dans l'inquiétude, les pensées vont de-ci de là sans repos, raidissant le dos pour faire face à ce qui pourrait arriver [51].

L'expérience subjective, le ressenti des émotions, est largement le reflet des tendances à l'action, comme le montrent les recherches portant sur la description des expériences émotionnelles. Les émotions dites « de base » sont caractérisées par des mondes de préparation distincts et spécifiques : la peur par la tendance à s'éloigner ou à se protéger, la colère par l'opposition et l'hostilité, la honte et la culpabilité par la soumission, et les émotions de joie et de tristesse par

des tendances plus diffuses d'augmentation et de diminution de l'activation générale [87].

1.2.2 Modèles théoriques de l'émotion

Selon Scherer, « les émotions sont les interfaces de l'organisme avec le monde extérieur » et le processus émotionnel se décompose en trois principaux aspects [187] :

1. L'évaluation de la signification des stimuli par l'organisme (aspect cognitif) ;
2. La préparation aux niveaux physiologique et psychologique d'actions adaptées (aspect physiologique) ;
3. La communication par l'organisme des états et des intentions de l'individu à son environnement social (aspect expressif).

Ces trois aspects, cognitif, physiologique et expressif sont généralement acceptés comme constituants du phénomène émotionnel [48].

1.2.2.1 Théorie physiologique

Une première théorie proposée, à peu près en même temps, par le psychologue américain William James (1884) et le physiologiste danois Carl Lange (1885) mettent l'accent sur le rôle essentiel des réactions émotionnelles dans le déclenchement de l'expérience émotionnelle. Cette théorie trouvera plus tard un prolongement avec l'hypothèse de la rétroaction faciale qui met en lumière la façon dont la prise de conscience des modifications corporelles et faciales intervient dans les modifications de l'état d'esprit de la personne.

D'autres chercheurs, comme Walter Cannon (1927), dont la théorie sera développée par Phillip Bard (1934), jugent plutôt que les réactions émotionnelles résultent au même titre que le vécu psychologique, de l'activation de mécanismes sous-corticaux [93].

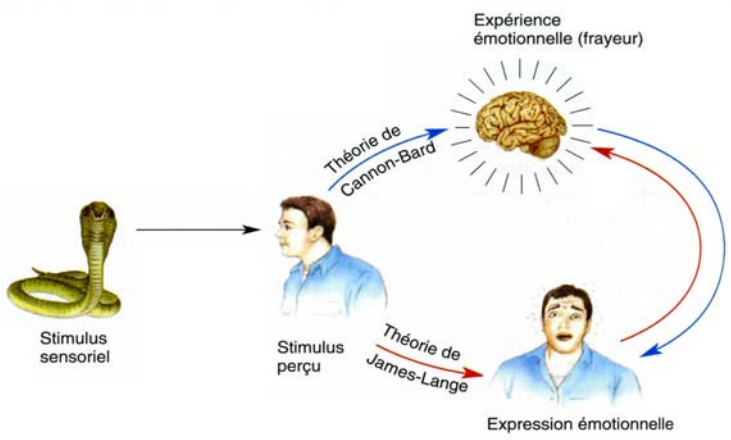


FIGURE 1.1 – Les théories physiologiques [208]

La figure 1.1 montre une comparaison schématique des théories de James-Lang et Cannon-Bard, des processus émotionnels. Selon la théorie de James-Lang (sens inverse des aiguilles d'une montre), l'individu perçoit la présence de l'animal effrayant, puis réagit. C'est ce comportement, déclenché en réponse à la perception de l'animal qui lui fait ressentir la frayeur. Selon la théorie de Cannon-Bard (sens des aiguilles d'une montre), la frayeur résulte de la perception du stimulus, et ensuite seulement il y a une réaction comportementale.

1.2.2.2 Théorie Néo-Darwinienne

La perspective évolutionniste tire son origine des travaux de Darwin [56]. Elle étudie essentiellement la fonction communicative des émotions en donnant la prédominance aux expressions faciales.

Charles Darwin, en 1872, fut l'un des premiers à s'intéresser aux phénomènes émotionnels en publiant, dans le prolongement de son analyse évolutionniste de l'univers vivant, un ouvrage intitulé : l'expression des émotions chez l'homme et l'animal.

Pour Darwin, les expressions émotionnelles de l'adulte humain sont le reflet de la continuité de systèmes comportementaux complexes dérivés des autres espèces animales [44]. Darwin a eu recours à trois principes de base afin d'expliciter sa démarche :

1. Les habitudes associées : les expressions émotionnelles sont à l'origine des actes utilitaires qui rempliraient une fonction adaptative par rapport à l'environnement ;
2. L'antithèse : les états émotionnels sont souvent caractérisés par des manifestations motrices antagonistes ;
3. L'action directe sur le cerveau : effet de débordement et de dérivation de la force nerveuse engendrée par la stimulation.

Les théories néo-darwiniennes se sont essentiellement focalisées sur la détermination des émotions de base en étudiant les expressions faciales émotionnelles. Les diverses catégorisations des émotions de base proposées dans la littérature indiquent qu'il existe d'importantes divergences entre les auteurs. Ces diverses conceptions théoriques ont en commun de mettre l'accent sur la relation entre une configuration expressive faciale et une émotion spécifique. Les expressions faciales permettent aussi de communiquer à autrui son état émotionnel interne [44].

1.2.3 Neurophysiologie des émotions : Le système limbique

De nombreuses structures du cerveau participent à la physiologie des émotions, nous nous concentrons sur le système limbique et des structures voisines. La figure 1.2 montre une conception du système limbique comprenant un réseau de structures interconnectées qui contrôlent l'expression émotionnelle. Les principales structures comprennent le cortex cingulaire, l'hippocampe, l'amygdale et ses connexions étendues avec l'hypothalamus et le cortex, les corps mammillaires de l'hypothalamus et le cortex préfrontal.

L'hypothalamus est le lieu où sont générés les comportements et le système limbique crée les émotions. Les régions préfrontales et sensorielles établissent des contacts avec le cortex cingulaire, l'hippocampe et l'amygdale. Les deux dernières structures établissent des connexions avec l'hypothalamus, qui à son tour, par le thalamus, établit des connexions avec le cortex cingulaire.

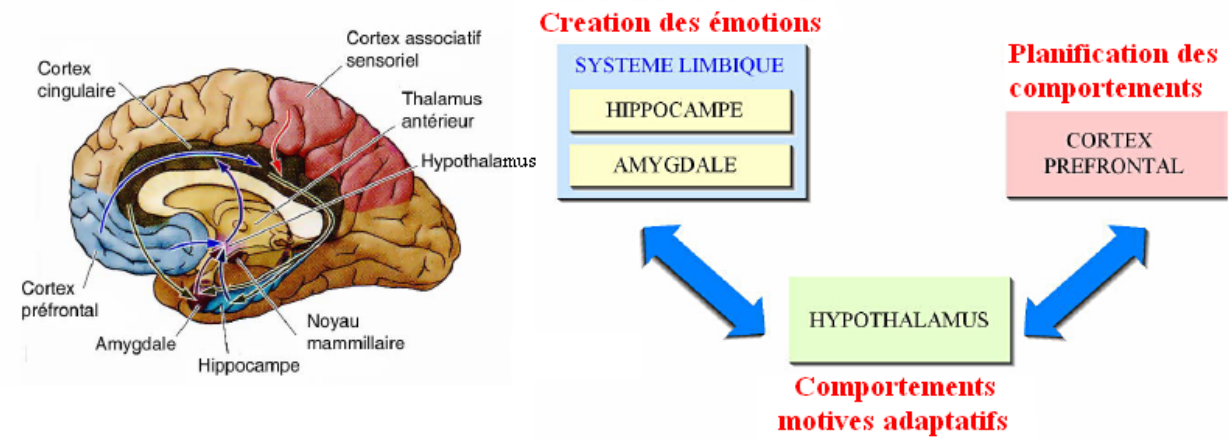


FIGURE 1.2 – Représentation schématique des connexions principales du système limbique [208]

1.2.4 Types d'émotion

Les émotions sont des séquences courantes et vives qui interviennent tout au long de nos journées et en donnent le ton. Il existe trois types d'émotions : émotions primaires ou dites de base, émotions secondaires et émotions sociales.

1.2.4.1 Émotions primaires

Les émotions primaires sont déclenchées par des événements particuliers ou bien elles se manifestent dans des circonstances précises en provoquant des comportements spécifiques (tableau 1.1). Elles sont à la base de nos réactions qui ne sont pas seulement déterminées par notre jugement rationnel ou notre passé individuel mais aussi par notre passé ancestral [55].

En fait, ces émotions primaires sont comme une matière première, à partir de laquelle on peut fabriquer toutes les autres émotions [52].

1.2.4.2 Émotions secondaires

Les émotions secondaires ont pour base, au départ, un processus de pensée et sont l'aboutissement de l'apprentissage des émotions primaires.

Les émotions secondaires sont celles qui sont engendrées à l'évocation de souvenirs et arrivent à maturation à l'âge adulte [55].

1.2.4.3 Émotions sociales

On parle aussi des émotions sociales qui sont inhérentes à la relation aux autres, comme la

culpabilité, la honte, la jalousie, la timidité, l'humiliation, etc. Toutes ces émotions sont apprises et sont constituées à partir des émotions primaires. L'éducation et la culture sont fortement impliquées dans l'acquisition des émotions sociales [52].

Émotion	Déclencheurs et circonstances d'apparition	Comportement
Joie	Désir Réussite Bien-être Accomplissement	Approche
Tristesse	Perte Deuil	Repli sur soi
Colère	Obstacle Injustice Dommage Atteindre à son intégrité physique ou psychique Limites de la personne Atteinte au système de valeurs	Attaque
Peur	Menace Danger Inconnu	Fuite Sidération Évitement Parfois attaque
Dégoût	Substance ou personne nuisible Aversion physique ou psychique Contre quelqu'un Rejet	
Surprise	Danger immédiat Inconnu Imprévu	Retrait Sursaut

TABLE 1.1 – Les 6 émotions de base [52]

1.2.5 Représentation des émotions

La manipulation des émotions par ordinateur soulève de nombreuses problématiques. D'abord au niveau de leurs représentations, il s'agit de trouver un formalisme qui soit en accord avec les résultats psychologiques existants, tout en permettant une manipulation simple. Ensuite, pour un événement donné, il faut pouvoir déterminer le potentiel émotionnel qui lui est associé. En se fondant sur les travaux en psychologie, certaines mesures considèrent les états émotionnels comme des catégories, d'autres comme un construit multidimensionnel.

1.2.5.1 Approche catégorielle

C'est l'approche la plus répandue, qui consiste à considérer les émotions comme des caractéristiques épisodiques et universelles [70]. Il suffit ensuite d'associer un mot du langage à ces caractéristiques. 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 ethnie ou leur culture (Table 1.2). Cette approche fait essentiellement la distinction entre ces émotions et propose de les classer sous forme de catégories discrètes. Ainsi les dénominations affectives qui ne trouvent pas leur place dans ces classifications sont considérées comme des mélanges d'émotions primaires. La justification principale de cette approche réside dans le fait que ces émotions basiques sont clairement identifiables chez la majorité des individus, notamment à travers la communication non verbale. Toutefois, leur nombre, le nom qu'il faut leur attribuer et leur caractérisation comme émotion basique, restent des questions ouvertes [154].

L'intérêt principal de l'approche catégorielle est qu'une fois que les émotions à traiter sont clairement identifiées, il devient simple de les manipuler, aussi bien pour les hommes que pour les machines.

Auteur	Émotions basiques
Ekman et al. [77]	colère, dégoût, peur, joie, tristesse, surprise.
Izard [105]	colère, mépris, dégoût, détresse, peur, culpabilité, intérêt, joie, honte, surprise.
Plutchik [170]	acceptation, colère, anticipation, dégoût, peur, joie, tristesse, surprise.
Tomkins [205]	colère, intérêt, mépris, dégoût, détresse, peur, joie, honte, surprise.

TABLE 1.2 – Liste des émotions basiques selon différents auteurs

1.2.5.2 Approche dimensionnelle

Approche dimensionnelle est une autre approche théorique très populaire en psychologie des émotions humaines [182, 60] qui propose une représentation continue sur plusieurs axes ou dimensions (par contraste aux catégories discrètes des émotions de base) (Table 1.3).

Trois facteurs ont été utilisés afin de mieux rendre compte des effets psychophysiologiques des différentes émotions : la valence, le degré d'activation physiologique (ou l'arousal) et la dominance (contrôle). En général, deux dimensions principales sont mises en avant (figure 1.3). D'une part, la valence émotionnelle, c'est-à-dire le caractère positif ou négatif de l'expérience émotionnelle et, d'autre part, la dimension de l'intensité ou le degré d'activation de l'expérience émotionnelle (l'arousal).

L'approche dimensionnelle permet de représenter facilement des émotions nuancées mais également des transitions entre différents états émotionnels.

Auteur	Axe choisi
Russel [183]	Arousal/Valence
Cowie et al.[54]	Activation/ Évaluation

TABLE 1.3 – Quelques axes choisis par différents auteurs

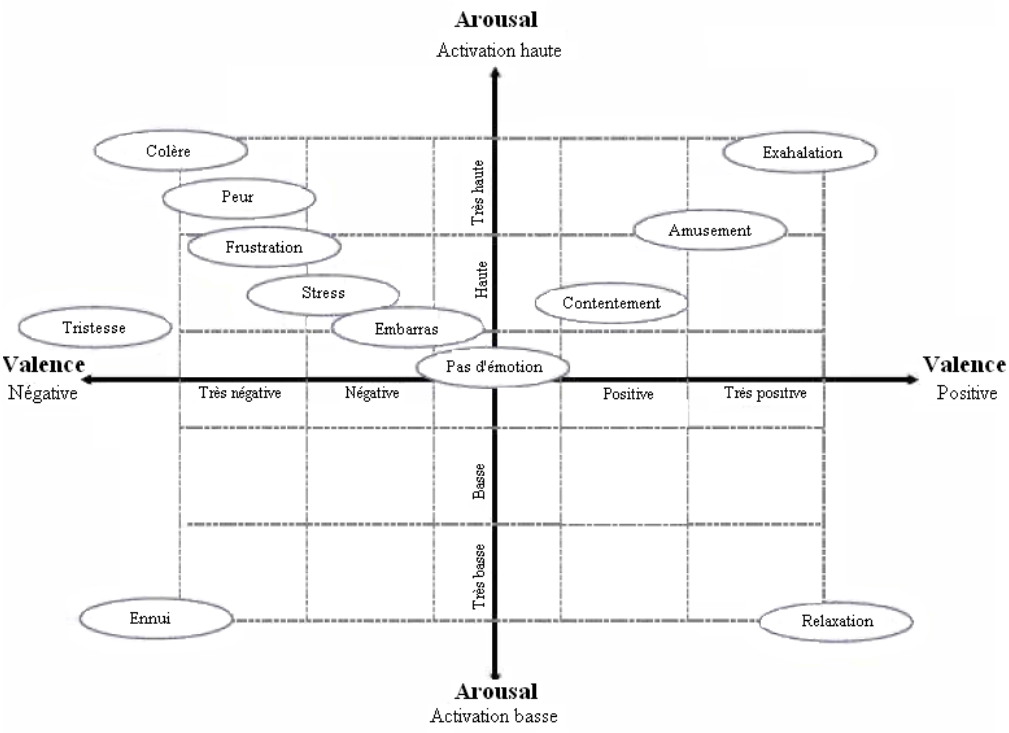


FIGURE 1.3 – La représentation de quelques émotions sur deux axes [171]

Plutchik en 1980 place les émotions primaires sur les différents secteurs d'un cercle. Dans les encadrés de forme rectangulaire, on trouve les dyades primaires qui correspondent à des émotions secondaires. Elles résultent de la combinaison de deux émotions primaires représentées par des secteurs adjacents sur le cercle [170]. Par exemple, la déception résulte de la tristesse et de la surprise (voir figure 1.4).

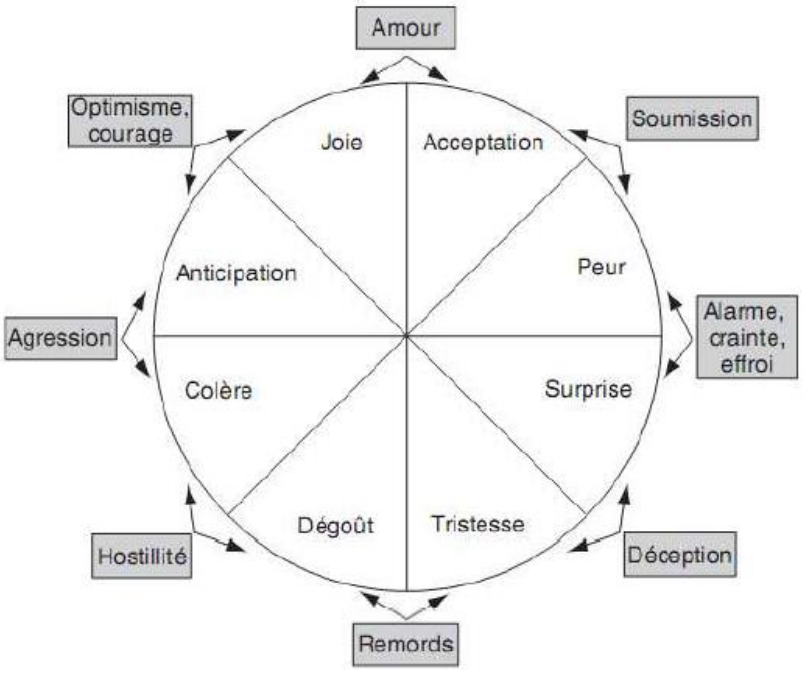


FIGURE 1.4 – La représentation des émotions mixtes [170]

La figure 1.5 montre le modèle multidimensionnel de Plutchik où les 8 émotions primaires sont présentées sur une section dans le plan horizontal. Sur une section verticale, sont reportées les différentes intensités d'une même émotion primaire (par exemple : appréhension, peur et terreur).

Les deux approches, catégorielle et dimensionnelle, loin d'être opposées, sont complémentaires pour l'étude des émotions.

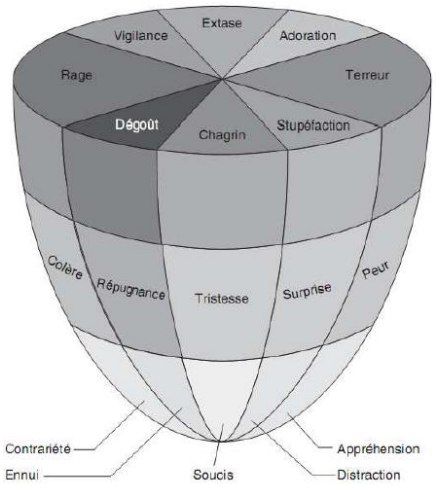


FIGURE 1.5 – La représentation de diverses émotions selon leurs intensités [170]

1.2.6 Composantes d'une émotion

D'après la définition de l'émotion, on peut noter la présence de trois composantes fondamentales : la composante comportementale, la composante physiologique et la composante cognitive/subjective [179].

1.2.6.1 Composantes physiologiques des émotions

Le comportement et l'état de l'organisme sont affectés par des réactions physiologiques périphériques qui accompagnent toute émotion. En effet, l'activité neurophysiologique établit une fonction adaptative attribuée à l'émotion. Il s'agit d'une activité sympathique qui permet à l'individu de réagir rapidement aux stimuli externes. Le système nerveux autonome (SNA) subit des modifications associées aux états émotionnels en commandant à nos viscères indépendamment de notre volonté.

Les centres d'hypothalamus qui sont une région du système nerveux sympathique commandent l'activation végétative. La mesure de cette dernière peut se révéler très précieuse dans l'étude de l'expression des émotions et des circuits cérébraux mis en jeu. Or, ces variations physiologiques sont des marqueurs temporels des variations somatiques dont les relations aux variations d'états émotionnels sont bien connues, même s'il n'existe aucun schéma établi pour une émotion donnée.

La plupart des émotions chez tous les sujets sont caractérisées par des manifestations neuro-végétatives incontestables. Mais s'il est prouvé qu'une même personne va réagir le plus souvent à un même processus, il est également sûr que deux individus, qui éprouvent une émotion semblable, ne réagissent pas physiologiquement de la même façon [179].

1.2.6.2 Composantes comportementales des émotions

L'émotion peut être révélée par un ensemble de traits comportementaux par lesquels elle se révèle, tel que l'intonation de la voix, les pleurs, le sourire ou les mimiques faciales. La fonction principale de l'expression émotionnelle est de générer un langage détectable par les autres individus.

Par la suite, nous détaillons deux composantes comportementales : les expressions faciales et la prosodie.

A. Les expressions faciales

Les visages véhiculent des informations riches qui constituent deux catégories, d'une part les indices de l'identité individuelle et d'autre part, des expressions de communication (verbale et non verbale), d'intentions et d'émotions entre individus, via, en particulier, la direction du regard et les expressions faciales.

L'être humain est aussi particulièrement doué pour reconnaître les émotions associées à des expressions faciales. Il peut donc communiquer avec d'autres personnes présentes d'une façon beaucoup plus rapide qu'avec le langage. Ceci lui permet entre autre de donner aux personnes de son entourage des feed-back sur leurs actions, pour qu'elles puissent savoir de quelle façon celles-ci sont perçues, et ainsi de modifier si nécessaire leur projet d'origine.

Ekman et al. [69] ont montré que chaque société possède des règles spécifiques qui décrivent l'expression en fonction des circonstances. Ainsi, on peut générer des mimiques d'émotions sans pour autant les ressentir. Mais l'expression faciale délibérée se distingue des émotions faciales spontanées par la séquence temporelle des unités musculaires mises en œuvre et le degré d'asymétrie faciale.

Parmi les méthodes qui permettent de mesurer l'expression faciale, il y a la technique électromyographique. Cette méthode consiste à mesurer directement l'activité électrique des muscles faciaux par électrodes appliquées sur la surface cutanée. Cette technique donne accès aux modifications latentes de l'activité faciale non visible pour l'observateur. Ces modifications se manifestent en correspondance avec l'imagerie mentale émotionnelle. La méthode de FACS (*Facial Action Coding System*) [76], également utilisée, permet le codage de toutes les unités d'actions visibles sur un visage photographié ou filmé, par exemple : baisser les paupières, bouger les lèvres, etc. Cependant, l'expression des émotions débute par les mouvements des muscles faciaux qui se produisent quelques millièmes de seconde à peine après l'évènement déclenchant [76].

B. La prosodie

Les systèmes de reconnaissance automatique de la parole donnent à la machine les capacités de transformer le signal sonore en une suite de mots. Le domaine du traitement automatique du langage permet d'accéder au sens de cette suite de mots. Partant de ces outils (relativement efficaces), il est nécessaire d'aller plus loin : la question n'est plus uniquement de savoir ce qui est dit mais aussi de connaître le contexte de prononciation de la phrase. C'est à ce niveau qu'intervient la dimension émotionnelle. Si on ne prend pas en compte l'intonation de la phrase, il est difficile de faire la différence entre une question et une affirmation. De la même façon, selon l'émotion, l'attitude, mais aussi selon la personnalité du locuteur, une même phrase peut avoir un sens différent.

Depuis quelques années, les études sur la parole émotionnelle vont au delà d'une analyse des manifestations vocales des différents états émotionnels et commencent à développer des systèmes de classification automatique des émotions. Cette évolution est née de la prise de conscience des applications industrielles potentielles du domaine des sciences affectives avec l'apparition d'un nouveau champ de recherche, le domaine de l'«affective computing » [168].

Un survol rapide de la parole émotionnelle montre que la prosodie [1] est le vecteur privilégié des émotions dans la parole [22]. Elle est le siège de l'expression directe des émotions, du codage des attitudes et des stratégies expressives pour un même matériel acoustique. Les paramètres mesurés refléteraient essentiellement la dimension d'activation émotionnelle et l'utilisation d'autres paramètres -mieux choisis - permettrait une meilleure différenciation des différents états émotionnels sur le plan acoustique.

1.2.7 Conclusion

Aujourd'hui, il est possible de détecter et de mesurer les différentes manifestations qui déterminent la nature de l'émotion. Cette détection est imparfaite, du fait d'une part, de la complexité des signaux et d'autre part, la variabilité entre les individus, il n'existe aucun modèle complet qui détermine les émotions. Cependant, plusieurs études se sont intéressées au recueil de données afin d'améliorer ce modèle. En général, les modules de reconnaissance existants prennent en considération toutes les données nécessaires soit complètes ou incomplètes pour but de reconnaître au mieux l'état émotionnel du sujet. Pour cela nous allons survoler les travaux sur la reconnaissance des émotions à partir des expressions faciales, des signaux physiologiques et des systèmes multimodaux.

1. Prosodie : partie de la phonétique qui étudie l'intonation, l'accentuation, les tons, le rythme, les pauses, la durée des phonèmes

1.3 Les expressions faciales

1.3.1 Introduction

La question de savoir si les expressions faciales sont des signes invariants, tant biologiques que culturels, des émotions est une question récurrente en psychologie [51].

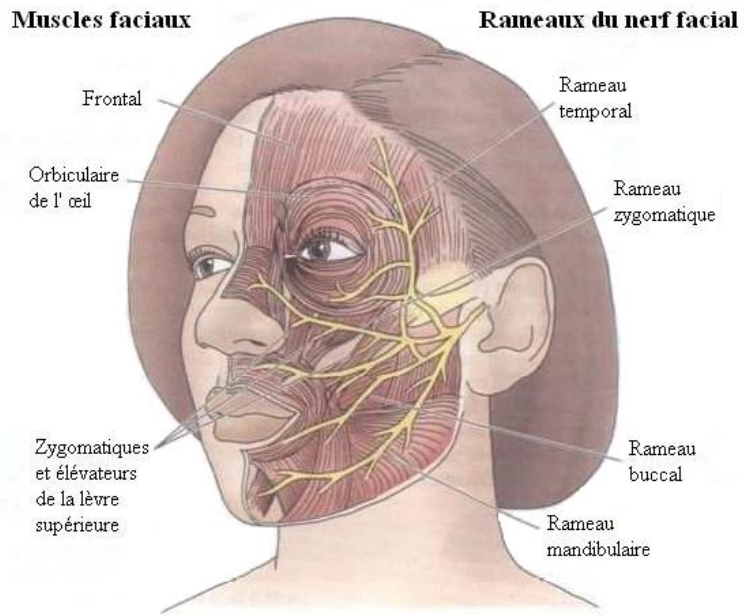


FIGURE 1.6 – Muscles faciaux et leur contrôle nerveux

L'expression faciale est un aspect important du comportement et de la communication non-verbale [8] où le changement dans le visage, perceptible visuellement, dû à l'activation (volontaire ou non) de l'un ou de plusieurs des 44 muscles composant le visage (250000 expressions possibles). L'expression faciale, déjà étudiée par Darwin et Duchenne de Boulogne au dix-huitième siècle, a joué un rôle majeur dans la recherche sur les émotions depuis les travaux de Sylvan Tomkins dans les années 1960. Ses élèves Paul Ekman et Caroll Izard ont défendu l'idée d'un nombre limité d'émotions de base auxquelles sont associées des expressions faciales automatiques, universelles et innées.

Durant la seconde moitié du dix-huitième siècle, le neurologiste Duchenne de Boulogne réalise une série d'expériences sur l'expression faciale de l'émotion. Il utilise la photographie et la stimulation électrique des muscles de la face pour mettre en évidence les mouvements associés à l'expression des émotions. Il remarque notamment que les sourires exprimant une joie sincère se différencient des sourires volontaires par la contraction d'orbicularis oculi, un muscle situé autour des yeux (figure 1.6).

Des recherches menées dans les années 1980 par Paul Ekman et son équipe ont permis de confirmer et de compléter ces résultats [72]. Ekman a mis en évidence le fait que nous sommes pour la plupart incapables de contracter volontairement l'orbicularis oculi et que ceux qui le peuvent n'arrivent généralement pas à contracter ce muscle de chaque côté au même moment.

En outre, les sourires de Duchenne sont généralement associés à une activité assymétrique dans le lobe frontal, considérée comme un signe d'affection positif [73, 74].

Par la suite, nous allons présenter les méthodes existantes pour l'analyse des expressions faciales dont l'objectif est de reconnaître les émotions associées.

1.3.2 Un système d'analyse des expressions faciales

En général, trois étapes principales peuvent être distinguées dans un système d'analyse d'expression faciale. La première étape consiste à détecter le visage, qui permet de limiter la zone de recherche. Par la suite, l'extraction des informations nécessaires qui décrivent au mieux l'expression. A la fin, en se basant sur ces informations, l'image sera affectée à une catégorie d'expressions à l'aide d'un classifieur.

La figure 1.7 schématise les différentes étapes d'un système de reconnaissance des expressions faciales.

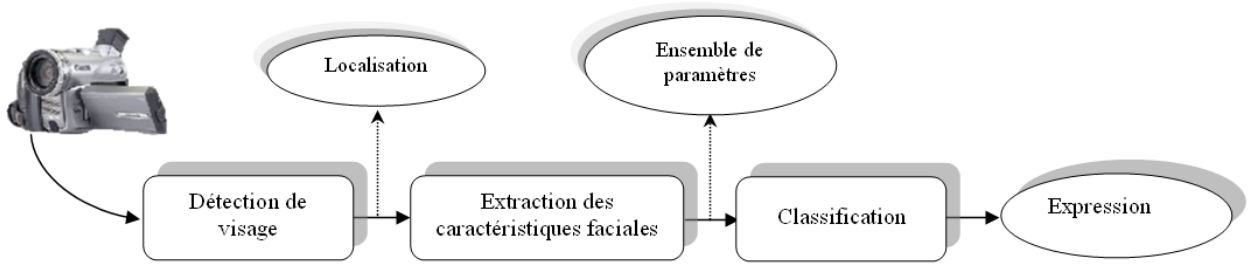


FIGURE 1.7 – Architecture d'un système de reconnaissance des expressions faciales

1.3.3 Les techniques de détection de visages

Plusieurs méthodes de détection de visages ont été proposées au cours des dernières années. On peut les classer selon quatre catégories décrites ci-dessous :

1. Méthodes basées sur les connaissances acquises ;
2. Méthodes basées sur les caractéristiques invariantes ;
3. Méthodes basées sur la mise en correspondance ;
4. Méthodes basées sur l'apparence.

Étudions chacune d'elle un peu plus en détail.

1.3.3.1 Méthodes basées sur les connaissances acquises

Les méthodes de cette catégorie se basent sur la connaissance des propriétés du visage et des relations existantes entre les différentes caractéristiques faciales. Voici quelques exemples qui peuvent être appliqués sur une image pour tenter de déterminer la présence d'un visage :

- ❒ La partie centrale d'un visage a des valeurs d'intensité lumineuse uniforme ;
- ❒ Un visage apparaît souvent avec deux yeux qui sont symétriques, un nez et une bouche ;
- ❒ Les rapports existants entre les différentes composantes faciales peuvent être représentés par leurs distances et positions relatives.

On peut citer comme exemple la méthode de G. Yang et T.S. Huang [226] basée sur une stratégie de multi-résolution.

Le problème de ces méthodes est la difficulté de traduire la connaissance des propriétés du visage humain en des règles bien définies. De plus, il est difficile d'étendre cette approche pour détecter des visages dans différentes orientations car énumérer tous les cas possibles ne semble pas réalisable. Toutefois, ces méthodes sont efficaces pour détecter des visages de face.

1.3.3.2 Méthodes basées sur les caractéristiques invariantes

Les algorithmes contenus dans cette catégorie se basent sur le fait que les caractéristiques faciales ne changent pas même quand l'orientation du visage, l'angle de prise de vue ou les conditions d'éclairage varient. On trouve ainsi dans cette catégorie des techniques basées sur la détection de la couleur de la peau, la détection des différentes caractéristiques faciales etc.

Citons comme exemple les travaux de S. McKenna, S. Gong, et Y. Raja qui se basent sur la détection de la couleur de la peau [148].

Le problème de ces méthodes est le manque d'efficacité si les images en entrée sont de faible luminosité ou si certaines parties d'un visage sont masquées. De plus, le fait qu'un visage soit ombragé rend l'extraction des caractéristiques faciales très difficiles.

1.3.3.3 Méthodes basées sur la mise en correspondance

Les techniques employées ici se font à l'aide de motifs de visage préalablement créés et judicieusement choisis. Le principe est de calculer la corrélation entre des zones de l'image d'entrée et les motifs de visages. On peut citer comme exemple les travaux de A. Lanitis, C.J. Taylor et T.F. Cootes basés sur des modèles de visage déformables [129].

L'avantage de ces techniques est qu'elles sont simples à implémenter. Toutefois, les premières méthodes proposées ne pouvaient pas faire face aux changements d'orientation, de forme et de taille des visages. C'est pourquoi des méthodes basées sur des modèles de visage multi-résolution, multi-échelles et déformables ont été proposées par la suite.

1.3.3.4 Méthodes basées sur l'apparence

Les méthodes de cette catégorie reviennent à traiter le problème de la détection de visages comme un problème de classification. Dans le but de déterminer si une image appartient à la classe des visages ou à celle des non-visages, on utilise des techniques d'apprentissage automatique supervisé. Pour cela, un ensemble de données d'apprentissage est constitué d'une part des images représentant des visages et d'autre part des images ne représentant pas des visages, qui

permettra de fixer une règle de séparation entre les deux classes. Un exemple de méthode appartenant à cette catégorie est celle développée par H. Rowley, S. Baluja et T. Kanade basée sur des réseaux de neurones [181].

Ce sont actuellement les techniques les plus efficaces et les plus adéquates pour une détection en temps-réel.

1.3.4 Extraction des caractéristiques faciales

Diverses techniques ont déjà été proposées pour l'extraction (ou bien la segmentation) des caractéristiques faciales. Elles peuvent être classées en 3 grandes familles selon les types d'informations et de contraintes qu'elles utilisent : analyse bas niveau, analyse intermédiaire et analyse haut niveau.

1.3.4.1 Analyse bas niveau

Les techniques de bas niveau supposent que les pixels de l'objet à segmenter possèdent des caractéristiques homogènes et différentes du fond. Or, la segmentation peut être effectuée par l'identification et la séparation des classes objets et fond. Pour cela, différentes solutions ont été proposées. Certaines utilisent un simple seuillage d'une grandeur colorimétrique, alors que d'autres mettent en œuvre des techniques de classification plus évoluées [81].

Zhang [240] s'est basé sur la corrélation entre l'image originale et des modèles de coins d'œil et de la bouche en respectant certaines conditions sur la géométrie du visage (pour vérifier l'exactitude) (figure 1.8) comme :

1. La ligne liant les deux coins de chaque œil et de la bouche sont parallèles avec la ligne liant des deux centres de l'œil ;
2. Les largeurs des deux yeux sont égales ;
3. Le centre de la bouche est le milieu de la ligne liant les deux coins de la bouche.

Kapmann et al. se sont basés sur le travail de Zhang [240] pour l'extraction des coins de l'œil afin de les représenter avec un modèle déformable (figure 1.9). Le rayon du cercle qui correspond à l'iris est proportionnel à la distance entre les deux coins de l'œil [111].

Cette méthode est basée uniquement sur le processus de la binarisation dans une zone approximative. Le résultat est très sensible à la valeur du seuil qui est difficile à définir.

Ko et al. proposent un seuillage adaptatif suivi par une sélection basée sur les contraintes morphologiques pour le processus de détection d'œil [119]. L'image binaire est calculée grâce à la méthode du seuil heuristique (P-title [193]). Un exemple est montré dans la figure 1.10. Les contraintes morphologiques sont utilisées pour permettre la sélection des blocs candidats tel que la taille et la forme des yeux. En se basant sur ces deux paramètres, les blocs candidats

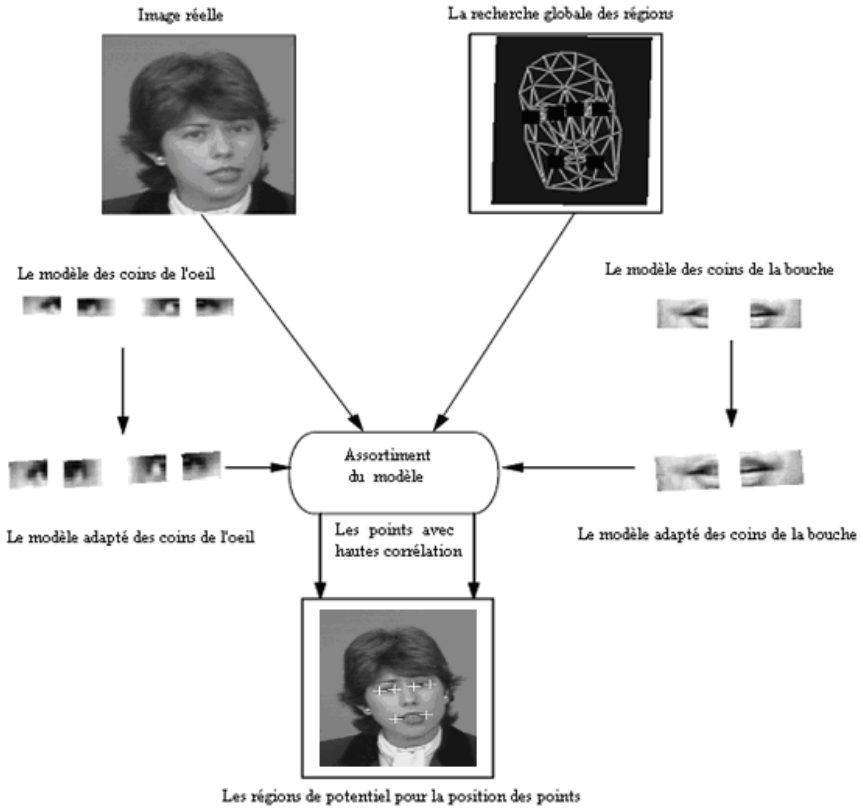


FIGURE 1.8 – La détection des zones potentielles pour la position des coins [240]

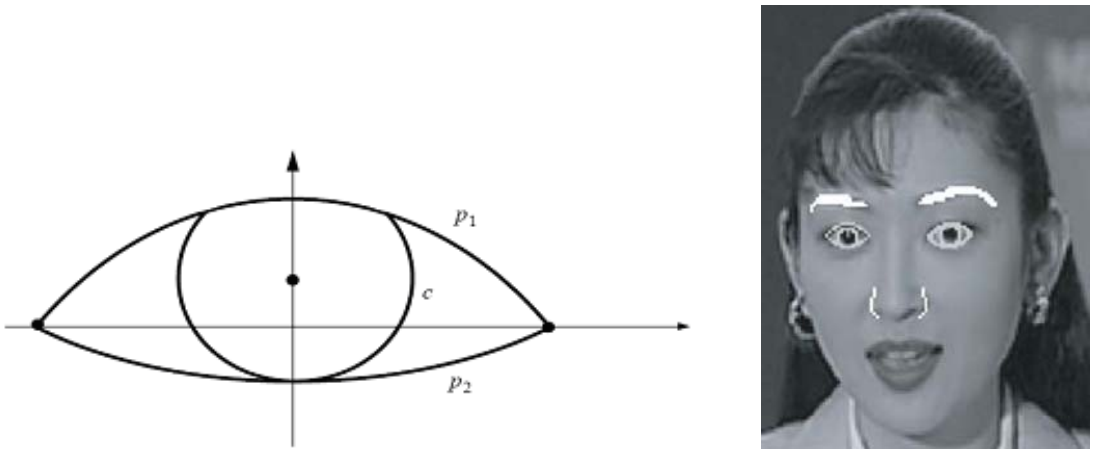


FIGURE 1.9 – Le modèle utilisé dans [111]

sélectionnés sont ceux qui vérifient certaines conditions comme la similarité entre les blocs. Pour la localisation de la bouche, les auteurs ont pris le bloc le plus large et pour les narines ils se sont basés sur l'information entre les yeux et la bouche.

Les deux paramètres (la taille et la forme des yeux) sont très dépendants des conditions de luminance, de la distance entre l'objet et la caméra et du processus de binarisation. Ainsi le processus de sélection mène parfois à des fausses détections.

Deng et al. combinent l'intensité et la luminance du gradient pour améliorer le processus de sélection [65]. Pour chaque visage détecté, la projection horizontale du contour vertical et la

(a) (b) (c) (d)

FIGURE 1.10 – Localisation des yeux : a- Image d'origine, b- Image binarisée, c- Localisation des yeux, d- Résultat d'extraction [119]

projection horizontale de l'intensité sont calculées (figure 1.11). En se basant sur les pics des deux projections, la région des yeux sera localisée. Cette méthode est sensible aux lunettes, les moustaches et les cheveux longs peuvent fausser le résultat à cause de leur contour vertical et de leur intensité.

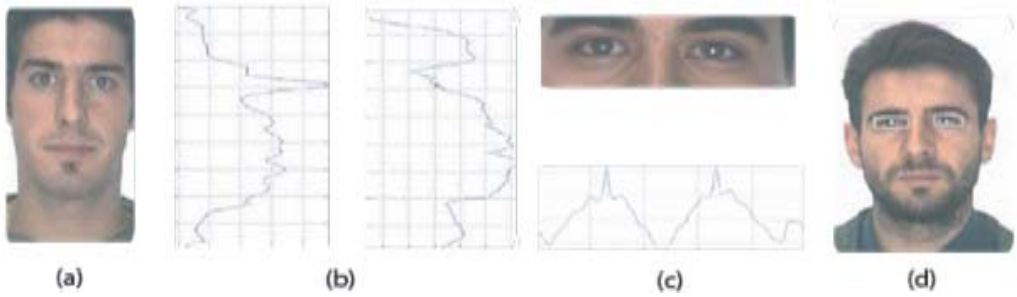


FIGURE 1.11 – a- Image d'origine, b- La projection horizontale, c- La projection verticale de la transition verticale, d- Extraction des yeux dans l'image [65]

Pour améliorer les résultats des techniques précédentes qui ne donnent pas une segmentation régulière des caractéristiques du contour, Vezhnevets et al. ont introduit un modèle pour le contour de l'œil basé sur un cercle pour l'iris et une courbe cubique pour la haute paupière et une courbe quadratique pour la basse paupière [212]. Leur méthode utilise des images en couleur contenant un seul œil. Cette méthode est constituée de trois étapes : la détection approximative du centre de l'œil basée sur le canal rouge, extraction de la forme de l'iris et l'extraction des courbes de la paupière (figure 1.12).

FIGURE 1.12 – a- Ensemble des points de vallée de la luminance, b- Ensemble des lignes principales, c- Fitting d'une courbe polynomiale cubique, d- Résultat de la segmentation [212]

Hammal a proposé pour la segmentation des traits du visage des algorithmes pour extraire les contours de l'iris, des yeux et des sourcils basés sur la maximisation d'un flux de gradient de

luminance autour des contours des traits recherchés [96]. Les modèles proposés sont flexibles et robustes aux conditions d'éclairage (figure 1.13), à l'origine ethnique, au port de lunettes et aux déformations de ces traits qui peuvent survenir sur le visage en cas d'expression faciale. Pour la segmentation des lèvres, elle a utilisé le travail proposé dans [81].

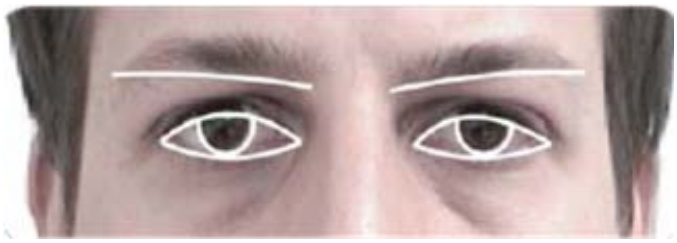


FIGURE 1.13 – Les résultats de la segmentation de l'œil et du sourcil [96]

1.3.4.2 Analyse intermédiaire

A un niveau d'analyse intermédiaire, on cherche à détecter des caractéristiques indépendantes des conditions lumineuses et de l'orientation des visages [81]. Les contours actifs (ou *snakes*) introduits par Kass et Witkin à la fin des années 80 [113] sont des courbes qui peuvent se déformer progressivement de manière à s'approcher au plus prés du contour d'un objet. Cette déformation est guidée par la minimisation d'une fonction d'énergie comprenant deux termes : une énergie intérieure E_{int} qui permet de régulariser le contour et une énergie extérieure E_{ext} reliée à l'image et aux contraintes particulières que l'on peut ajouter.

Radeva a utilisé le contour « active rubber » pour la segmentation qui nécessite une localisation initiale des caractéristiques faciales effectuée par le choix de la projection horizontale et verticale [175].

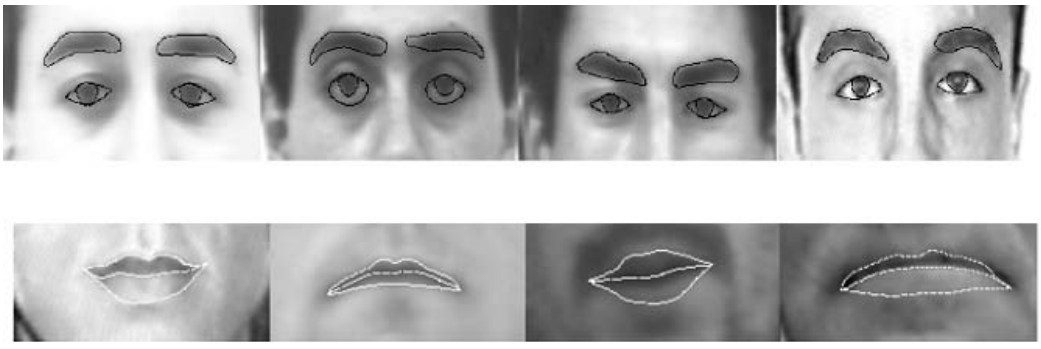


FIGURE 1.14 – Exemple de segmentation des yeux et de la bouche en utilisant les contours actifs de rubber [175]

Pardas et al. ont appliqué les contours actifs pour la sélection des positions des points et pour avoir l'information du mouvement, ils ont pris en considération l'image précédente et l'image courante. La sélection est calculée par l'implémentation du programme dynamique (DP) de l'algorithme du contour actif. Ils ont ajouté un nouveau terme pour introduire l'information du mouvement et donner plus de flexibilité à l'algorithme [163].

FIGURE 1.15 – Exemple de suivi des caractéristiques faciales [163]

1.3.4.3 Analyse haut niveau

Les méthodes de bas et de moyen niveau décrites dans les sections précédentes sont des processus à forme libre. Elles n'intègrent aucune connaissance a priori sur les formes admissibles. A l'opposé, les méthodes de haut niveau sont basées sur des modèles caractéristiques des formes à segmenter, obtenus de manière heuristique ou statique. Ces modèles génériques sont déformés de manière à être adaptés aux contours de l'objet.

Malciu et al. [145] ont proposé une méthode basée sur les modèles déformables. La fonction de l'énergie interne est calculée pour incorporer la rigidité et la contrainte de la symétrie interne. Le modèle est un système à base de ressorts connectant les coins des caractéristiques faciales (figure 1.16-a, b). La fonction de l'énergie externe sert à maintenir l'uniformité entre le modèle géométrique et l'image. La déformation optimale du modèle est estimée par la minimisation de la fonction d'énergie totale en utilisant la méthode de simplex [173].

FIGURE 1.16 – Les résultats de segmentation de la méthode de [145]

Botino [29], Tian et al. [202] ont adopté une approche rapide sans optimisation itérative de la fonction du coût. Ils ont utilisé trois points pour les sourcils avec deux segments connectés. Pour les yeux, ils ont utilisé un modèle multi états. Le modèle de l'œil ouvert correspond à deux

paraboles et un cercle pour l'iris, par contre l'œil fermé est modélisé par un segment liant les deux coins de l'œil. La détection est basée sur l'étiquetage manuel des points caractéristiques dans la première image. Dans le reste de la séquence, la localisation devient automatique en utilisant l'algorithme de Kanade [141].

Plusieurs travaux [42, 8, 222] ont utilisé les modèles actifs d'apparence (*Active Appearence Models* AAM) qui permettent de construire un modèle statique de l'objet à segmenter incluant à la fois la forme et les niveaux de gris. La description de ce type de modèle est calculée à l'aide de l'ACP (analyse en composantes principales). Dans [8], les auteurs ont proposé à partir d'une seule photo, d'annuler l'expression d'un visage quelconque, puis de synthétiser une expression faciale artificielle sur ce même visage. Deux approches pour la modélisation d'expressions faciales par régression linéaire sur un ensemble d'apprentissage ont été abordées. Les modèles linéaires (direct et évolutif) ainsi obtenus ont été utilisés pour le filtrage et la synthèse d'expressions faciales.

1.3.4.4 Synthèse sur l'extraction des caractéristiques

Les techniques de segmentation présentées peuvent être classées en trois grandes familles selon les informations utilisées (locales ou globales), on distingue trois niveaux d'analyse (bas, intermédiaire et haut).

L'analyse de bas niveau permet une localisation rapide et assez précise des caractéristiques faciales mais ne permet absolument pas de détecter leurs contours d'une manière fiable. Dans le cas d'une analyse de niveau intermédiaire, la détection des formes très variées est assurée par la grande déformabilité des contours actifs mais les zones de faible gradient comme les commissures des lèvres sont difficiles à segmenter. Par contre, les méthodes haut niveau permettent d'obtenir des formes admissibles en se basant sur des modèles caractéristiques, l'inconvénient majeur de ceci réside dans la sensibilité aux conditions d'éclairage et à l'angle de prise de vue.

Afin d'obtenir une extraction précise et rapide des caractéristiques faciales, il semble nécessaire d'utiliser une méthode hybride basée sur le principe des trois niveaux. L'objectif de cette partie est de proposer un modèle anthropométrique qui permet d'aboutir toujours à une forme admissible (niveau 3) et une modélisation des muscles faciaux basée sur des techniques de bas niveau (niveau 1).

1.3.5 Classification des expressions basée sur des données statiques

1.3.5.1 Approches basées sur des modèles

Plusieurs systèmes de reconnaissance des expressions faciales ont employé les techniques basées sur des modèles. Certains d'entre eux appliquent un processus de déformation de l'image pour tracer le visage dans un modèle géométrique. D'autres réalisent une analyse locale là ou les nœuds (*kernel*) sont localisés spatialement pour filtrer les caractéristiques faciales extraites. De nombreux travaux appliquent l'analyse globale des niveaux de gris basée sur l'ACP (analyse

en composantes principales), l'analyse des ondelettes de Gabor ou les approches basées sur des valeurs propres et l'espace de Fisher.

Hung et Huang [101] en premier lieu, calculent les 10 paramètres d'action APs (*Action parameters*) (Figure 1.17-a) en se basant sur la différence entre les paramètres des caractéristiques du modèle dans un visage neutre et ceux des expressions faciales examinées des mêmes personnes. Les deux premières composantes de l'ACP ont été utilisées pour représenter la variation des APs. La classification de ces composantes est effectuée avec la distance minimale pour le traitement de six expressions.

(a) (b) (c)

FIGURE 1.17 – a- Les paramètres d'action APs [101], b- Graphe élastique *Gabor-labeled* pour une image faciale [142], c- La moyenne des régions de mouvements [18]

Lyons [142] se base sur la représentation des ondelettes de Gabor pour la classification des 6 expressions universelles plus l'expression neutre. Un modèle de 34 points des caractéristiques faciales est manuellement initialisé dans le visage (figure 1.17-b). Les coefficients de l'ondelette de Gabor sont calculés pour chaque point du maillage. Ensuite, une analyse discriminante linaire est utilisée afin d'agréger le résultat des vecteurs dans des groupes ayant des différents attributs faciaux. Enfin, la classification est effectuée par la projection du vecteur d'entrée de l'image test le long du vecteur de distinction afin de l'affecter au groupe le plus proche.

Anderson [18] utilise les mouvements faciaux pour caractériser des images monochromes d'une vue frontale du visage (figure 1.17-c). Le flux optique du visage est déterminé en utilisant un modèle de gradient pour une implémentation en temps réel. A la fin, les mouvements produits sont classés par un classifieur SVM (séparateurs à vastes marges) en une de six expressions universelles.

Dubuisson [66] adopte l'ACP (analyse en composantes principales) pour la réduction de la dimension du visage et l'ADL (analyse discriminante linéaire) pour générer l'espace de Fisher où le nouvel échantillon est classé. Figure 1.18 montre 5 espaces de Fisher générés et appliqués au corpus d'apprentissage contenant les 6 classes des expressions faciales. Ensuite, le classifieur de l'arbre de décision [43] est utilisé comme une mesure entre la projection des échantillons de test et chaque vecteur moyen d'estimation dans l'espace de Fisher. Les performances de la

méthode de la reconnaissance dépendent fortement de la précision de l'étape du post-traitement où la normalisation et l'enregistrement de chaque nouveau visage sont requis et manuellement réalisés.

FIGURE 1.18 – Les 5 espaces de Fisher correspondant aux 5 axes du sous espace généré par ACP et ADL[66]

Gao [90] propose une approche de classification de trois expressions : neutre, sourire et colère appliquant une architecture basée sur la ligne. Leur approche utilise la ligne du contour (LEM : *Line Edge Map*) [89] comme une expression descriptive (figure 1.19-a). La classification est obtenue en calculant 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 (figure 1.19-b).

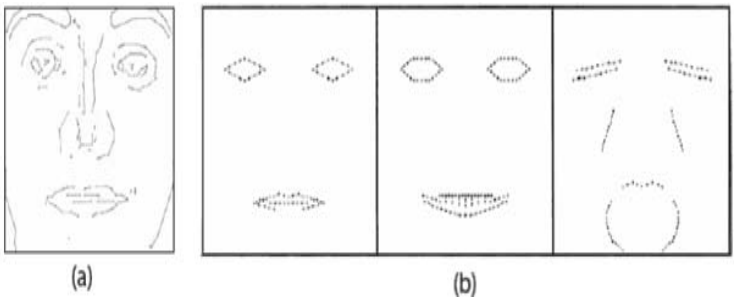


FIGURE 1.19 – a- LEM du visage, b- le modèle de l'expression faciale [90]

Abboud et al. [8] proposent un modèle automatique pour la classification des expressions faciales en se basant sur les modèles d'apparence active (MAA). Chaque image de visage est représentée par un vecteur de MAA correspondant. Ensuite, ils utilisent la distance de Mahalanobis [97] pour mesurer la distance entre le vecteur MAA et le vecteur moyen dans l'espace Fisher. Pour chaque configuration, le visage testé est assigné à la classe qui a la moyenne la plus proche.

1.3.5.2 Approches basées sur des points caractéristiques

La représentation de l'information faciale par l'analyse géométrique des caractéristiques a été utilisée fréquemment ces dernières années. Dans ces approches, le mouvement facial est quantifié en mesurant le déplacement géométrique de points des caractéristiques entre les images courantes et l'initiale.

Tian et al. [204] utilisent deux réseaux de neurones séparés pour la reconnaissance des 6 hauts et 10 bas des AUs (*Action Units*) basés sur les caractéristiques faciales permanentes (œil, sourcil et la bouche) et transitoires (les rides des expressions faciales, sillons). Les descripteurs de ces caractéristiques permanentes et transitoires sont l'entrée du réseau de neurones (figure 1.20). Celles-ci sont extraites manuellement dans la première image et tracées dans le reste de la séquence. La reconnaissance des expressions faciales est réalisée par la combinaison du haut et du bas des AUs.

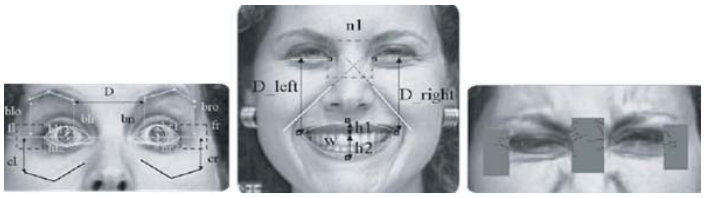


FIGURE 1.20 – Les paramètres choisis par Tian dans [204]

Pantic et Rothkrantig [159] utilisent le modèle des points du visage avec les deux vues frontale et latérale pour la classification des expressions faciales (figure 1.21). Ils codent plusieurs points caractéristiques (comme les coins de l'œil, les coins de la bouche etc.) dans AUs en utilisant un ensemble de règles. Par la suite, FACs (*Facial Action Coding System*) [76] est utilisé pour reconnaître les six expressions universelles. La classification est réalisée par une combinaison du descripteur de code AU de l'expression observée par rapport aux règles du descripteur FACs.

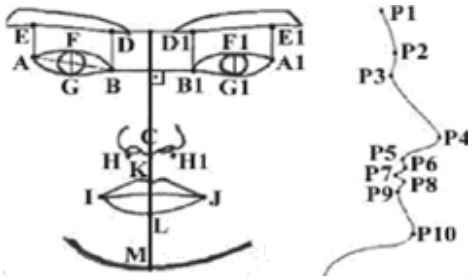


FIGURE 1.21 – Le modèle des points caractéristiques pour la vue frontale et pour la vue de face [159]

Pandas et al. [162] et Transpatsoulis et al. [207] proposent une description des six expressions faciales universelles en utilisant l'ensemble des paramètres de la définition faciale (*Facial definition parameter set* FDPs) MPEG-4 [201] (figure 1.22-b). Tsapatsoulis et al. [207] utilisent tous les FAPs (définie dans [201]) et proposent une classification basée sur un système d'inférence flou. Ils se basent uniquement sur la segmentation du contour des sourcils et de la bouche.

Pardas et al. [162] utilisent 8 points faciaux pour les sourcils et 10 pour la bouche pour le processus de classification qui est basé sur les chaînes de Markov qui assignent à l'entrée l'expression avec la probabilité la plus élevée.

A partir de l'information portée par les contours, il a été montré qu'il est possible de re-

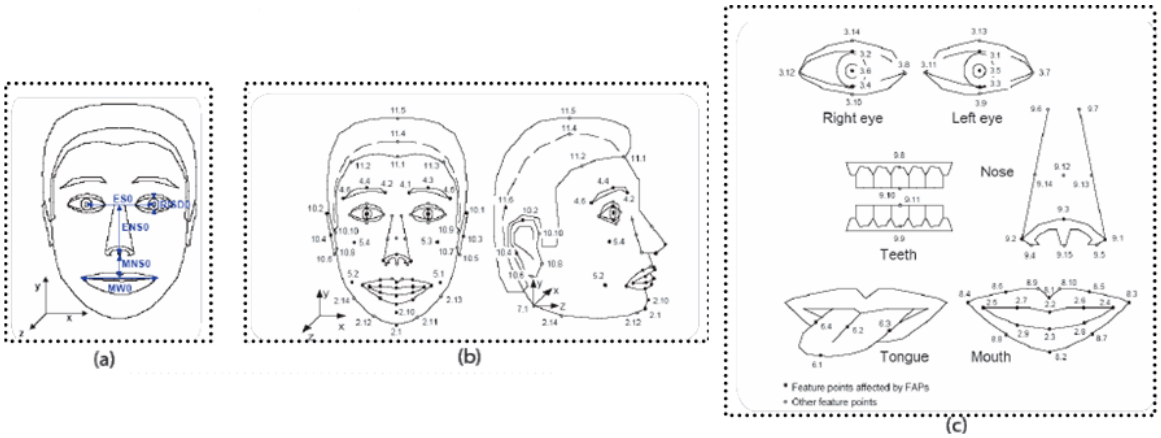


FIGURE 1.22 – a- Le modèle du visage dans le cas neutre, b- et c- Les paramètres de l'animation faciale [162]

connaître les expressions faciales [96]. Cinq distances caractéristiques sont définies à partir des traits permanents du visage. Les déformations par rapport à l'état neutre sont utilisées pour la modélisation des expressions faciales à partir du Modèle de Croyance Transférable (MCT) afin d'assurer l'invariabilité entre les individus.

Conclusion :

D'une manière générale, les approches décrites sont similaires. Elles extraient quelques caractéristiques du visage dans un premier temps puis réalisent une classification. Elles diffèrent uniquement dans la façon d'extraire les caractéristiques faciales.

1.3.6 Classification basée sur des données dynamiques

Bassili [24] a montré que la reconnaissance des expressions faciales dans des séquences vidéo est plus précise que celle des images statiques. En se basant sur ce principe, quelques auteurs ont travaillé sur des modèles dynamiques pour la déformation des caractéristiques faciales.

Coh et al. [50] proposent un cadre hiérarchique des modèles de Markov cachés MMC (figure 1.23) pour une segmentation automatique et une reconnaissance des expressions faciales dans une séquence vidéo. Le premier niveau de l'architecture est composé des MMCs indépendants correspondants aux six expressions universelles. Dans chaque image, les caractéristiques du mouvement par rapport à l'état neutre sont utilisées comme entrée des MMCs. Leurs sorties sont utilisées comme une observation pour le niveau haut de l'architecture qui est constitué de 7 états, 6 pour les expressions universelles et un pour l'état neutre. La transition entre les états est imposée par le passage par l'état neutre. Le modèle permet d'obtenir la probabilité de l'expression montrée sur la séquence. La reconnaissance est donnée par le décodage à chaque instant de l'état des MMCs du haut niveau. Chaque séquence d'expressions contient plusieurs instances

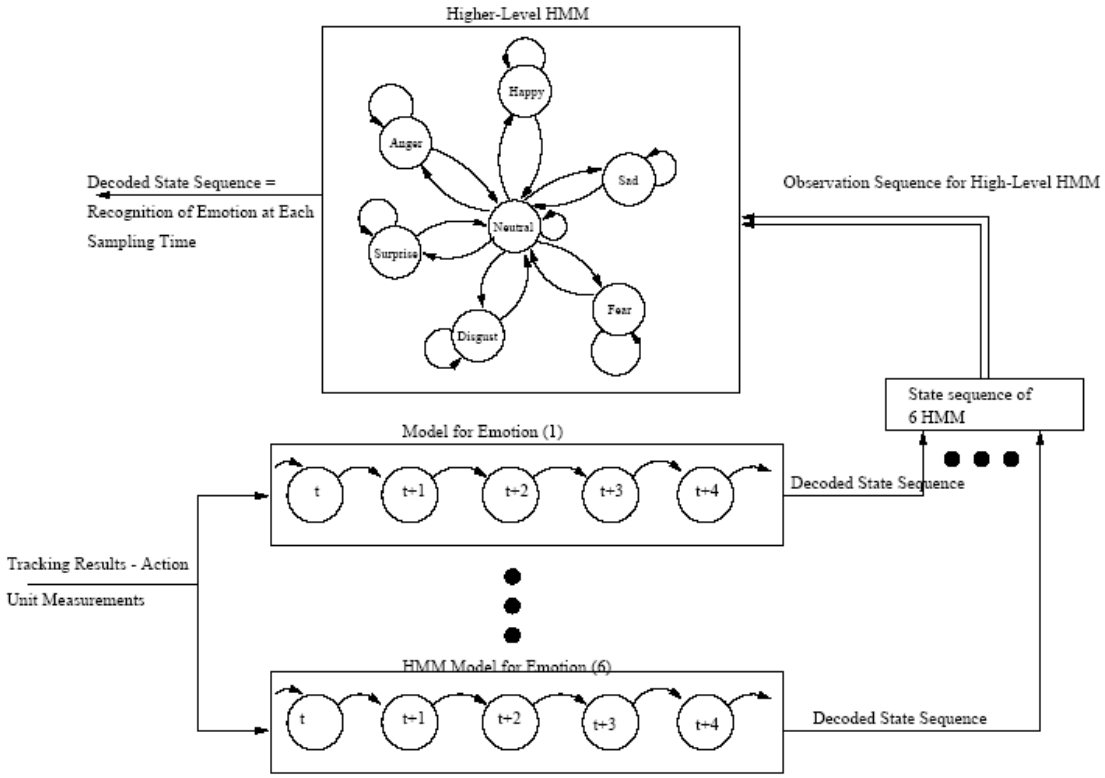


FIGURE 1.23 – Architecture multi niveau des MMC pour la reconnaissance dynamique des émotions [50]

séparées par un état neutre.

Zhang et Qiang proposent une technique pour la fusion de l'information multi-sensorielle basée sur le réseau Bayésien [241]. Les caractéristiques faciales permanentes sont automatiquement détectées dans la première image et suivies dans le reste de la séquence. Le rapport géométrique entre les caractéristiques faciales permanentes (les yeux, les sourcils, le nez et la bouche) et les caractéristiques transitoires est utilisé pour caractériser les unités d'action AUs détectées (figure 1.24). Leur modèle du réseau Bayésien est constitué de trois couches : classification, AUs et la couche de l'information sensorielle. La reconnaissance des expressions faciales est réalisée en fusionnant le résultat actuel avec les résultats des images précédentes. Par conséquent, la reconnaissance devient plus robuste et précise grâce à la modélisation temporelle des expressions.

Pantic et Pantras [157, 158] proposent une méthode pour la reconnaissance des segments temporels (début, apex, fin) des unités d'action faciales (AUs) durant la séquence de l'expression faciale. Ils initialisent 20 points caractéristiques frontaux (figure 1.25-b) et 15 points faciaux du profil (figure 1.25-a) dans la première image et les suivent dans le reste de la séquence. La déformation faciale est codée en AUs et elle est divisée en trois segments : onset ->début, apex->pic, offset->fin. En se basant sur la vue frontale, la méthode basée sur des règles est définie pour coder un segment temporel de 27 AUs ou une combinaison des AUs pour chaque groupe de 5 images consécutives.

Dans [96], Hammal a proposé d'introduire une information temporelle dans le modèle de

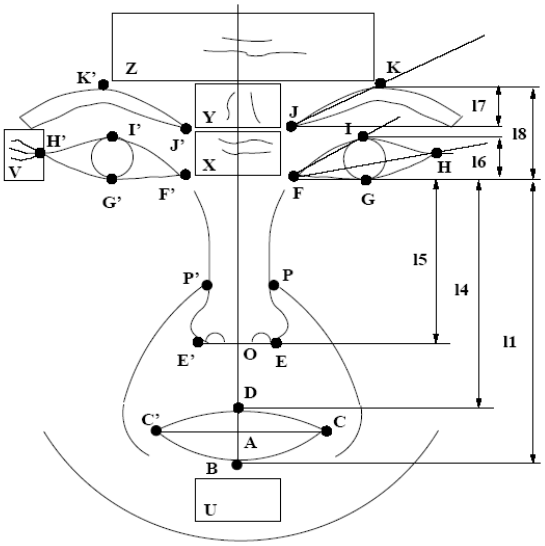
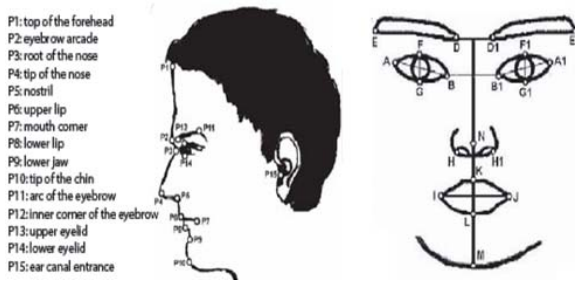


FIGURE 1.24 – Le rapport géométrique des points caractéristiques faciaux où les rectangles représentent la région des sillons et les rides [241]



(a) (b)

FIGURE 1.25 – a- Les points de profil du visage [158], b- Les points de face du visage [157]

croyance transférable. La reconnaissance est basée sur la combinaison de l'ensemble des déformations des traits du visage entre le début et la fin de la séquence de l'expression. Cette amélioration permet d'être plus robuste aux erreurs ponctuelles de segmentation et aux déformations asynchrones des traits du visage.

1.3.7 Synthèse sur la classification des expressions faciales

Il est difficile de concevoir un modèle physique déterministe qui représente les propriétés géométriques du visage et les activités du muscle. Les approches globales impliquent un temps intensif pour l'apprentissage. Le modèle appris est souvent incertain pour une utilisation pratique due à la variation interpersonnelle.

Les méthodes basées sur des points sont probablement les plus rapides mais elles présentent deux inconvénients [12] :

1. L'utilisation unique de la position de certains points (information locale) pour calculer le modèle est non robuste vis-à-vis du bruit, contrairement au calcul intégral basé sur l'information globale ;

2. Ces méthodes sont basées sur la précision de l'algorithme de suivi.

Afin d'obtenir un système adapté à une application en temps réel, nous avons opté pour une méthode rapide basée sur les points. Pour assurer plus d'informations sur l'expression faciale, notre modèle anthropométrique basé sur les points caractéristiques permet d'avoir une information plus globale que locale en utilisant par exemple 10 points autour de la bouche au lieu de 4 [15].

Référence	Caractéristiques	Classifieur	classe	Type	Sujet	Taux
Cohen et al. 04 [49]	Ondelettes de Gabor & modèle de la forme	*LDC*	3 AUs	Im	21	76 %
El Kaliouby et al. 04 [109]	24 points faciaux	*RBD*	6	Vi	30	77.4 %
Fasel et al. 04 [83]	Intensité des niveau de gris	*NN*	7	Im	?	38 :68 %
Sebe et al. 04 [191]	Mouvement de 12 unités	*KNN*	4	Im	Ck :53 SD :28	93 % (CK) 95 % (SD)
Gunes et al. 05 [94]	Caratéristiques de la forme & flux optic	C4.5 , réseau Bayésien	8	Im	FABO :4	80 %
Loannou et al. 05[104]	FAPs	Réseau neuroflou	3	Im	?	78 %
Lee et al.05 [132]	L'intensité des pixels dans la région du visage	Modèle décomposable	6	Im/Vi	CK :8 OD :16	39.58 % 61.85 %
Pantic et al. 06 [161]	Points faciaux du profil	Méthode basée sur des règles	27 AUs	Vi	MMI :19	86.3 %
Yeasin et al. 06 [228]	L'intensité des pixels du visage	*KNN+HMM*	6	Im	CK :97 OD :21	90.7 % (CK) 72.82 % (OD)
Whitehill et al. 06 [221]	Descripteurs de HAAR	*Adaboost*	11 AUs	Im	?	92.35 %
Wang et al. 06 [217]	Labels de la surface 3D	*LDA*	6	Im	BU :60	83.6 %
Zeng et al. 06 [232]	Texture avec LPP	*SVDD*	2	Im	AAI :2	79 % (homme) 87 % (femme)
Littlewort et al. 07 [138]	Ondelettes de Gabor	*Adaboost* & *SVM*	2	Vi	26	72 %
Tong et al.07 [206]	Ondelettes de Gabor	*Adaboost* & RBD	14 AUs	Vi	CK :100 OD :10	93.2 % (OD) 93.3 % (CK)
Valstar et al. 07 [209]	20 points faciaux	*Gentle SVM-sigmoide*	2	Vi	MMI :52	94 %

Im : reconnaissance basée sur l'image ;
Vi : reconnaissance basée sur la vidéo ;
AAI,CH, SAL, FABO et SD sont les noms des bases de données.

TABLE 1.5 – Comparaison des algorithmes de reconnaissance des expressions faciales [235]

1.3.8 Conclusion

La table 1.5 résume les approches de classification des expressions faciales. En général, les performances varient en fonction de la base de données utilisée, le nombre de personnes utilisées et le type de l'émotion traitée.

La recherche vise à identifier les émotions avec des modalités différentes, le plus souvent à partir de l'analyse des expressions du visage [160, 54]. Toutefois, il est assez facile de cacher une expression faciale. En outre, cette chaîne n'est pas disponible en permanence, car les utilisateurs ne sont pas toujours en face d'une caméra. Enfin, ces mesures résultent d'un processus émotionnel qui pourrait avoir débuté bien avant l'expression extérieure de l'émotion et empêche le système d'anticiper des réactions. Nous croyons que l'utilisation de signaux physiologiques permet de faire face à ces problèmes. Ces signaux proviennent du système nerveux périphérique (principalement le système nerveux autonome) et du système nerveux central (le cerveau).

1.4 Les signaux physiologiques

1.4.1 Introduction

L'expression des émotions est étroitement liée au système nerveux végétatif. Elle met en jeu, de ce fait, tous les centres cérébraux qui contrôlent les neurones préganglionaires du tronc cérébral et de la moelle. Les signes les plus nets de l'excitation émotionnelle concernent les changements d'activité du système nerveux végétatif. Selon les émotions, on pourra ainsi observer des augmentations ou des diminutions de la sudation, de la fréquence cardiaque, du débit sanguin cutané (rougissement ou pâleur), de la piloérection et de la motilité intestinale.

Les indices physiologiques qui sont couramment utilisés pour caractériser les deux composantes valence (le caractère positif ou négatif de l'expérience émotionnelle) et intensité (le degré d'activation de l'expérience émotionnelle) d'une émotion, sont :

1. la réponse électrodermale RED (*Skin Conductance* SKC) ;
2. le volume sanguin impulsionnel (*Blood Volume Pulse* BVP) ;
3. le signal du volume respiratoire (VR) ;
4. l'activité électromyographique (EMG) ;
5. la température cutanée (*Skin Temperature* SKT) ;
6. la fréquence cardiaque (Fc) ;
7. le rythme électroencéphalogramme (EEG).

L'activation de ces différents indicateurs varie en fonction de l'émotion considérée et des sujets, ce qui induit un patron de réponse complexe permettant de distinguer les différentes émotions. Dans ce travail, les cinq premiers indices sont particulièrement utilisés.

1.4.2 Les modifications physiologiques concomitant des émotions

Il est très difficile de trouver un lien univoque et systématique entre une émotion donnée et une activation physiologique caractéristique.

On peut en effet observer qu'un grand nombre de ces modifications sont communes à plusieurs émotions, ce qui a suscité l'idée que ces activations seraient spécifiques et générales. 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.

Il a était montré aussi que la peur, la colère et la tristesse sont associées à une augmentation du rythme cardiaque plus que la joie. Le dégoût qui abaisse le rythme cardiaque est associé à une diminution de la température cutanée. La conductance de la peau augmente après un état d'amusement et diminue après l'état neutre et reste la même après une tristesse [199].

1.4.3 L'activité physiologique et l'activation émotionnelle

1.4.3.1 Activité électrodermale

Cette activité constitue un des indices physiologiques les plus fréquemment utilisés dans un grand nombre d'explorations en psychologie, psychophysiologie et neuroscience cognitive.

L'activité électrodermale ou AED est une donnée physiologique consistant en une évaluation du niveau de conductibilité électrique de la peau. Cette activité électrique de la peau varie très sensiblement dans les situations mettant en jeu les émotions [179].

Cependant, lorsque l'on est soumis à une émotion, l'activité électrodermale témoigne de l'existence de courants électriques cutanés, associés à la sudation qui va améliorer la conductibilité de la peau. Cette sudation résulte de la sécrétion des glandes sudoripares dites glandes eccrines, qui ont la particularité de répondre à des expériences émotionnelles. Ces glandes sont localisées dans la paume des mains et la plante des pieds [199].

Sur un enregistrement de conductance dermale (figure 1.26), le niveau électrodermal (NED) correspond à la ligne de base. C'est un indicateur de l'activation générale de l'organisme. Il peut présenter des dérives lentes et des variations transitoires, consécutives ou non à une stimulation ou à une action du participant (mouvements, respiration forte). Une variation transitoire, survenant une à trois secondes après le début d'une cause identifiée, est appelée réponse électrodermale (RED). L'amplitude des RED traduit l'importance de la réaction phasique à une stimulation, souvent de nature émotionnelle. Toute variation transitoire, survenant en dehors de cette fenêtre de latence est considérée comme une fluctuation spontanée (FS). Le NED et la fréquence des FS sont des paramètres toniques. L'AED peut avoir une valeur de trait individuel stable : la labilité (antonyme : stabilité) est caractérisée par une fréquence élevée de FS, souvent

associée à un défaut d'habituation des RED en cas de stimulations répétées [152].

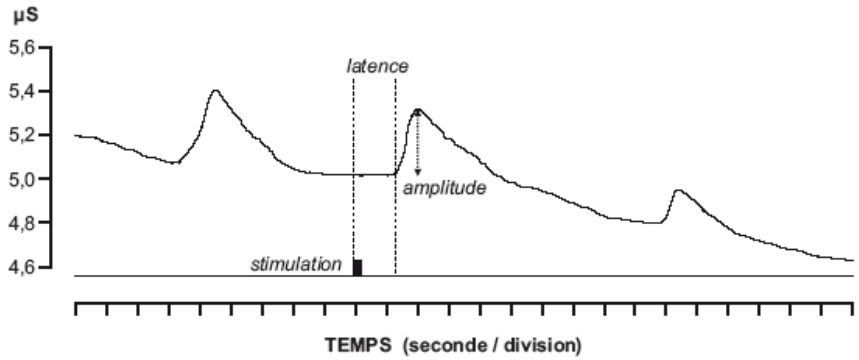


FIGURE 1.26 – Tracé d'activité électrodermale (conductance exprimée en MicroSiemens) montrant une dérive lente du niveau de base sur lequel se greffent trois fluctuations transitoires : au centre, une réponse électrodermale induite par une stimulation et, de part et d'autre, une fluctuation spontanée [152]

1.4.3.2 Pression sanguine volumique (*Blood volume pulse* BVP)

La pression sanguine volumique est un indicateur de l'écoulement du sang à travers le corps humain. Le BVP diminue sous l'effort et le stress puisque le sang est détourné vers les muscles qui travaillent afin de les irriguer et les préparer à une action imminente. Ceci signifie que l'écoulement de sang est réduit aux extrémités et, donc, aux doigts.

Il n'est pas nécessaire d'employer un autre capteur pour mesurer la fréquence cardiaque (HR), elle se déduit de la mesure du BVP. Sous l'effet du stress, la fréquence cardiaque augmente puisque le rythme du cœur s'accélère pour envoyer davantage de sang vers les muscles et les préparer à l'action [68].

1.4.3.3 Volume et rythme respiratoire (VR)

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).

A chaque inspiration normale, 0,5 litres d'air entre dans les poumons. Le volume d'air qui pénètre « en plus » au cours d'une inspiration forcée est de 2,5 à 3 litres. En fin d'expiration normale, on peut encore « chasser en plus » 1 litre d'air : on effectue alors une expiration forcée. En fin d'expiration forcée, il reste encore 1,5 litres d'air dans les poumons ; on ne peut donc jamais les vider complètement [2].

En effet, un état de repos et de relaxation est caractérisé par une respiration plus lente et plus superficielle. Par contre, les respirations plus profondes sont engendrées par des excitations émotionnelles et des activités physiques. Un état de stress sera donc décelable par une respiration fréquente, cependant, des agents stresseurs ponctuels, comme le sursaut, provoque un arrêt momentané de la respiration [87]. Mais pour la peur, on trouve une accélération des mouvements

de respiration. Les émotions à valence négative généralement causent des respirations irrégulières.

1.4.3.4 Activité électromyographique (EMG)

Le signal élecromyographie représente l'enregistrement d'une série d'évènements électriques (potentiel membranaire du muscle) produits par la fibre musculaire lorsque les muscles se contractent. Cette mesure est à prendre en compte pour déterminer l'état émotionnel du sujet.

En effet, le tonus émotionnel est une contraction involontaire, permanente et modérée des muscles, entretenue par des influx nerveux. Cette légère tension qui affecte constamment tout muscle au repos est l'expression des variations de l'émotion comme un état de stress mental. Il est montré que l'activité musculaire augmente durant le stress et les émotions à valence négative [35].

1.4.3.5 Température cutanée (*Skin Temperature* SKT)

La température périphérique, telle que mesurée aux extrémités du corps, varie en fonction de l'irrigation sanguine dans la peau.

En effet, les variations de la température de la peau sont liées à la vasodilatation des vaisseaux sanguins périphériques induite par une augmentation de l'activité du système sympathique. Cette variation dépend de l'état de la personne. Si elle est stressée, la température des extrémités de son corps diminue, car le sang est acheminé en priorité vers les organes vitaux [2], par mesure de protection. Ses doigts ont alors tendance à être plus froids et si la personne est relaxée la température des doigts augmente. Si la personne a peur, le sang va se diriger vers les muscles qui commandent le mouvement du corps comme le muscle de la jambe préparant la fuite et entraînant des températures basses aux extrémités du corps à cause de la vasoconstriction [19].

L'activation de ces différents indicateurs varie en fonction de l'émotion considérée et des sujets, ce qui induit un patron de réponse complexe permettant de distinguer les différentes émotions. La question de savoir si ces variations de paramètres physiologiques sont ou non spécifiques d'une émotion donnée a fait l'objet de longs débats.

1.4.4 Recherche antérieure sur la reconnaissance des émotions à partir des signaux physiologiques

Dans l'ensemble, le principe est le même, induire l'émotion à partir des images, des films ou de la musique, pour traiter les informations issues de certains capteurs physiologiques. Trois étapes sont nécessaires pour la reconnaissance des émotions à partir des signaux physiologiques :

2. organes vitaux comme par exemple le cœur, les poumons, l'estomac, les intestins, le foie et la vésicule bilière, le pancréas, les reins et les différentes glandes.

1. Le choix des capteurs pour l'acquisition ;
2. Le traitement des données et l'extraction des caractéristiques ;
3. Le choix de la méthode de classification.

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.

Lanzetta et al. [130] ont choisi le ton vocal, des expressions faciales et le choc électrique pour induire la joie et la peur. Ils ont traité la conductance de la peau de 37 hommes et 23 femmes (60 personnes) avec la méthode d'ANOVA (*ANalysis Of VAriance*). Leur travail a montré que la peur produit un niveau plus élevé de l'excitation tonique et de la conductance de la peau.

Vrana et al. [215] 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. 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.

Sinha et al. [197] ont utilisé des scénarios d'images développées pour induire 5 émotions (neutre, peur, joie, action, tristesse et colère). Ils se sont basés sur les données de cinq capteurs (le rythme cardiaque, conductance de la peau, température des doigts, pression sanguine, electro-oculogramme [3] et électromyographie faciale) de 27 hommes âgés entre 21 et 35 ans, et sur les fonctions d'analyse discriminante et la méthode d'ANOVA. Ils ont obtenu un taux de 99% de classification correcte. Cela indique que les modèles des réponses émotionnelles pour la peur et la colère sont différenciés correctement les unes des autres par rapport à l'état neutre.

Scheirer et al. [186] ont utilisé une interface d'un jeu d'ordinateur lent pour induire la frustration à 24 personnes (figure 1.27). 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.

Picard et al. [169] ont développé un système qui peut reconnaître avec précision huit émotions à partir des signaux physiologiques. 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 : aucune 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 réparties dans 20 jours. Les cinq signaux physiologiques enregistrés sont : électromyographie (EMG), la pression sanguine volumique (BVP), la fréquence cardiaque (HR), la conductance de la peau (RED), et la respiration. Les signaux physiologiques

3. Œil et Physiologie de la Vision

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 SFFS avec la projection de Fisher (SFFS-FP). Les résultats obtenus en utilisant les algorithmes SFFS-FP ont indiqué une précision de reconnaissance de 81% pour les huit catégories d'émotions qui est plus élevé par rapport à la reconnaissance des émotions avec la parole (60-70%) et proche de la précision des expressions faciales (80% à 98%)[169, 27].

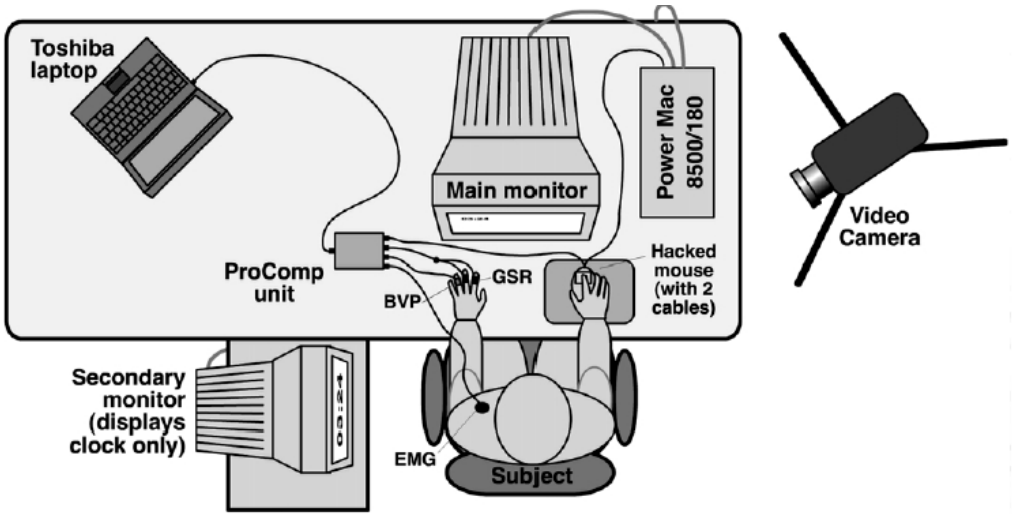


FIGURE 1.27 – L'environnement de l'acquisition utilisé dans [186]

Rani et al. [177] ont utilisé des participants jouant à des jeux vidéo et évaluent leur niveau de stress par l'auto-évaluation et la variation de la fréquence cardiaque et de l'intervalle Interbeat (IBI) [4]. Une analyse dans le domaine fréquentielle a été effectuée au signal IBI pour détecter si les participants éprouvaient un stress. L'architecture développée comprenait une collection de signaux de l'électrocardiographie (ECG) des participants et le calcul d'IBI. La transformée en ondelettes a été utilisée par la suite. Les auteurs ont calculé l'écart type de la bande fréquentielle sympathique [5] et de la bande parasympathique [6]. Ces écarts types sont l'entrée d'un système basé sur la logique floue, la sortie de ce système est la valeur de l'indice du stress. Si la valeur de l'indice de stress était plus grande qu'une valeur de seuil, un signal d'alarme sera envoyé à un robot qui prend des mesures pour aider les participants. Certains problèmes ont été rencontrés afin de simuler des situations stressantes qui suscitent des réponses appropriées : la variabilité du participant pendant la journée et la variabilité entre les participants [177].

4. Intervalle Interbeat est l'intervalle de temps entre les battements du cœur des mammifères [1].

5. La bande de basse fréquence correspond à l'activité du système nerveux sympathique, régissant la réponse de fuite ou de lutte (que ce soit une activité physique ou intellectuelle) par dilatation des bronches, accélération de l'activité cardiaque, dilatation des pupilles, augmentation de la sécrétion de la sueur et de la tension artérielle, mais diminution de l'activité digestive [6].

6. La bande de haute fréquence correspond à l'activité du système nerveux parasympathique, régissant le ralentissement général des fonctions de l'organisme afin de conserver l'énergie. Tout ce qui était augmenté, dilaté ou accéléré par le système nerveux sympathique est diminué, contracté et ralenti. Il est associé au neurotransmetteur acétylcholine. Dans la bande de fréquence au système nerveux parasympathique, la fréquence présentant le maximum de puissance spectrale est liée à la fréquence respiratoire. Si ces deux fréquences ne sont pas synchronisées, on observe alors un problème cardiorespiratoire [6].

Rani, Sarkar, Smith et Adams [176] ont étudié la reconnaissance de l'état affectif en se basant sur les mesures physiologiques de six participants. Trois problèmes à résoudre (une anagramme [7], problèmes mathématiques et des sons) de difficultés variables à travers six sessions expérimentales ont été utilisés pour induire l'anxiété aux participants. Ils ont mesuré l'ECG, BVP, EMG (du front à gauche et des muscles de la mâchoire), la température de la peau et le temps de transit du pouls TTP [8]. Le rapport d'auto-évaluation a également été utilisé pour corréler aux données physiologiques acquises avec les niveaux d'anxiété déclarés par les participants. Les signaux physiologiques ont été traités en utilisant la logique floue et un arbre de décision. Les données ont été divisées en deux ensembles, un pour l'apprentissage et l'autre pour le test du système. Les résultats indiquent qu'ils ont été en mesure de détecter de façon fiable l'anxiété chez les participants impliqués dans la résolution de problèmes pendant les sessions. Ils ont constaté que la décision du système basée sur l'arbre de classification était plus fiable que le système basé sur la logique floue.

Kulic et Croft [122] ont utilisé l'arousal et la valence pour l'évaluation de l'intention des participants. Ils ont mesuré la pression du volume sanguin, la conductance de la peau, la respiration et l'EMG (sourcils). Ils ont traité les signaux à l'aide d'un système d'inférence floue à cinq jeux de règles. La première série de règles évalue la relation entre le RED et l'arousal. Le deuxième ensemble de règles porte sur la relation entre l'EMG et la valence. Le troisième ensemble de règles entre l'activité cardiaque et valence/arousal. La quatrième série de règles relie les réponses vasomotrices à l'arousal. La cinquième série de règles de corrélation entre l'activité respiratoire avec l'état émotionnel [122]. Les auteurs ont utilisé pour l'induction des images et des tests psychophysiologiques développés par Lang et al. [128]. L'étude a été réalisée en utilisant trois participants. En moyenne, l'arousal a été correctement détectée 94%. Le changement de la valence a été correctement détecté en moyenne de 80%. Ils ont constaté que le taux de respiration n'a pas été utile pour déterminer les réactions de l'arousal des participants parce que le temps de réponse est trop long pour une application en temps réel. En outre, ils ont déterminé que les changements dans la fréquence cardiaque ont été difficiles à associer à un événement spécifique ou contexte [122].

Kim et al. [118] ont proposé un système de reconnaissance des émotions à partir des signaux physiologiques en utilisant trois capteurs pour la reconnaissance de quatre émotions (tristesse, colère, stress et la surprise), voir la figure 1.29. Les auteurs ont utilisé un protocole d'induction composé de trois aspects : l'audition, la vision et la cognition, la figure 1.28 montre un exemple pour induire la tristesse où les auteurs utilisent une lumière bleue associée à une musique triste en racontant une histoire triste par un acteur qui joue sur sa voix. Ils concluent que le protocole d'induction par la stimulation visuelle avec les images IAPS (*International affective picture system*) n'est pas suffisant. Le taux obtenu avec leur protocole est de 78.43% et 61.76% pour trois et

7. Jeu littéraire qui inverse ou permute les lettres d'un mot ou d'un groupe de mots pour en extraire un sens ou un mot nouveau

8. TTP est le temps écoulé entre l'éjection systolique et l'arrivée de l'onde de pouls en périphérie (ou temps de propagation de l'onde de pression) [3]. La mesure du TTP nécessite l'enregistrement de l'électrocardiogramme et de l'onde de pouls grâce à un capteur photopléthysmographique placé au niveau d'un doigt et est donc totalement non invasive [131].

quatre émotions respectivement, pour 50 personnes.

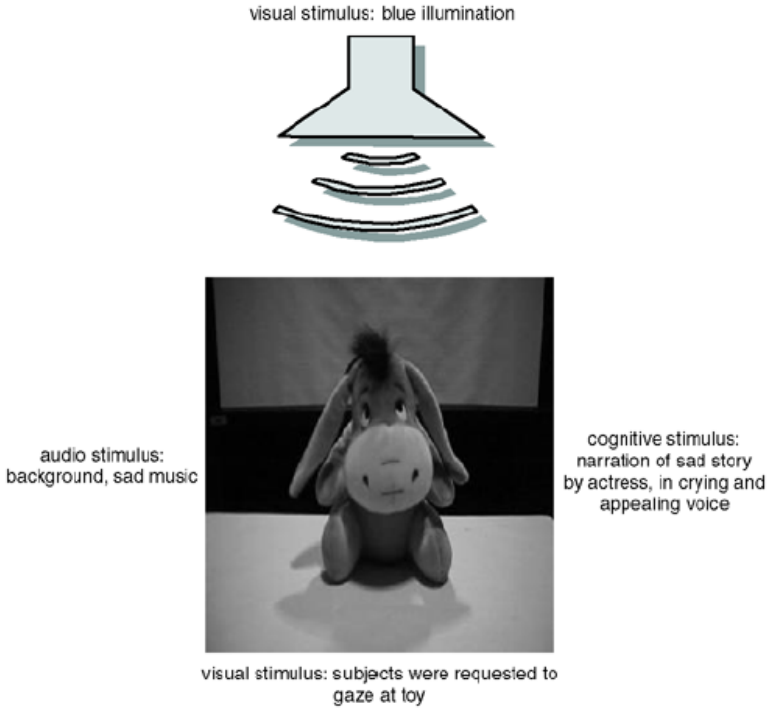


FIGURE 1.28 – Exemple d'induction de la tristesse avec le protocole de [118]

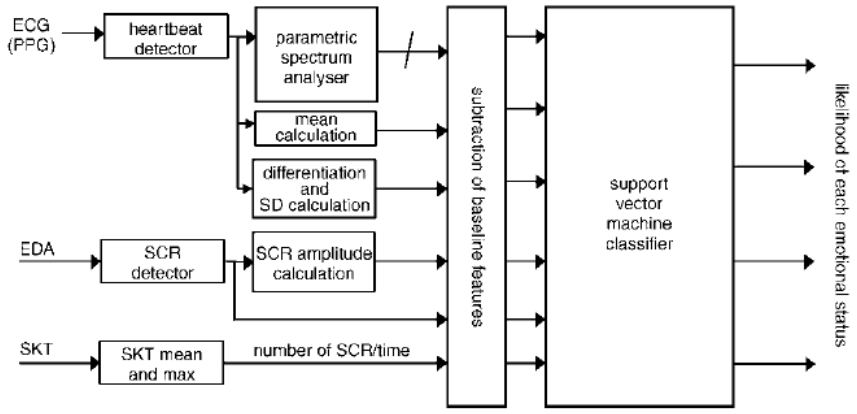


FIGURE 1.29 – Architecture du système de reconnaissance des émotions [118]

Chanel et al. [38] ont proposé un système pour la reconnaissance de trois états émotionnels positif, calme et négatif. Afin d'induire les émotions, ils ont demandé aux participants de vivre deux situations déjà passées dans leurs vies. Pour l'état neutre, ils demandent aux participants de se relaxer et de se calmer. Les auteurs ont utilisé quatre signaux : la pression sanguine, la fréquence cardiaque, la respiration et la réponse galvanique de la peau, et ils ont utilisé les signaux EEG. Afin d'enlever le bruit des signaux, ils ont utilisé un filtre moyen en changeant la fenêtre pour chaque signal (512 pour RED, 128 pour la pression sanguine et 256 pour la respiration). Leur résultat montre l'importance du signal EEG par rapport aux autres signaux.

Ils ont obtenu un taux de 67% pour trois classes (neutre, positive, négative) et un taux de 79% pour deux classes avec les SVM. Les résultats obtenus avec l'analyse discriminante linéaire ADL sont mauvais par rapport aux SVM. La méthode de sélection *Fast Correlation Based Filter* FCBF n'a pas amélioré les résultats dus à la non linéarité du problème traité. Les auteurs ont trouvé que les SVM favorisent la dimensionnalité élevée pour un meilleur taux de reconnaissance.

1.4.5 Conclusion

Les signaux physiologiques sont de bons marqueurs temporels. Leur principal avantage est que les participants ne peuvent pas manipuler consciemment les activités de leur système nerveux autonome [115, 169, 139].

Un des principaux inconvénients des signaux physiologiques qui peut affecter l'efficacité de la reconnaissance du système est le fait qu'ils sont intrusifs, par exemple, un capteur sur la tête et l'autre autour du torse [200]. Afin de remédier à ce problème, nous envisageons d'utiliser un système multimodal basé sur les expressions faciales et sur les signaux physiologiques.

1.5 Les systèmes multimodaux

1.5.1 Introduction

Bien que les systèmes de la reconnaissance automatique des émotions ont exploré l'utilisation soit des expressions faciales [28, 79, 147, 203, 225] ou de la parole [62, 153, 133] ou des signaux physiologiques [118, 122, 177, 169] pour détecter l'état émotionnel de l'homme, relativement peu d'efforts ont été concentrés sur la reconnaissance des émotions en utilisant deux ou plusieurs modalités [41, 194], notamment entre les signaux physiologiques et les expressions faciales. 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 bruité [160].

La fusion d'informations peut prendre différentes formes selon le moment ou elle est effectuée. La figure 1.30 montre trois possibilités de fusion à différents stades de la reconnaissance. Il est possible de fusionner les données directement après extraction des signaux, fusionner les attributs provenant des différentes modalités ou fusionner les informations durant la phase de décision.

1.5.2 Fusion de données

Il y a deux catégories de fusion de données brutes :

1. Niveau signal : la fusion consiste à combiner des signaux de même nature pour obtenir un signal de meilleure qualité. Les signaux doivent être bien recalés dans le temps ;

2. Niveau pixel : à ce niveau, il s'agit d'améliorer la connaissance sur chacun des pixels d'une image, par la combinaison de plusieurs images originales, ceci afin de pouvoir réaliser par la suite une meilleure segmentation.

Ce type de fusion n'est pas adapté aux données dont on dispose (un mélange entre signal et image).

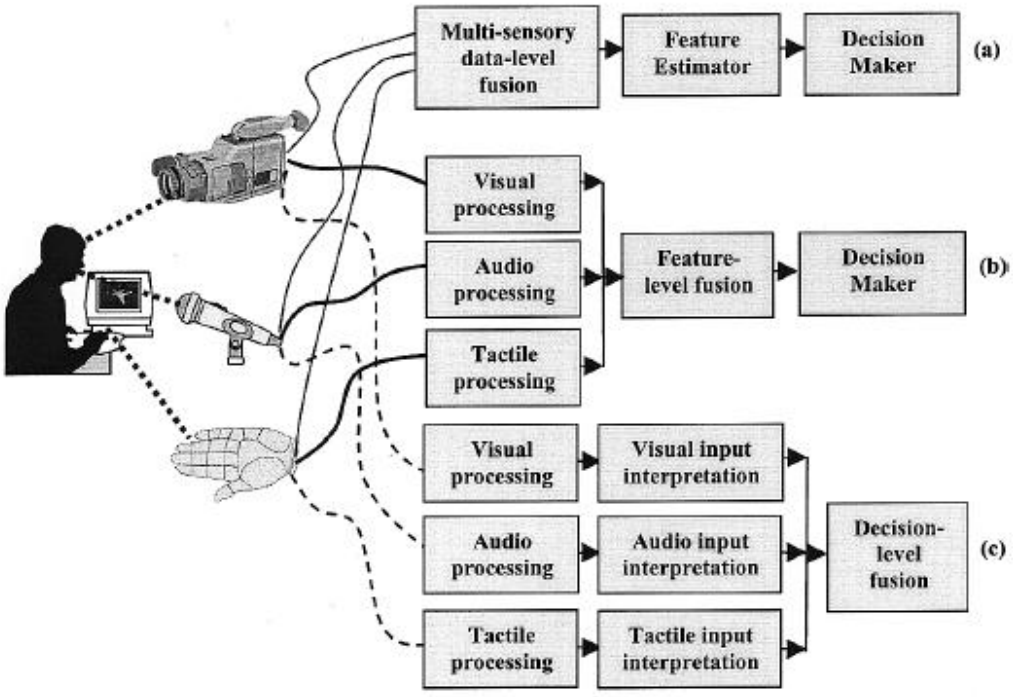


FIGURE 1.30 – Fusion de plusieurs modalités : a- Fusion des données, b- Fusion des caractéristiques, c- Fusion des décisions [160]

1.5.3 Fusion des caractéristiques

Huang et al. [102] ont proposé une simple concaténation des caractéristiques pour leur système audio-visuel. Ils ont remarqué que leur système basé sur la parole présente une confusion entre la tristesse et l'aversion (*dislike*), et d'un autre coté leur système basé sur l'image a une confusion entre la tristesse et la surprise et entre la colère et l'aversion (*dislike*). Les confusions rencontrées avec les deux modalités traitées séparément ont mené les auteurs à fusionner les deux modalités. Ils ont utilisé la méthode du plus proche voisin (*nearest-neighbor leave one out cross validation*) pour la classification. Avec cette fusion et ce classifieur, ils ont trouvé de bons résultats : audio 75%, la vidéo 69.4% et le bimodal 97.2%. Les auteurs ont utilisé la base de *Spanish and Sinhala* qui contient 36 clips.

La concaténation des caractéristiques est une méthode simple pour la fusion. La chaîne d'analyse des caractéristiques peut être poursuivie par une phase de réduction de dimension supposée de grande dimension et/ou redondante. Théoriquement, l'augmentation du nombre des descrip-

teurs pourra améliorer les performances du système. Néanmoins, en pratique, l'utilisation d'un grand nombre de descripteurs, en plus du problème de complexité engendré par la dimension élevée de l'espace de représentation des données, peut en fait aboutir à une baisse des performances [103].

Les méthodes de réduction de dimensions des données utilisées peuvent être divisées en deux classes :

- ❒ Les méthodes de sélection ;
- ❒ Les méthodes de transformation ou d'extraction.

1.5.3.1 Méthodes de sélection

Les méthodes de sélection visent à trouver d'une manière optimale, un sous-ensemble de caractéristiques à partir d'un ensemble initial. Formellement, supposons que nous disposons d'un ensemble F de n éléments, dont nous voulons sélectionner un sous-ensemble S de m caractéristiques. Le nombre total des sous ensembles possibles est très élevé, le traitement de chaque sous ensemble est généralement impossible. L'objectif est d'obtenir un sous-ensemble S qui conserve autant de renseignements que possible en F.

Les méthodes de sélection diffèrent principalement sur deux aspects. La première est la stratégie de recherche qui précise la façon dont le sous-ensemble S est construit, tandis que la seconde est la mesure utilisée pour évaluer la qualité de chaque caractéristique, ou même le sous-ensemble S.

A. Backward selection

Les méthodes de la rétro-sélection *backward selection* utilisent à la première itération toutes les variables, retirent celle qui est la moins discriminante en termes du critère donné. Par la suite, on réitère sur le reste des variables et ainsi de suite, on élimine la variable la plus faible. La recherche s'arrête lorsque toutes les variables dans le modèle sont significatives. Cette approche a pour inconvénient de ne pas permettre de réévaluer la variable éliminée. Elle est efficace pour un nombre important de variables [103].

B. Forward selection FFS

À l'inverse des premières, cette approche commence par un ensemble vide de variables, puis à chaque itération, la variable qui donne un taux de discrimination maximal couplée avec celles déjà incluses est sélectionnée. Son inconvénient est qu'elle ne permet pas d'éliminer une variable qui devient inefficace après l'ajout d'autres variables. Par contre, elle est efficace pour un petit nombre de paramètres [103].

C. Floating forward and backward selection FBS

Pour la méthode FFS, au début, le sous-ensemble optimal est initialisé à l'ensemble vide et à chaque itération, une nouvelle variable est ajoutée [174]. Après chaque insertion, on vérifie si on peut enlever des variables une après l'autre tout en conservant un bon taux de discrimination.

Cette approche est une combinaison des deux premières. Pour la méthode FBS, elle est similaire à la méthode FFS sauf qu'elle débute avec l'ensemble complet de variables.

Une fois la méthode de recherche appropriée à l'application visée est choisie, on doit exécuter un algorithme de sélection des variables pour extraire le meilleur sous-ensemble. Deux catégories d'approches existent : les algorithmes dits filtres [57, 230] et les algorithmes *wrappers* [120].

A. Approches à base de filtres

Ces algorithmes mesurent la pertinence du sous-ensemble à sélectionner indépendamment du classifieur ou un critère donné. Il s'agit d'ordonner les différentes variables selon la pertinence du score qui leur a été attribué durant l'évaluation des sous-ensembles [114, 100]. Plusieurs algorithmes sont proposés, nous avons choisi d'étudier l'information mutuelle (MIFS) [25, 166, 85, 125, 124, 53].

B. Approches à base de wrappers

Au contraire de la première approche, les algorithmes wrappers définissent la pertinence des attributs par l'intermédiaire d'une prédiction de la performance du système final (classification par exemple) comme ReliefF [114, 180], Fisher ou analyse discriminantes [184, 242], les arbres de décision [219] et les algorithmes génétiques [165, 227].

Quelques exemples sur la sélection des caractéristiques pour la reconnaissance des émotions

Kim et André ont proposé un système bimodal basé sur la parole et les signaux physiologiques [117] en supposant que la combinaison de ces deux modalités donne de meilleures performances au système de reconnaissance d'émotion. Ils se sont intéressés à quatre émotions (positive élevée, positive basse, négative élevée et négative basse). Ils précisent que la réduction de la dimensionnalité permet la diminution du temps de calcul ainsi que la suppression du bruit améliore la séparation des classes. Ils ont obtenu une amélioration de 30% en utilisant une rétro-sélection séquentielle SBS. Le succès de la méthode dépend du classifieur utilisé. Parmi les valeurs sélectionnées par cette méthode : l'entropie du spectre du BVP (pression sanguine), le nombre d'occurrence du maximum de la conductance de la peau et le signal EMG et la valeur moyenne du MFCC (*mel frequency cepstral coefficient*) pour la parole. Ils ne peuvent pas généraliser ce qu'ils ont trouvé à cause du petit ensemble d'apprentissage. Ils ont testé plusieurs classifieurs : k plus proches voisins, perceptron multicouche et l'analyse discriminante linéaire, et ont opté pour l'analyse discriminante linéaire ADL qui a donné des bons résultats.

Busso et al. [34] ont utilisé la rétro-sélection séquentielle pour la fusion des caractéristiques dans leur système audio-visuel. Les auteurs ont obtenu un taux de 89.1% pour la classification de quatre émotions (colère, tristesse, joie et neutralité). Ils ont conclu leur travail par une très bonne reconnaissance de la colère et la neutralité en utilisant un système bimodal par rapport à l'utilisation des expressions faciales.

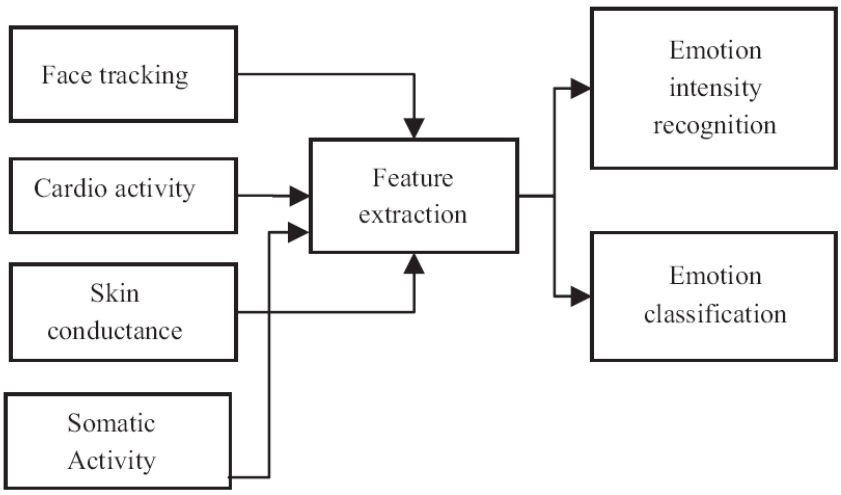


FIGURE 1.31 – L'architecture du système de reconnaissance des émotions [20]

La figure 1.31 montre l'architecture du système de reconnaissance des émotions en fusionnant les caractéristiques de deux modalités : les signaux physiologiques et les expressions faciales utilisées par Bailonson et al. [20]. Les émotions considérées dans ce travail sont l'amusement et la tristesse. Les auteurs ont opté pour des films afin d'induire les émotions. Ils ont utilisé plusieurs paramètres physiologiques en se basant sur la fréquence cardiaque, la pression sanguine, la conductance de la peau et la température du doigt. Concernant les expressions faciales, ils ont utilisé les coordonnées de 22 points faciaux. Les auteurs ont utilisé la méthode du CHI2 afin d'évaluer la contribution de chaque caractéristique et le classifieur utilisé est le perceptron multicouche. Les auteurs ont obtenu environ 99% pour la base individuelle. Pour la base globale, ils ont focalisé leur travail sur l'effet du sexe vis-à-vis de l'expression de l'émotion, ils ont trouvé que les femmes expriment l'amusement mieux que les hommes avec un taux de 95% par rapport à 93%. Par contre, les hommes expriment la tristesse mieux que les femmes avec taux de 85% par rapport à 84% avec leur système bimodal.

1.5.3.2 Transformations des caractéristiques

Les méthodes de transformation des caractéristiques, ou bien les méthodes d'extraction des caractéristiques [140], produisent un ensemble de nouvelles caractéristiques basé sur l'ensemble initial. Cela signifie que toutes les caractéristiques originales sont nécessaires et il n'y a aucune réduction dans les conditions de collecte de données. Les nouvelles caractéristiques sont obtenues à la suite de l'application d'une transformation linéaire ou non linéaire sur le vecteur initial des caractéristiques. Cela conduit à un changement dans la représentation de la donnée elle-même, dans le sens que l'on peut mieux visualiser et comprendre. La transformation des données peut également être supervisée ou non supervisée.

Plusieurs techniques sont proposées dans ce cadre et sont subdivisées selon deux aspects : linéaire et non-linéaire. Parmi les méthodes linéaires [31, 80], on trouve :

- ❒ L'analyse en composantes principales (ACP) ;
- ❒ L'analyse linéaire discriminante (ALD) ;
- ❒ L'analyse factorielle des correspondances (AFC).

Pour les méthodes non-linéaires, on trouve :

- ❒ L'analyse en composantes curvilinéaires (ACC) [63, 64] ;
- ❒ *Kernel PCA* [188] ;
- ❒ *Data-driven High Dimensional Scaling* (DD-HDS) [134].

Nous présentons dans ce qui suit les deux transformées, les plus utilisés pour réduire la dimensionnalité des données.

Analyse en composantes principales (ACP)

C'est une transformée non-supervisée, aucune information sur les classes n'est présente dans les données utilisées, même si cette information est disponible. L'ACP vise à trouver les directions de l'espace dans lequel la variance des données est plus grande. Il effectue une rotation telle que les nouveaux axes de coordonnées sont ces directions. Aussi connu sous le nom de la transformée de Karhunen-Loeve depuis 1901[164]. L'ACP est calculée à partir de la matrice de covariance des données et le tri du vecteur des valeurs propres en ordre décroissant. En suite, toutes les données sont projetées sur cette nouvelle base. Pour réduire la dimensionnalité des données obtenues, les paramètres qui correspondent à une assez grande valeur propre sont conservés. Cela se fait par la fixation d'un seuil à la valeur propre. L'ACP est la meilleure transformée linéaire, dans le sens de l'erreur quadratique moyenne, pour la réduction des dimensions des données, puis sa reconstruction [95].

Analyse discriminante linéaire (ADL)

C'est une transformée supervisée, car elle utilise des étiquettes de la classe visant à trouver la transformation qui sépare le mieux les classes. L'ADL est également une technique très ancienne et populaire, dérivée de la discrimination linéaire de Fisher présentée en 1936 [84], pour des problèmes à deux classes et étendue à des problèmes multiclasses par Rao [178].

Le critère qui est maximisé par l'ADL est le rapport entre la dispersion inter-classe et extra-classe. Cela requiert une simple matrice arithmétique. Le résultat de la transformation est une réduction de la représentation des dimensions, tout comme pour l'ACP, mais où les classes devraient idéalement être séparables et compactes.Toutefois, il existe plusieurs limites de l'ADL, comme l'hypothèse que les distributions de probabilités des classes sont normales et les classes sont homoscedastiques, c'est-à-dire ils ont la même covariance. Heteroscedatic ADL [123][67] a été développée pour traiter spécifiquement le problème de la distribution des classes avec différentes covariances.

1.5.3.2.1 Exemples sur la transformation des caractéristiques pour la reconnaissance des émotions

Busso et al. [34] ont utilisé l'ACP pour réduire l'information visuelle de leur système à 10 dimensions. Les résultats obtenus pour la classification des expressions faciales actées sont de

l'ordre de 85%.

Chuanget al. [47] ont conçu un système multi-modal pour la reconnaissance des émotions en se basant sur la parole et sur la saisie du texte. L'information textuelle est un autre moyen de communication important et peut être récupérée à partir de nombreuses sources telles que livres, journaux, pages web, e-mails, etc. Avec les techniques de traitement du langage naturel, les émotions peuvent être extraites d'un texte en analysant la ponctuation, les mots clés émotionnels, la structure syntaxique, l'information sémantique, etc., les auteurs ont développé un réseau sémantique pour effectuer la reconnaissance des émotions d'un contenu textuel. Leur système traite six émotions : joie, tristesse, colère, peur, surprise et dégoût. Pour évaluer l'information acoustique, ils ont extrait 33 paramètres et grâce a l'ACP ils sont passés à 14 composantes qui sont classées par la suite avec les SVM. La décision du système de reconnaissance des émotions est basée sur la décision des deux modalités.

1.5.3.3 Synthèse sur la fusion au niveau des caractéristiques

Une des limites des algorithmes de sélection cités précédemment est le traitement individuel de chaque paramètre sans tenir compte de l'interaction entre eux. Une solution alternative est donnée par l'algorithme MIFS (*Mutual Information Feature Selection*) proposé par Battiti [25]. L'algorithme MIFS consiste à chercher le sous-ensemble S, de dimension d inférieure à celle de l'ensemble total des descripteurs initiaux, qui maximise l'information mutuelle IM (C, S) entre cet ensemble et la variable C des classes d'appartenance.

L'utilisation de l'information mutuelle dans le cas où la dimension de l'espace des attributs est assez importante $D >>$ présente le problème du temps de calcul. Dans notre travail nous avons un $D = 51$ qui ne pose pas ce problème. Donc notre choix s'est porté sur l'information mutuelle pour la sélection des caractéristiques et sur l'ACP pour leur transformation afin de voir l'effet de chacune sur la fusion.

1.5.4 Fusion des décisions

Nous présentons ici quelques cadres théoriques de fusion d'informations de haut niveau. Nous considérons le problème de la fusion de m sources S_j afin de déterminer une des n classes C_i possibles.

1.5.4.1 Principe du vote

Le principe du vote est la méthode de fusion d'informations la plus simple à mettre en œuvre. Plus qu'une approche de fusion, le principe du vote est une méthode de combinaison particulièrement adaptée aux décisions de type symbolique.

Dans [234], Zeng et al. ont utilisé le principe de vote pour combiner les résultats de la classification des mouvements du front du visage, le pitch et l'énergie de la prosodie. Cette fusion

a permis d'améliorer le taux de reconnaissance de 7.5% par rapport à l'utilisation de chaque modalité à part. Pour mémoire, les auteurs ont utilisé SNoW (*Sparse Network of Window*) pour la classification.

1.5.4.2 Les règles

Dans [196], De Silva et al. ont proposé un système audio-visuel de la reconnaissance des émotions basé sur la fusion des décisions avec des règles. Du coté audio, ils ont utilisé les caractéristiques de la prosodie et du coté vidéo ils ont utilisé le maximum des distances de six points caractéristiques. Une approche similaire a également été présentée par Chen et al. [41], dans laquelle la modalité dominante a été utilisée pour résoudre les divergences entre les sorties des systèmes uni-modaux selon l'expérience subjective effectuée [195]. Dans les deux études, ils ont conclu que la performance du système augmente lorsque les deux modalités ont été utilisées ensemble.

1.5.4.3 Méthodes empiriques

Yoshitomi et al. ont proposé un système multimodal en considérant la parole, l'information visuelle et aussi la distribution thermique acquises par une caméra infrarouge [229]. L'utilisation des images infrarouges surmonte le problème des conditions d'éclairage qui est posé par l'acquisition des expressions faciales avec des caméras classiques. Ils ont utilisé une base de données enregistrée à partir d'une locutrice qui lit en actant les cinq émotions. Ils ont intégré ces trois modalités avec la fusion de décisions en utilisant des poids déterminés d'une façon empirique. Les performances du système ont été meilleures en considérant les trois modalités.

1.5.4.4 Distance euclidienne

La figure 1.32 montre le diagramme de reconnaissance des émotions en fusionnant la décision de deux modalités : les signaux physiologiques et les expressions faciales. Les émotions considérées dans ce travail sont la peur, l'amour, la joie et la surprise [46]. Les auteurs ont opté pour des films afin d'induire les émotions. Ils ont utilisé la conductance de la peau, la température du doigt et la fréquence cardiaque. Du coté image, ils ont utilisé 12 distances faciales. Pour les deux modalités, le classifieur utilisé est le LVQNNs (*Learning Vector quantization neural net work*). Chaque vecteur d'entrée des caractéristiques est comparé en utilisant la distance euclidienne dans le LVQNNs. Pour une émotion inconnue, les distances entre les différentes classes des deux LVQNNs sont additionnées. L'émotion est classée selon la catégorie dont la sommation est le minimum.

Notons que les auteurs ont travaillé avec 4 personnes et ils ont utilisé les échantillons impairs pour l'apprentissage et les échantillons pairs pour le test, ils ont trouvé un taux de 90% pour les expressions faciales, 88.33% pour les signaux physiologiques et 95% avec la fusion de décisions des deux modalités.

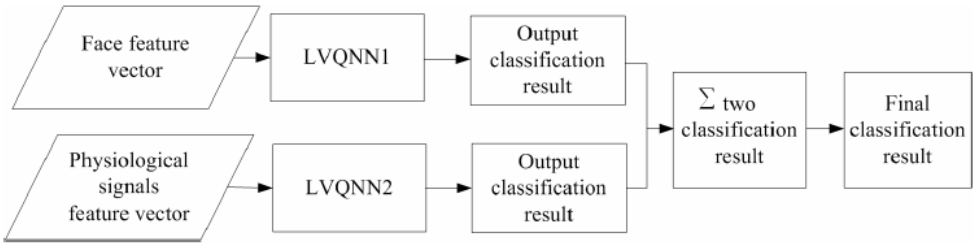


FIGURE 1.32 – Le diagramme de la reconnaissance des émotions en considérant les signaux physiologiques et les expressions faciales [46]

1.5.4.5 Modèle graphique de probabilité

A. Probabilité postérieure

Kim et André ont proposé un système bimodal basé sur la parole et sur les signaux physiologiques [117]. Ils ont fusionné les décisions de chaque modalité en utilisant la probabilité postérieure sachant qu'ils ont utilisé l'analyse discriminante linéaire LDA pour la classification dans chaque modalité. Les auteurs ont conclu leur article par le problème posé par la probabilité postérieure quand on fusionne deux modalités avec des précisions de haute disparité. Ils favorisent leurs méthodes de sélection des caractéristiques par rapport à la méthode de fusion de données.

Également, Busso et al. [34] ont utilisé la probabilité postérieure avec plusieurs critères pour fusionner les deux décisions issues des deux SVC (*separator vaste clustering*). Le meilleur taux obtenu est 89% pour quatre émotions (colère, tristesse, joie et neutralité). Les auteurs concluent leur travail par une très bonne reconnaissance pour la joie et la tristesse avec leur système bimodal à base de fusion de décision. Donc la meilleure fusion dépend de l'application. Notons que les résultats sont validés sur un seul sujet.

B. Réseau Bayésien

Dans [189], Sebe et al. ont proposé une topologie du réseau Bayésien permettant de reconnaître les émotions à partir d'expressions faciales et des informations audio présentées dans la figure 1.33. 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 est influencée par des expressions du visage reconnues, par les expressions vocales reconnues et par le contexte dans lequel le système fonctionne. La reconnaissance est également affectée par une variable qui indique si la personne parle ou pas. En utilisant cette topologie, la reconnaissance de l'expression émotionnelle humaine peut être effectuée même lorsque certains éléments de l'information sont manquants, par exemple, lorsque le son est trop bruyant ou l'image perd le visage [189]. Les résultats expérimentaux des auteurs montrent que la moyenne du taux de reconnaissance de la personne dépendante est très bien améliorée lorsque les deux informations visuelle et audio sont utilisées à la fois dans la classification.

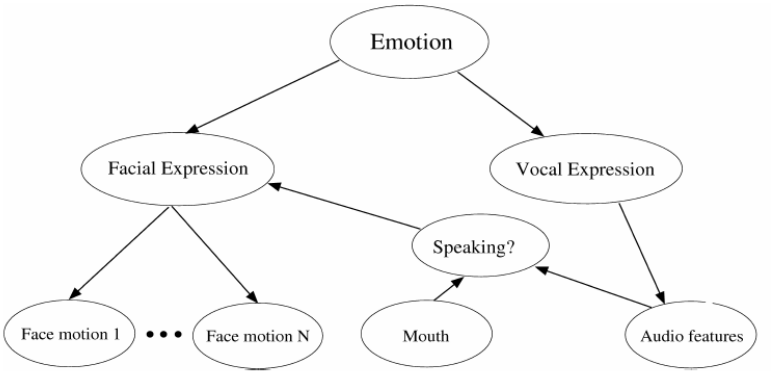


FIGURE 1.33 – La topologie d'un réseau Bayésien pour la reconnaissance bimodale des émotions [189]

Datcu et al. [59] utilise un réseau Bayésien dynamique dont le modèle de fusion vise à déterminer l'émotion la plus probable du sujet en considérant celles déterminées dans les images précédentes. Une fenêtre de données contient la décision actuelle et les n précédentes pour l'analyse. La figure 1.34 montre l'architecture du réseau Bayésien dynamique RBD utilisée pour la fusion des données. Les auteurs ont utilisé les SVM pour la classification uni-modale.

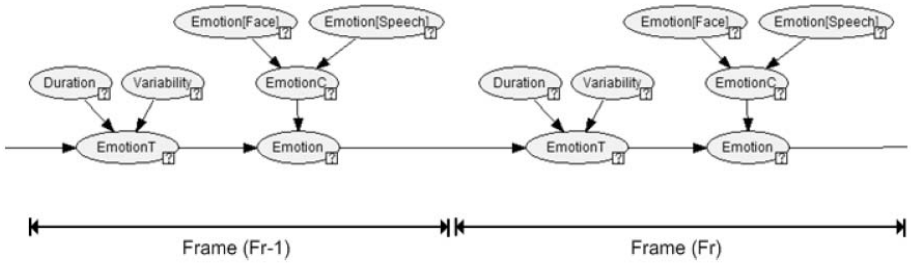


FIGURE 1.34 – Le modèle du réseau Bayésien dynamique pour la reconnaissance multimodale des émotions

1.5.5 Conclusion

La problématique de la fusion de données a des contours scientifiques flous par essence. En effet, la fusion est toujours associée à un contexte applicatif particulier pour atteindre un objectif précis. Le cadre de la fusion d'informations se différencie suivant les domaines d'application, par exemple Busso et al. [34] ont obtenu un très bon taux avec la colère et la neutralité en utilisant la fusion des caractéristiques et un très bon taux pour la joie et la tristesse en utilisant la fusion des décisions.

Dans le cadre de cette thèse, nous proposons de comparer les approches basées sur une fusion de caractéristiques et les approches basées sur la fusion de décisions à des fins de choix. Le tableau 1.7 résume quelques travaux pour les systèmes multimodaux (Audio-visuel).

1.6 Conclusion

Ce chapitre portait sur les défis rencontrés lors de la reconnaissance des émotions. Nous

avons présenté les émotions d'une manière générale et par la suite des expressions faciales et des signaux physiologiques en terminant par les systèmes multimodaux.

Le chapitre suivant se focalise sur une étude détaillée de l'analyse des expressions faciales pour la reconnaissance des émotions de notre système.

Référence	Caractéristiques	Fusion	Classifieur	Exp	classe	Taux
Go et al. 03 [92]	Visages propres, MFCC	D	*LDA*	P	6	95 %
Busso et al. 04 [34]	102 marqueurs, prosodie	F,D	*SVM*	P	4	89 %
Zeng et al. 04 [237]	Mouvement des unités, prosodie	D	*SNoW*	P	11	89 %
Song et al. 04 [198]	54 FAPS, prosodie	M	*THMM*	P	7	84.7 %
Zeng et al. 05 [239]	Mouvement des unités, prosodie	F	*Fisher-boosting*	P	4	84 %
Fragopanagos et al. 05 [86]	17 FAPS, prosodie	M	*ANNA*	S	4	44-71 %
Hoch et al. 05 [99]	Descripteurs de Gabbor, prosodie	D	*SVM*	P	3	90.7 %
Wang & Guan 05 [218]	Ondelettes de Gabor, prosodie, MFCC, *formants*	D	*FLDA*	P	6	82.14 %
Zeng et al. 05 [238]	Mouvement des unités, prosodie	M	*MFHMM*	P	11	80.61 %
Caridakis et al.06 [37]	Points faciaux, prosodie	M	*RNN*	S	4	79 %
Pal et al. 06 [155]	Niveau de gris vertical, F0-F3	D	*les règles, K-means*	S	5	75.2 %
Sebe et al. 06 [190]	Mouvement de 12 unités, prosodie	M	*BN*	P	11	90 %
Zeng et al. 06 [233]	Mouvement de 12 unités, prosodie	M	*MFHMM*	P	11	83 %
Karpouzis et al. 07 [112]	19 FPS, prosodie	M	*RNN*	S	4	82 %
Zeng et al. 07 [234]	Texture avec LLP, prosodie	D	*Adaboost + MHMM*	S	2	89 %
Zeng et al. 07 [236]	Mouvement des unités, prosodie, *formants*	D	*HMM*	P	11	72.42 %
Petridis et al. 08 [167]	Points faciaux, caractéristiques du spectre	F,D	*Adaboost + NN*	S	2	86.9 %

Exp : expression ;
S : expression **s**pontanée ;
P : expression actée ;
Im : reconnaissance basée sur l'image ;
Vi : reconnaissance basée sur la vidéo ;
F : fusion des caractéristiques ;
D : fusion des décisions ;
M : niveau de modalité.

TABLE 1.7 – Comparaison des systèmes multimodaux (audio-visuel) [235]

Chapitre 2

Analyse des expressions faciales

2.1 Introduction

Du point de vue physiologique, une expression faciale résulte de la déformation des traits du visage provoquée par une émotion. L'essentiel de l'information d'une expression est contenu dans la déformation des traits permanents principaux du visage, à savoir les yeux, le nez, les sourcils et la bouche. Cette constatation a été validée d'un point de vue psychologique par différents travaux [71, 24]. Ces derniers ont démontré par des expériences psychologiques que l'information sur une expression faciale est contenue non pas dans un trait particulier mais dans la combinaison de ces derniers.

Notre système de reconnaissance d'émotions à partir des expressions faciales, illustré dans la figure 2.1, analyse le mouvement de différentes caractéristiques faciales durant une séquence vidéo pour déterminer l'émotion d'une personne.
L'analyse automatique des expressions faciales s'effectue généralement selon les étapes suivantes :

1. Localisation des caractéristiques faciales dans la première image de la séquence vidéo ;
2. Suivi des caractéristiques faciales dans le reste des images de la séquence ;
3. Codage et classification des expressions faciales pour la reconnaissance des émotions.

La première partie du chapitre fournit une vue détaillée de notre approche de l'extraction des caractéristiques faciales. La reconnaissance des expressions est expliquée dans la deuxième partie.

2.2 Extraction des caractéristiques faciales

L'extraction des caractéristiques du visage est une étape clé dans notre système. Elle consiste à détecter des points d'intérêt qui décrivent les expressions faciales. Notons qu'il y a trois paramètres à considérer comme hypothèses nécessaires pour notre application :

1- Le visage du sujet doit être en face de la caméra ;

2- Le sujet commence toujours par une expression faciale neutre pour la phase d'initialisation ;

3- Les conditions d'éclairage sont stables.

La variation de l'un de ces paramètres peut influencer les résultats de notre système.

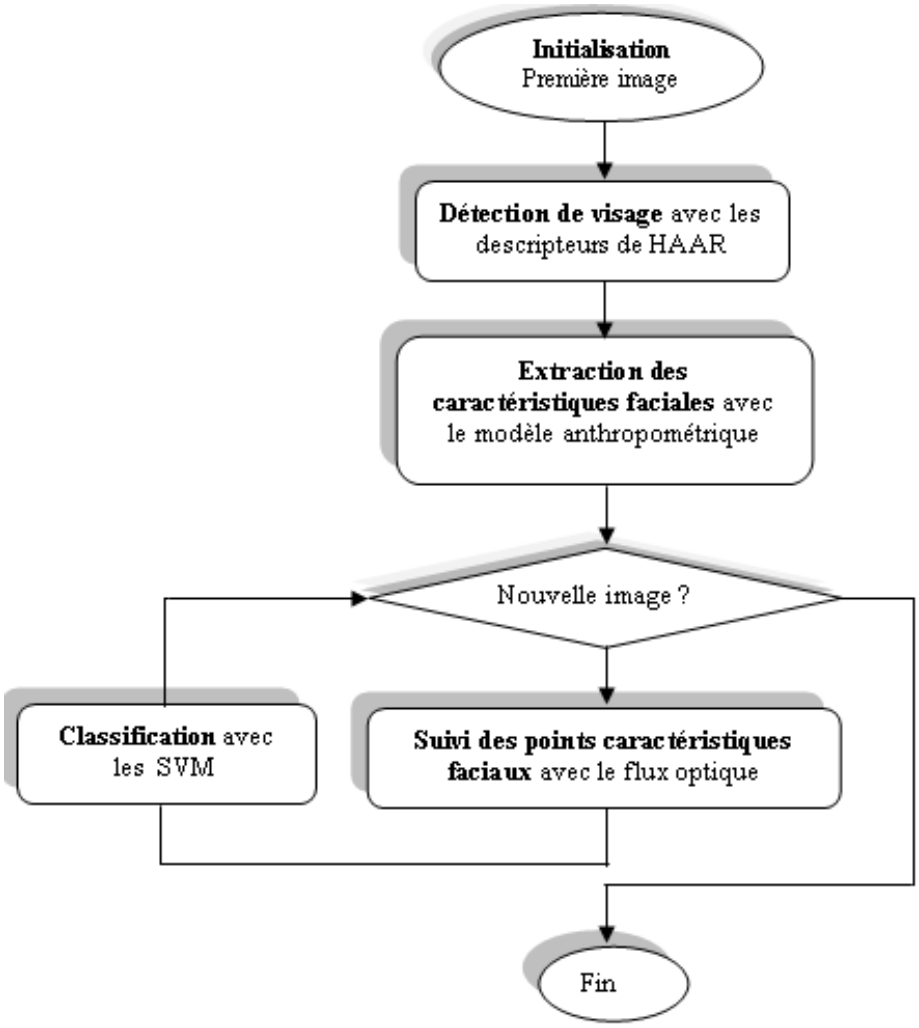


FIGURE 2.1 – Organigramme d'un système de reconnaissance des expressions faciales

2.2.1 Détection de visage

La détection de visage est la première étape de notre système de reconnaissance d'expressions faciales qui consiste à délimiter la zone d'intérêt par un rectangle. Pour réaliser cette tâche, nous avons utilisé un détecteur de visage rapide et robuste implémenté dans la bibliothèque OpenCV [33]. Il a été initialement élaboré par P. Viola et M. Jones [214] puis amélioré et implémenté par R. Lienhart [136]. Ce détecteur est basé sur les descripteurs de HAAR et des classifieurs en cascade.

2.2.1.1 Les descripteurs de HAAR

Les valeurs d'un pixel ne nous informent que sur la luminance et la couleur d'un point donné. Il est donc plus judicieux de trouver des détecteurs fondés sur des caractéristiques plus globales de l'objet. C'est le cas des descripteurs de Haar qui sont des fonctions permettant de connaître la différence de contraste entre plusieurs régions rectangulaires contiguës dans une image. Ils codent ainsi les contrastes existants dans un visage et les relations spatiales (figure 2.2).

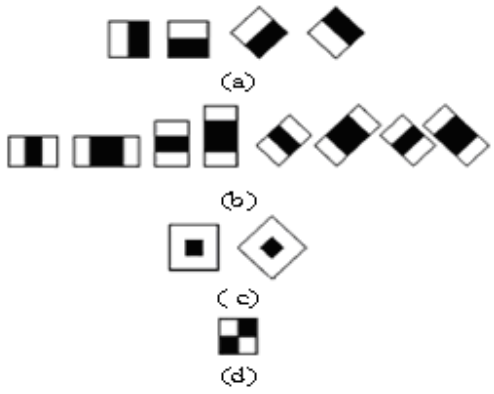


FIGURE 2.2 – Les descripteurs de HAAR : a- Les descripteurs de contour, b- Les descripteurs de ligne, c- Les descripteurs du centre, d- Les descripteurs de ligne diagonale.

En effet, ces descripteurs permettent de calculer la différence entre la somme des pixels dans les zones blanches et la somme des pixels des zones noires.

$$f_i = Sum(r_i, blanche) - Sum(r_i, noire) \tag{2.1}$$

Tel que : r_i est l'intensité des pixels d'une région rectangulaire adjacente de l'image.
Un descripteur de HAAR est caractérisé par :

- ❒ Le nombre de rectangles ;
- ❒ La position (le sommet supérieur gauche) (x,y) de chaque rectangle ;
- ❒ La largeur w et la hauteur h de chaque rectangle ;
- ❒ Les poids positifs ou négatifs de chaque rectangle.

La figure 2.3 montre le principe de calcul de la valeur d'un descripteur.

La taille de la fenêtre de détection est égale à la taille des images utilisées comme exemple d'apprentissage dans le processus de création du classifieur.

Par la suite, une cascade de classifieurs est utilisée pour classer une région comme un visage ou non visage selon la valeur de son descripteur.

2.2.1.2 Cascade de classifieur

Chaque classifieur est en réalité une « cascade de classifieurs dopés basés sur les descripteurs

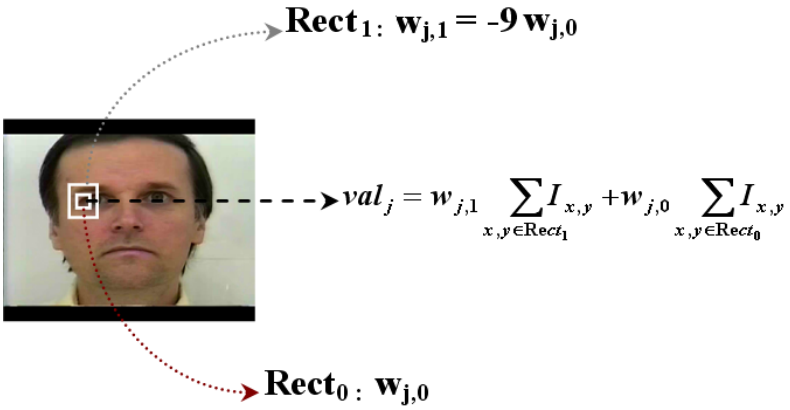


FIGURE 2.3 – La forme et la localisation d'un descripteur j dans une fenêtre de recherche

de Haar » (*cascade of boosted classifiers working with Haar-like features*). Le terme « cascade » signifie que le classifieur est constitué de plusieurs classifieurs simples (appelés étages) qui sont appliqués les uns à la suite des autres sur une région d'intérêt d'une image (figure 2.4).
Le terme « dopé » (*boosted*) consiste à combiner le résultat obtenu par plusieurs classifieurs "faibles" pour obtenir un classifieur plus efficace. L'algorithme de boosting le plus populaire est l'algorithme d'AdaBoost, mis au point en 1997 par Freund et Shapire [185].
Le fonctionnement général d'un algorithme de boosting est itératif, un ensemble de N classifieurs complémentaires est constitué à partir d'un ensemble d'exemples d'apprentissage distribués selon une loi de probabilité. Le classifieur final revient à une combinaison linéaire pondérée des classifieurs faibles déterminés à chaque itération (figure 2.4).

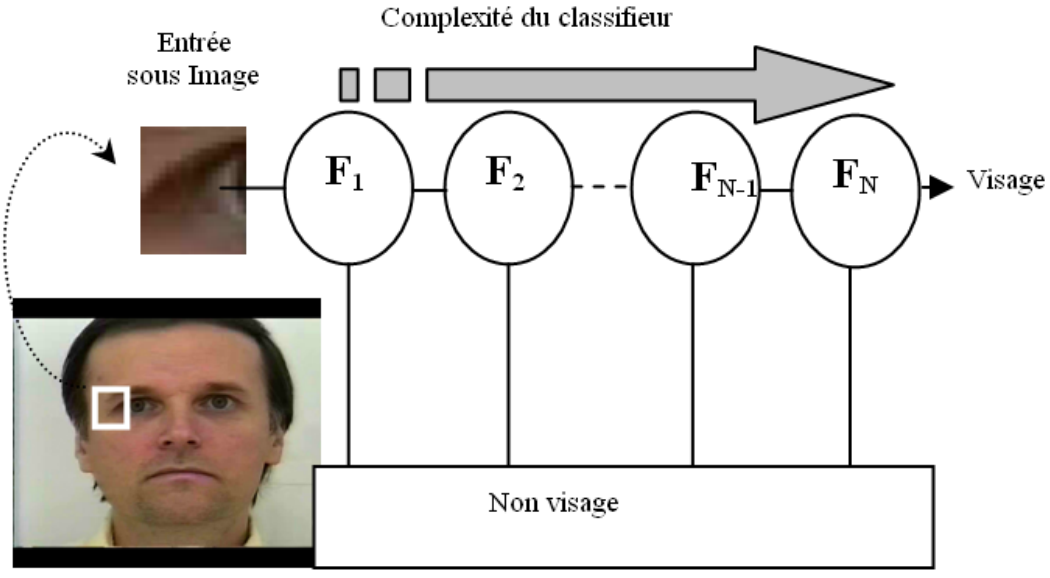


FIGURE 2.4 – Cascade de classifieurs

L'implémentation du détecteur de Viola et Jones sur un échantillon de la base de données de FEEDTUM [216] et en temps réel (en utilisant une simple webcam) montre que ce détecteur de visage est rapide et robuste vis-à-vis des variations de lumière (figure 2.5).
La figure 2.6 illustre les limites de ce détecteur qui répond mal quand il s'agit d'un visage trop

incliné. Cependant, cet inconvénient est sans importance en considérant l'hypothèse citée au début de la section, où la personne doit être face à la caméra pour la détection de visage dans la première image.

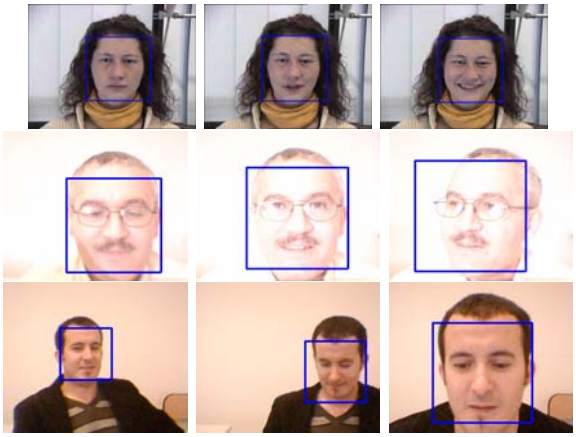


FIGURE 2.5 – La détection de visage avec le détecteur de Viola-Jones

FIGURE 2.6 – Les limites du détecteur de visage

2.2.2 La localisation des points caractéristiques faciaux

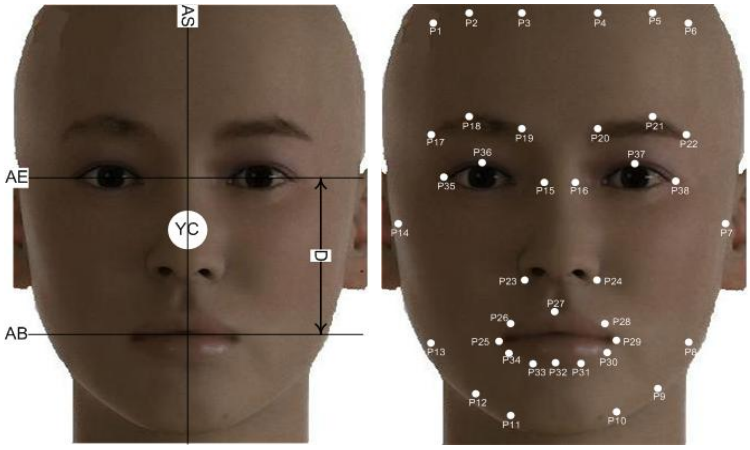
Après la détection de visage avec le détecteur de Viola-Jones, nous avons localisé les points caractéristiques faciaux dans le cadre qui englobe le visage.
Il existe de nombreuses techniques différentes consacrées à l'extraction des caractéristiques faciales [231]. Notre choix s'est porté sur une méthode rapide et adaptée au traitement en temps réel qui se base sur les points d'intérêt. Pour réaliser cette étape, nous avons procédé à la localisation des axes principaux (figure 2.7-a) : l'axe des yeux, l'axe de la bouche et l'axe vertical du nez [9, 17, 10].

1. La localisation de l'axe des yeux est basée sur le calcul de gradient dans le cadre qui délimite le visage. La ligne qui contient les yeux correspond à la ligne qui a le maximum de gradient. Cette ligne correspond à plusieurs transitions : peau vers l'œil, blanc de l'œil vers l'iris, l'iris vers la pupille et la même chose de l'autre côté.

2. La localisation de l'axe de la bouche est basée sur le même principe que l'étape précédente, où, l'axe de la bouche correspond à la ligne qui a le plus fort gradient situé dans la partie en dessous de l'axe des yeux.

3. L'axe médian est déterminé par l'axe qui divise le rectangle englobant le visage en deux parties égales.

Par la suite, nous avons localisé les points caractéristiques en appliquant notre modèle anthropométrique (figure 2.7-b). Ce dernier suppose que les distances entre les différents points caractéristiques sont proportionnelles à la distance verticale entre la bouche et les yeux. Le tableau 2.1 montre les coordonnées des points d'intérêt en fonction de l'axe de symétrie AS, l'axe de la bouche AB et le centre du cadre du visage YC. La valeur D est la distance entre l'axe des yeux et l'axe de la bouche.



(a) (b)

FIGURE 2.7 – Modèle anthropométrique des points caractéristiques faciaux

Afin de calculer le taux de détection des points faciaux de notre modèle anthropométrique, nous avons comparé les positions des points calculés et les positions des points marqués par un étiquetage manuel, en respectant un certain seuil pour considérer si le point est bien localisé ou pas. Le seuil utilisé est proportionnel à la distance D (10% D).
La localisation des points caractéristiques faciaux avec notre modèle anthropométrique a atteint un taux de 82.74% pour tous les points détectés. En revanche, la figure 2.8-b montre un exemple pratique d'une mauvaise localisation des points faciaux au niveau de la bouche.

Afin d'améliorer ces résultats, nous avons appliqué la méthode de Shi-Thomasi au voisinage des points d'intérêt. La combinaison du modèle anthropométrique et la méthode de Shi-Thomasi a donné un taux de 90.83%. Cette méthode nous a permis de régler le problème de la mauvaise localisation des points autour de la bouche (figure 2.8-b), plus particulièrement les commissures des lèvres (figure 2.8-d). La position de ces points est très importante pour la reconnaissance des expressions faciales car elle donne une information plus précise sur la déformation de la bouche.

La figure 2.9 illustre le résultat de la sélection automatique des points caractéristiques.

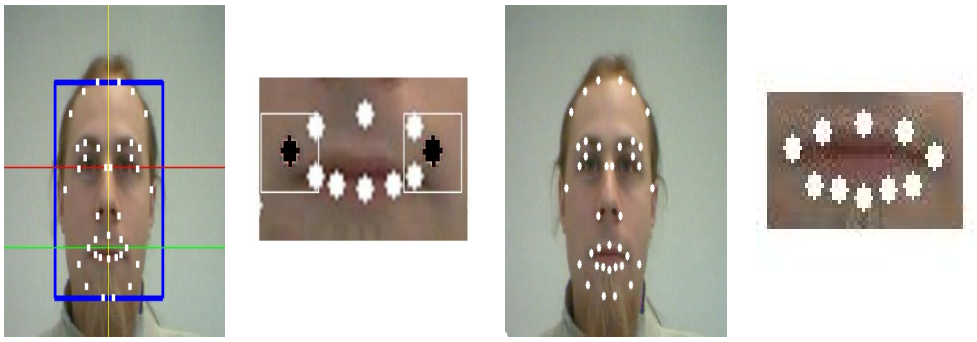
Les résultats expérimentaux montrent une bonne localisation des caractéristiques faciales indépendamment des conditions d'éclairage, de la couleur de la peau ou des accessoires : lunettes, barbes...

Point	X position	Y position	Point	X position	Y position
P1	≈ AS-0.91*D	≈ YC-0.91*D	P20	≈ AS+0.26*D	≈ YC-0.52*D
P2	≈ AS-0.58*D	≈ YC-1.17*D	P21	≈ AS+0.58*D	≈ YC-0.58*D
P3	≈ AS-0.26*D	≈ YC-1.3*D	P22	≈ AS+0.78*D	≈ YC-0.52*D
P4	≈ AS+0.26*D	≈ YC-1.3*D	P23	≈ AS-0.26*D	≈ YC+0.32*D
P5	≈ AS+0.58*D	≈ YC-1.17*D	P24	≈ AS+0.26*D	≈ YC+0.32*D
P6	≈ AS+0.91*D	≈ YC-0.91*D	P25	≈ AS-0.65*D	≈ AB
P7	≈ AS+1.04*D	≈ YC	P26	≈ AS-0.32*D	≈ AB-0.1*D
P8	≈ AS+0.71*D	≈ YC+0.91*D	P27	≈ AS	≈ AB-0.15*D
P9	≈ AS+0.52*D	≈ YC+1.17*D	P28	≈ AS+0.32*D	≈ AB-0.1*D
P10	≈ AS+0.13*D	≈ YC+1.30*D	P29	≈ AS+0.45*D	≈ AB
P11	≈ AS-0.13*D	≈ YC+1.3*D	P30	≈ AS+0.32*D	≈ AB+0.1*D
P12	≈ AS-0.52*D	≈ YC+1.17*D	P31	≈ AS+0.19*D	≈ AB+0.13*D
P13	≈ AS-0.71*D	≈ YC+0.91*D	P32	≈ AS	≈ AB+0.15*D
P14	≈ AS-1.04*D	≈ YC	P33	≈ AS-0.19*D	≈ AB+0.13*D
P15	≈ AS-0.06*D	≈ YC-0.26*D	P34	≈ AS-0.32*D	≈ AB+0.1*D
P16	≈ AS+0.06*D	≈ YC-0.26*D	P35	≈ AS-0.74*D	≈ YC-0.26*D
P17	≈ AS-0.78*D	≈ YC-0.52*D	P36	≈ AS-0.58*D	≈ YC-0.39*D
P18	≈ AS-0.58*D	≈ YC-0.58*D	P37	≈ AS+0.58*D	≈ YC-0.39*D
P19	≈ AS-0.26*D	≈ YC-0.52*D	P38	≈ AS+0.74*D	≈ YC-0.26*D

AS : axe de symétrie - *AB* : axe de la bouche - *YC* : centre du cadre du visage -*D* : distance entre l'axe des yeux et l'axe de la bouche

TABLE 2.1 – Proportion des positions des points caractéristiques du visage

Après la sélection des points d'intérêt, le suivi de ces derniers est réalisé par un algorithme du flux optique qui est l'algorithme pyramidal de Lucas-Kanade [30].



(a) (b) (c) (d)

FIGURE 2.8 – Exemple en temps réel : a- Résultat du modèle anthropométrique, b- Détection des points de la bouche avec le modèle anthropométrique, c- Résultat de la détection avec la méthode de combinaison, d- Détection des points de la bouche avec la méthode de combinaison

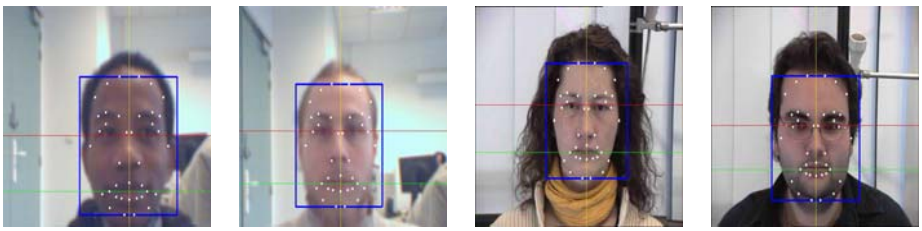


FIGURE 2.9 – Extraction des points caractéristiques dans la première image

2.2.3 Le suivi des points caractéristiques avec le flux optique

L'utilisation du flux optique procure une information de mouvement pour chaque pixel de l'image. Ainsi, il mesure les vecteurs de déplacement à partir de l'intensité des pixels de deux images consécutives ou temporellement rapprochées. Dans un contexte de détection de mouvement, les pixels inactifs posséderont alors une vélocité nulle contrairement aux pixels appartenant à des objets dynamiques.

Si l'on considère l'hypothèse que l'intensité lumineuse ne varie pas avec le temps et que les déplacements entre deux images consécutives sont faibles, on a alors [30] :

$$I(x+dx, y+dy, t+dt) \simeq I(x,y,t) \tag{2.2}$$

où $I(x,y,t)$ est la valeur (niveau de gris) du pixel (x,y) de l'image à l'instant t.

Calculer le flux optique revient à calculer en chaque point de l'image l'équation suivante :

$$-\frac{\delta I}{\delta t} = \frac{\delta x}{\delta t} \cdot \frac{\delta I}{\delta x} + \frac{\delta y}{\delta t} \cdot \frac{\delta I}{\delta y} \tag{2.3}$$

Lucas et Kanade ont ajouté de nouvelles contraintes pour garantir l'unicité de la solution de l'équation 2.3. La méthode de Lucas et Kanade consiste à rechercher le vecteur vitesse $\vec{U}$ $(\frac{\delta x}{\delta t}, \frac{\delta y}{\delta t})$ au point $p_i(x,y)$ en appliquant un calcul de moindres carrés pour minimiser la contrainte. On définit préalablement un voisinage et on cherche à optimiser $\vec{U}$ de sorte qu'il soit la solution du système suivant pour n points :

$$\begin{bmatrix} \frac{\delta I}{\delta x}(p1) & \frac{\delta I}{\delta y}(p_1) \\ \vdots & \vdots \\ \frac{\delta I}{\delta x}(p_i) & \frac{\delta I}{\delta y}(p_i) \\ \vdots & \vdots \\ \frac{\delta I}{\delta x}(p_n) & \frac{\delta I}{\delta y}(p_n) \end{bmatrix} \bullet \begin{bmatrix} \frac{\delta x}{\delta t} \\ \frac{\delta y}{\delta t} \end{bmatrix} = \begin{bmatrix} -\frac{\delta I}{\delta t}(p1) \\ \vdots \\ -\frac{\delta I}{\delta t}(p_i) \\ \vdots \\ -\frac{\delta I}{\delta t}(p_n) \end{bmatrix}$$

L'algorithme de Lucas et Kanade est implémenté en utilisant la bibliothèque d'OpenCV. Cet algorithme utilise des fenêtres de recherche autour des points d'intérêt afin de les retrouver dans l'image suivante.

Chacun des 38 points est recherché dans l'image suivante dans une fenêtre de taille p*q centrée sur la position du point dans l'image précédente. Le flux optique est estimé dans chaque fenêtre afin de trouver au mieux le déplacement de ce point [16, 14, 11].

Si la taille de la fenêtre est petite (par exemple $10*10$ pour une image de taille $640*480$), l'algorithme risque de converger vers un minimum local qui ne correspond pas au vrai déplacement. L'image 5 de la figure 2.10 montre un mauvais suivi du point de l'œil gauche à cause du clignement, ainsi dans l'image 51, l'un des points de la bouche perd sa position à cause de

l'ouverture.

Le choix d'une taille de fenêtre plus grande (par exemple $50 * 50$ pour une image de taille $640 * 480$) peut, dans certaines conditions, améliorer cette situation mais devient pénalisant dans les zones à fort gradient. L'image 67 de la figure 2.11 montre un mauvais suivi d'un point de la bouche (coté droit) qui s'est déplacé vers la dent.

Ce paramètre reste délicat à préciser et dépend de la taille et de la qualité de l'image utilisée. Pour notre application, nous avons utilisé une taille de $30 * 30$ pour une image de taille $640 * 480$. La figure 2.12 montre l'opportunité de ce choix qui a pu résoudre les deux problèmes précédents.

image2 image4 image5 image50 image51

FIGURE 2.10 – Le suivi des points caractéristiques dans une séquence vidéo avec une fenêtre de recherche de taille 10*10

image2 image4 image5 image67 image68

FIGURE 2.11 – Le suivi des points caractéristiques dans une séquence vidéo avec une fenêtre de recherche de taille 50*50

image1 image51 image54 image56 image61

FIGURE 2.12 – Le suivi des points caractéristiques dans une séquence vidéo avec une fenêtre de recherche de taille 30*30

Pour calculer le taux d'efficacité du suivi, nous avons comparé les positions des points estimés par l'algorithme et les positions des points marqués par un étiquetage manuel en respectant un certain seuil pour considérer si le point est bien localisé ou pas. Le seuil utilisé est proportionnel à la distance D (10% D). Le tableau 2.2 représente les résultats obtenus du suivi des points pour différentes taille de fenêtre de recherche, où nous avons un meilleur suivi pour une fenêtre de

taille 30*30.

Taille de la fenêtre	Taux du suivi (%)
10x10	96.82
30x30	100
50x50	98.21

TABLE 2.2 – Taux d'efficacité du suivi en fonction de la taille de la fenêtre de recherche

Après avoir détecté le visage et extrait les informations pertinentes, nous passons à l'étape suivante qui consiste à identifier et à reconnaître l'expression faciale affichée.

2.3 Reconnaissance des expressions faciales

Un grand nombre de systèmes d'analyse d'expressions faciales proposés dans la littérature visent à reconnaître et à mesurer l'amplitude d'unités d'actions faciales à partir du visage vu de face, fixe ou en mouvement [76]. D'autres systèmes cherchent plutôt à reconnaître un ensemble limité d'expressions « prototypes » telles que la joie, la colère, le dégoût, la tristesse, la peur, la surprise, ou d'autres actions telles qu'un clignement d'œil ou un cri. Dans le cadre de ce travail, nous nous sommes intéressés à la reconnaissance des expressions faciales définies par Ekman [70]. Dans ce chapitre, nous proposons une méthode basée sur la variation des distances caractéristiques par rapport à l'état neutre pour la représentation des expressions faciales. Nous montrons que cette méthode est simple et rapide permettant le traitement en temps réel. Pour la classification, nous utilisons la méthode de séparateur à vaste marge (SVM) qui présente les avantages suivants :

- ❒ Un nombre de paramètres faible à régler ;
- ❒ Une grande vitesse d'apprentissage ;
- ❒ Un nombre restreint d'échantillons suffit à la détermination des vecteurs supports permettant la discrimination entre les classes ;
- ❒ Traitement des problèmes linéaires ou non linéaires selon la fonction du noyau.

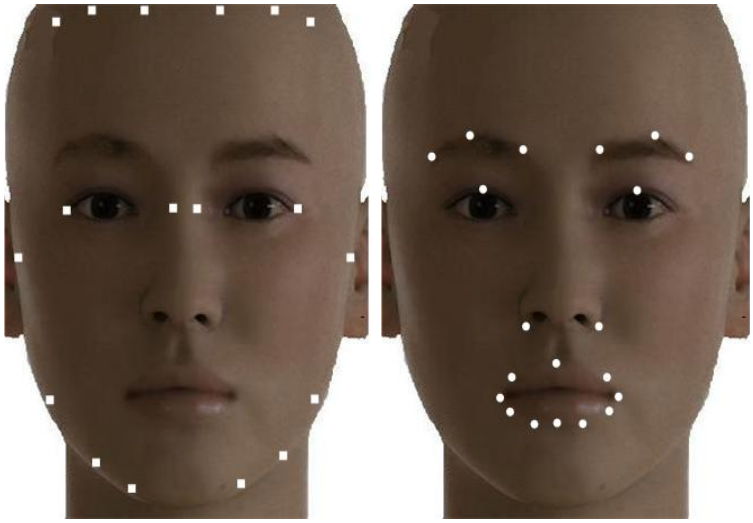
2.3.1 Codage des expressions faciales

Ekman et Izard [75] ont développé des méthodes de mesure des comportements du visage. En particulier, ils ont créé le système FACS (*Facial Action Coding System*) qui se base sur l'universalité. Ce système utilise une quarantaine de caractéristiques anatomiques indépendantes et définit une taxonomie de toutes les expressions faciales. Le système FACS est probablement le modèle le plus populaire utilisé pour classifier systématiquement les expressions physiques

des émotions du visage et est utilisé par des psychologues mais aussi par des infographistes. Ce système définit 46 unités d'actions, qui sont autant de contraction ou de relaxation d'un ou plusieurs muscles et dont l'association définit une expression faciale.

Lors de la production d'une expression faciale, il apparaît sur le visage un ensemble de transformations au niveau des traits du visage. Nous faisons l'hypothèse qu'il est possible de classer différentes expressions à partir de la variation de quelques distances calculées à partir des coordonnées des points d'intérêt qui délimitent les yeux, le nez, la bouche et les sourcils par rapport à l'état neutre [13].

Pour le traitement des mouvements faciaux, nous représentons chaque muscle par deux points, un point statique et un point dynamique [91]. Les points statiques sont des points qui ne bougent pas pendant la production d'une expression faciale (figure 2.13-a), par contre les points dynamiques subissent un mouvement selon le type de l'expression (figure 2.13-b).



a :Points statiques b :Points dynamiques

FIGURE 2.13 – Modèle des points faciaux

La figure 2.14 présente les différentes distances euclidiennes utilisées pour caractériser le mouvement des muscles faciaux, tel que :

- ❒ Les mouvements des sourcils sont décrits par les distances de D1 à D7 ;
- ❒ Les mouvements des yeux sont décrits par les distances D8 et D9 ;
- ❒ Les mouvements du nez sont décrits par les distances D10 et D11 ;
- ❒ Les mouvements de la bouche sont décrits par les distances de D12 à D21.

Ces distances sont calculées pour chaque image de la séquence.

Les figures 2.15 et 2.16 montrent l'évolution de quelques distances caractéristiques durant une séquence vidéo qui commence par un état neutre. Par exemple, la distance $D1$ diminue dans le cas de la surprise où les sourcils sont courbés vers le haut. En revanche, cette distance ne subit aucune variation dans le cas d'une joie où les sourcils sont décontractés.

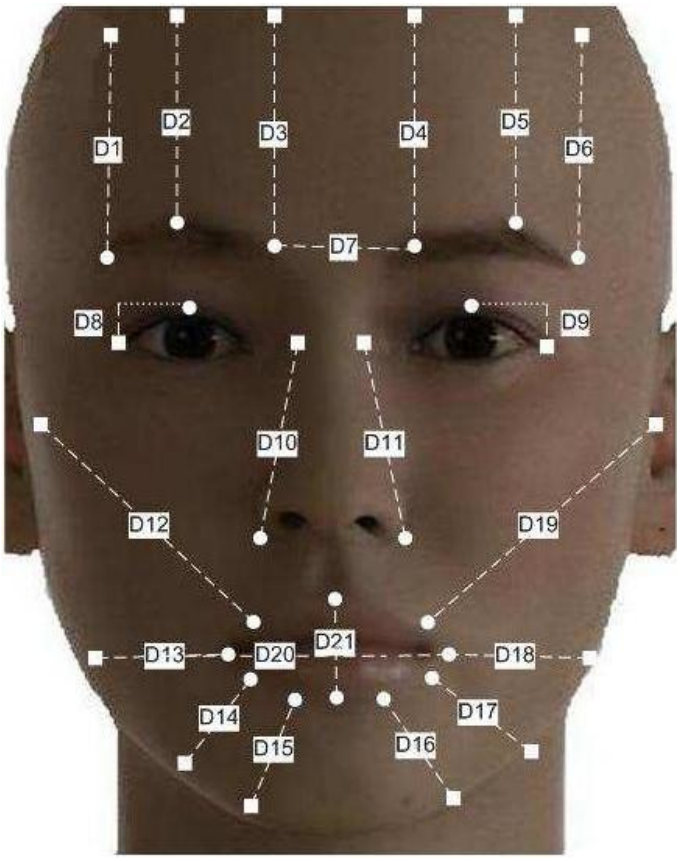


FIGURE 2.14 – Distances utilisées pour le codage des expressions faciales

Ces courbes vérifient bien la description des transformations, subies par chacun des traits du visage lors de la production des six expressions faciales (Tableau 2.3), fournie par la norme MPEG-4 [144].

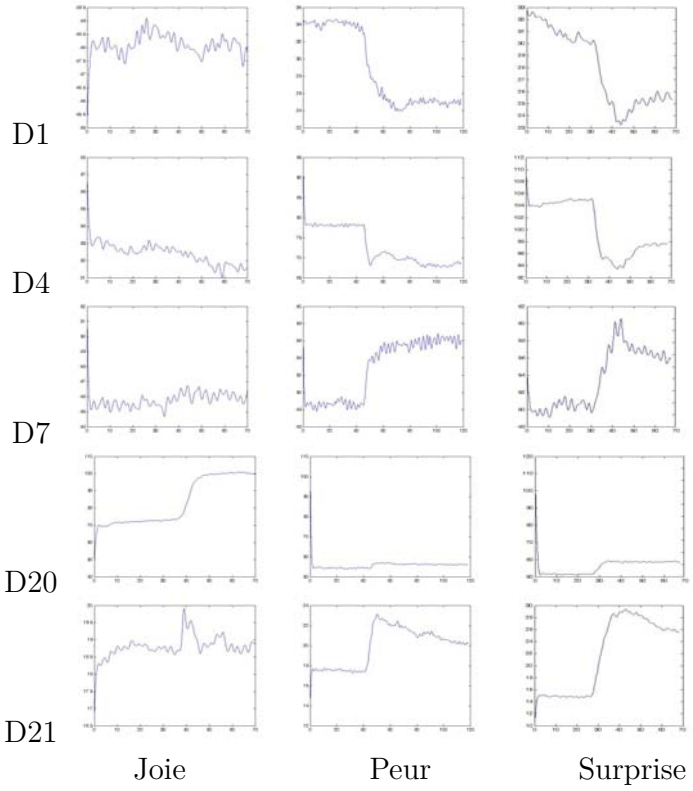
L'analyse de ces courbes montre que ces distances peuvent être utilisées pour encoder les expressions faciales. Si $D_T = (d_1, d_2, ..., d_n)$ est le vecteur des paramètres extraits à partir d'une séquence vidéo à l'instant T, ΔD est la variation des distances par rapport à la première image (une expression neutre).

En comparant entre les deux figures 2.15 et 2.16, nous pouvons constater que les distances varient dans le même sens pour les deux personnes et pour la même émotion mais avec des grandeurs différentes. Ce qui implique que ces distances ne peuvent pas être utilisées directement pour le codage des expressions faciales (conclusion vérifiée dans la section suivante).

2.3.2 Classification des expressions faciales

Après l'extraction des paramètres caractéristiques décrits dans la section précédente, nous avons utilisé un classifieur statistique supervisé SVM (Séparateur à Vaste Marge) pour faire la correspondance entre les paramètres calculés et les émotions correspondantes, pour plus de détail sur les SVM, voir annexe A.

La réalisation d'un programme d'apprentissage par SVM se ramène essentiellement à résoudre



Abscisse de la courbe : le numéro de l'image
Ordonnée de la courbe : la distance en pixel

FIGURE 2.15 – La variation des distances dans une séquence vidéo pour différentes émotions (Personne1 de la base FEEDTUM)

un problème d'optimisation impliquant un système de résolution de programmation quadratique dans un espace de dimension conséquente. La bibliothèque LibSVM [39] donne une implémentation de l'algorithme des SVM. L'utilisation de ce programme revient surtout à sélectionner une bonne famille de fonctions noyau et à régler les paramètres de ces fonctions (par exemple l'exposant pour les fonctions à noyau polynomial, ou bien l'écart type pour les fonctions à base radiale). Ces choix sont le plus souvent faits par une technique de validation croisée.

L'implémentation des SVM dans notre application est détaillée dans la section suivante.

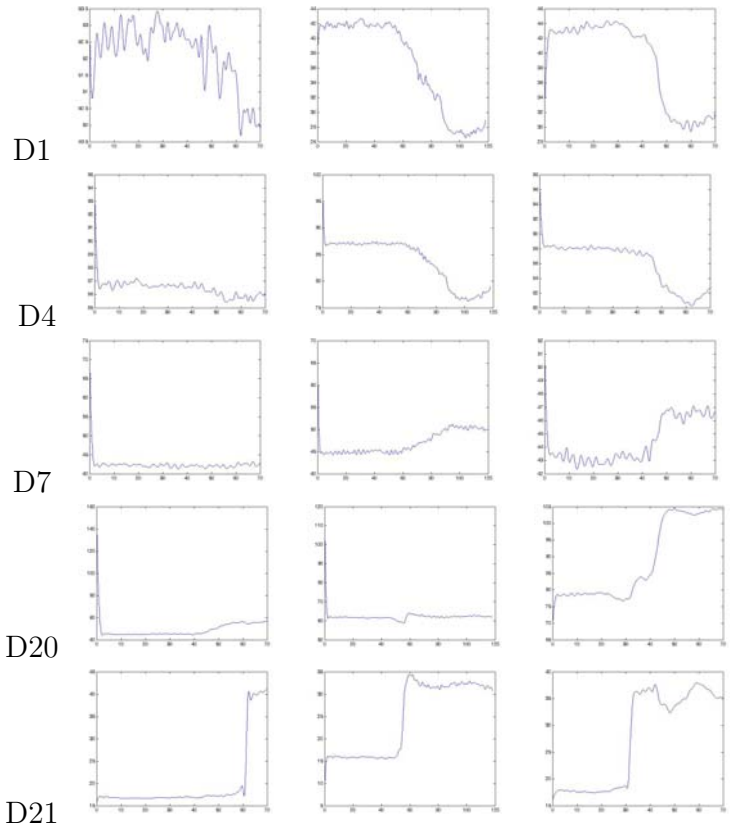
2.4 Résultats et discussions

2.4.1 Description des bases de données utilisées

Pour l'évaluation de notre travail nous avons utilisé deux bases publiques.

2.4.1.1 La base de Cohn-Kanade

Il s'agit d'une base de données de l'université Carnegie Mellon [110] qui contient des séquences d'images en niveaux de gris des expressions faciales pour hommes et femmes de différentes origines ethniques. L'orientation de la caméra est frontale. Les petits mouvements de la tête sont présents. La taille de l'image est 640 par 490 pixels (figure 2.17). Cette base est très utilisée pour



Abscisse de la courbe : le numéro de l'image
Ordonnée de la courbe : la distance en pixel

FIGURE 2.16 – La variation des distances dans une séquence vidéo pour différentes émotions (Personne 2 de la base FEEDTUM)

la reconnaissance des expressions faciales [58, 224, 45].

2.4.1.2 La base de FEEDTUM

Il s'agit d'une base de données de l'université technique de Munich [216] qui contient des séquences vidéo en couleur des expressions faciales pour hommes et femmes. L'orientation de la caméra est frontale. Les petits mouvements de la tête sont présents. La taille de l'image est 640 par 480 pixels (figure 2.18). Les expressions faciales sont générées en utilisant des films pour l'induction d'émotion.

Bien qu'il y ait 97 personnes à l'origine pour la base de Cohn-Kanade et 18 personnes pour la base de FEEDTUM, nous n'avons choisi que les images des personnes exprimant les 7 expressions faciales étudiées (tableau 2.4).
Notons que, pour le calcul du taux de reconnaissance, la sélection des images se fait d'une manière aléatoire afin de construire le corpus d'apprentissage ainsi que le corpus de test. Le tirage aléatoire est fait de telle sorte que les deux corpus d'apprentissage et de test contiennent des échantillons de toutes les classes et de toutes les personnes. Tous les résultats des tableaux du taux de reconnaissance sont obtenus en répétant l'opération cinq fois.

Expression	Description textuelle
Joie	Les sourcils sont décontractés. La bouche est ouverte et les commissures des lèvres retirées en arrière, vers les oreilles.
Surprise	Les coins intérieurs des sourcils sont courbés vers le haut. Les yeux sont légèrement fermés. La bouche est décontractée.
Dégoût	Les coins intérieurs des sourcils sont abaissés ensemble. Les yeux sont largement ouverts. Les lèvres sont serrées l'une contre l'autre ou ouvertes pour montrer les dents.
Colère	Les sourcils sont levés ensemble et leur partie intérieure est courbée vers le haut. Les yeux sont contractés et en état d'alerte.
Tristesse	Les sourcils et les paupières sont décontractés. La lèvre supérieure est levée et courbée, souvent de manière asymétrique.
Peur	Les sourcils sont levés. Les paupières supérieures sont ouvertes, les paupières inférieures, décontractées. La bouche est ouverte.

TABLE 2.3 – Les expressions faciales définies dans la norme MPEG-4 et leur description textuelle

	Cohn-Kanade	FEEDTUM
Nbr de Sujets	10	10
Nbr d'images par expression	6	20
Expressions faciales	7	7
Nbr total d'image	420	1400
Nbr d'images utilisées pour l'apprentissage	300	1000
Nbr d'images utilisées pour le test	120	400

TABLE 2.4 – Présentation des bases de données utilisées pour l'évaluation de notre système

neutre joie peur dégoût colère tristesse surprise

FIGURE 2.17 – Exemple des différentes expressions faciales de la base de Cohn-Kanade

neutre joie peur dégoût colère tristesse surprise

FIGURE 2.18 – Exemple des différentes expressions faciales de la base de FEEDTUM

2.4.2 Implémentation et Résultats

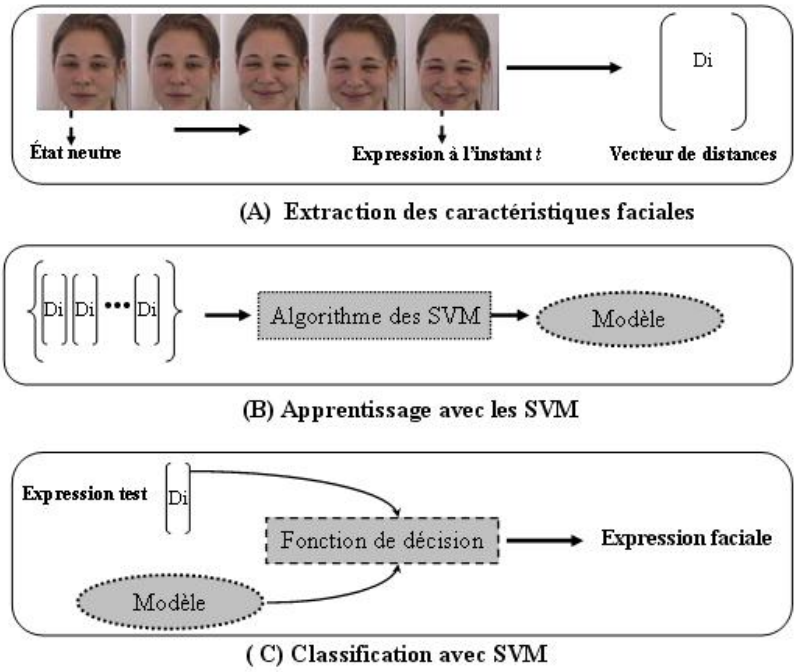


FIGURE 2.19 – Schéma de notre système de reconnaissance des expressions faciales

La figure 2.19 donne un descriptif de notre système de reconnaissance des expressions faciales :

1. La figure 2.19(A) montre l'étape de l'extraction des caractéristiques faciales qui a pour entrée une séquence vidéo et qui donne en sortie le vecteur de distances qui est l'entrée de l'étape suivante ;
2. La figure 2.19(B) montre l'étape de l'apprentissage qui a pour entrée une série de vecteurs de distances et qui donne en sortie le modèle des SVM ;
3. La figure 2.19(C) montre l'étape de la classification qui a pour entrée un vecteur de distances et qui donne en sortie l'expression reconnue en utilisant le modèle des SVM construit pendant l'étape d'apprentissage.

L'implémentation de la méthode des SVM est basée sur la bibliothèque libSVM [39]. Le noyau utilisé est la fonction de base radiale (RBF) donnée par :

$$K(x, x_i) = exp(-\frac{\| x - x_i \|}{\sigma}) \tag{2.4}$$

Les paramètres du SVM sont calculés en utilisant la méthode de la validation croisée qui permet d'acquérir une certaine robustesse dans le choix des paramètres.

Comme chaque expression faciale est représentée par un ensemble de distances, nous avons testé plusieurs combinaisons de ces distances pour choisir celle qui donne un meilleur taux de

reconnaissance (Tableau 2.5). Ces distances sont regroupées de telle sorte que chaque partie du visage est représentée séparément (les sourcils, les yeux, le nez et la bouche).

	les distances utilisées
Les sourcils	de $D1$ à $D7$
Les yeux	$D8$ et $D9$,
Le nez	$D10$ et $D11$
La bouche	de $D12$ à $D21$

TABLE 2.5 – Différentes combinaisons de distances utilisées pour le codage des expressions faciales

Pour le traitement de ces distances, nous avons testé deux types d'informations géométriques séparément :

A) Les 21 distances utilisées directement comme paramètres caractéristiques de l'expression faciale.

B) La variation ΔD par rapport à l'état neutre, en utilisant :

1. La différence : $\Delta D = D_i - D_{0i}$
   ou
2. Le rapport : $\Delta D = \frac{D_i}{D_{0i}}$

avec $i \in [1, 21]$.

Tel que :
D_i est la i^{eme} distance de l'image à traiter,
D_{0i} est la i^{eme} distance à l'état neutre présentant l'état de référence (hypothèse nécessaire pour la réalisation de notre application),

Combinaison choisie	Cohn-Kanade (%)	FEEDTUM (%)
Les sourcils :7D	61.66	79.10
Les yeux :2D	33.21	27.86
Le nez :2D	31.38	22.87
La bouche :10D	75.83	43.36

TABLE 2.6 – Taux de reconnaissance des expressions faciales pour chaque caractéristique faciale

Le tableau 2.6 représente les taux de reconnaissance des émotions obtenus pour différentes combinaisons de distances. Ces résultats montrent l'opportunité de chaque partie du visage, tel que :

❒ Pour la base de Kohn-Kanade :

1. La bouche est la partie la plus descriptive d'une expression faciale avec un taux $> 75\%$;
2. Les sourcils donnent une description importante avec un taux $> 60\%$;
3. Les yeux donnent une description moins importante avec un taux $> 33\%$;
4. Le nez est la partie la moins descriptive avec un taux de 31%.

❒ Pour la base de FEEDTUM :

1. Les sourcils sont la partie la plus descriptive d'une expression faciale avec un taux $> 79\%$;
2. La bouche donne une description importante avec un taux $> 43\%$;
3. Les yeux donnent une description moins importante avec un taux de 27% ;
4. Le nez est la partie la moins descriptive avec un taux $> 22\%$.

Cependant, l'expression faciale est bien décrite par la bouche dans une base et par les sourcils dans une autre. On ne peut donc pas conclure qu'une partie est plus descriptive des expressions faciales par rapport à une autre, ce qui mène à utiliser l'ensemble des parties du visage pour éviter toute ambiguïté.

Nous avons combiné les différentes parties du visage afin d'avoir plus d'informations sur l'expression faciale, en fusionnant les données (toutes les distances issues de chaque partie du visage), et en utilisant un seul classifieur SVM pour chaque base. Le tableau 2.7 montre les résultats de reconnaissance des expressions faciales obtenus pour différentes combinaisons (sourcils+nez+bouche et sourcils+yeux+nez+bouche). Nous remarquons que les meilleurs taux de reconnaissance sont obtenus en utilisant la deuxième combinaison (sourcils+yeux+nez+bouche), i.e. en fusionnant plus d'information (de distances).

Combinaison choisie	Cohn-Kanade	FEEDTUM
sourcil+ nez+ bouche :19D	81.10%	82.94%
sourcil+yeux+nez+bouche :21D	89.44%	95.32%

TABLE 2.7 – Taux de reconnaissance des expressions faciales pour différentes combinaisons de distances

Afin d'améliorer les résultats obtenus avec l'utilisation directe des distances, nous avons testé la variation de distance par rapport à l'état neutre (tableau 2.8). Nous remarquons que les résultats obtenus avec le rapport sont meilleurs que ceux obtenus avec la différence. L'utilisation du rapport pour calculer la variation des distances garantie l'homogénéisation des paramètres et permet d'assurer l'invariance par rapport au changement d'échelle, comme par exemple dans les cas du zoom de la caméra.

	Cohn-Kanade	FEEDTUM
Variation des distances par rapport à l'état neutre en utilisant le rapport	95.83%	97.5%
Variation des distances par rapport à l'état neutre en utilisant la différence	93.54%	96.81%

TABLE 2.8 – Taux de reconnaissance des expressions faciales en utilisant la variation des distances par rapport à l'état neutre

Les paramètres de notre modèle SVM obtenus avec la méthode de la validation croisée sont $\sigma = 21$ et $C = 1$ pour les deux bases.

Pour évaluer la classification de notre système, nous avons utilisé la matrice de confusion qui est un outil servant à mesurer la qualité d'un système de classification. La matrice de confusion est un tableau à double entrée [36]. En ligne s'expriment les résultats par rapport aux différentes classes définies. Les colonnes expriment les résultats par rapport aux expressions faciales de références. La cellule de croisement indique par conséquent le nombre d'expressions faciales appartenant à la classe i et assigné à la classe j. Les cellules correspondant à $i = j$ expriment le nombre d'expressions correctement affectées.

Un des intérêts de la matrice de confusion est qu'elle montre rapidement si le système parvient à classifier correctement les données.
Les tableaux 2.9, 2.10, 2.11 et 2.12 représentent les matrices de confusion de différentes émotions pour un noyau RBF avec la base de Cohn-Kanade et la base de FEEDTUM respectivement.
Cette matrice montre l'efficacité de la méthode de classification avec le séparateur à vaste marge, où le taux de reconnaissance le plus élevé correspond toujours à la bonne émotion.

Pour notre classifieur, nous avons obtenu de bons résultats pour la majorité des émotions avec un taux $\simeq 90\%$, à l'exception de quelques confusions entre les classes. Par exemple la confusion entre l'émotion neutre et l'émotion tristesse.

	Joie	Peur	Dégoût	Colère	Tristesse	Surprise	Neutre
Joie	**98.33%**	0%	0%	0%	0%	0%	1.66%
Peur	0%	**100%**	0%	0%	0%	0%	0%
Dégoût	0%	0%	**98.33%**	0%	0%	0%	1.66%
Colère	0%	0%	0%	**100%**	0%	0%	0%
Tristesse	0%	0%	0%	0%	**85%**	0%	15%
Surprise	0%	0%	0%	0%	0%	**100%**	0%
Neutre	0%	0%	0%	0%	5%	0%	**95%**

TABLE 2.9 – Matrice de confusion des émotions en utilisant la différence pour la base de Kohn-Kanade

Pour le choix du noyau du classifieur SVM, nous avons testé la fonction du noyau linéaire donnée par :

$$K(x, x_i) = x^T.x_i \tag{2.5}$$

En comparant entre les résultats obtenus avec les deux noyaux choisis (tableaux 2.13 et

	Joie	Peur	Dégoût	Colère	Tristesse	Surprise	Neutre
Joie	**98.33%**	0%	0%	0%	0%	0%	1.66%
Peur	0%	**100%**	0%	0%	0%	0%	0%
Dégoût	0%	0%	**98.33%**	0%	0%	0%	1.66%
Colère	0%	0%	0%	**100%**	0%	0%	0%
Tristesse	0%	0%	0%	0%	**91.66%**	0%	8.33%
Surprise	0%	0%	0%	0%	0%	**100%**	0%
Neutre	0%	0%	0%	0%	5%	0%	**95%**

TABLE 2.10 – Matrice de confusion des émotions en utilisant le rapport pour la base de Kohn-Kanade.

	Joie	Peur	Dégoût	Colère	Tristesse	Surprise	Neutre
Joie	**98%**	0%	0%	0.5%	0%	0%	1.5%
Peur	0%	**100%**	0%	0%	0%	0%	0%
Dégoût	0%	0%	**97.5%**	0%	2.5%	0%	0%
Colère	0%	0%	0%	**100%**	0%	0%	0%
Tristesse	0%	0%	0%	0%	**96%**	0%	4%
Surprise	0%	0%	0%	0%	0%	**100%**	0%
Neutre	0%	0%	0%	0%	1.5%	0%	**98.5%**

TABLE 2.11 – Matrice de confusion des émotions en utilisant la différence pour la base de FEEDTUM

	Joie	Peur	Dégoût	Colère	Tristesse	Surprise	Neutre
Joie	**98.5%**	0%	0%	0.5%	0%	0%	1%
Peur	0%	**100%**	0%	0%	0%	0%	0%
Dégoût	0%	0%	**99.5%**	0%	0.5%	0%	0%
Colère	0%	0 %	0 %	**100%**	0%	0%	0%
Tristesse	0 %	0 %	0 %	0%	**97%**	0 %	3%
Surprise	0%	0%	0%	0%	0 %	**100%**	0%
Neutre	0%	0%	0%	0%	2.5%	0%	**97.5%**

TABLE 2.12 – Matrice de confusion des émotions en utilisant le rapport pour la base de FEEDTUM

2.14), nous constatons que le noyau RBF améliore légèrement les résultats par rapport au noyau linéaire [82] car les classes traitées sont non linéairement séparables.

	Noyau linéaire	Noyau RBF
Variation des distances par rapport à l'état neutre en utilisant le rapport	94.44%	95.83%

TABLE 2.13 – Comparaison du taux de reconnaissance des expressions faciales en utilisant différents noyaux SVM (la base de Cohn-Kanade)

Pour bien évaluer les performances de notre système, nous avons mélangé les deux bases de données dans l'objectif d'avoir une base de données plus grande. Notons que les deux bases n'ont pas les mêmes conditions d'acquisition (éclairage et résolution).

	Noyau linéaire	Noyau RBF
Variation des distances par rapport à l'état neutre en utilisant le rapport	94.25%	97.5%

TABLE 2.14 – Comparaison du taux de reconnaissance des expressions faciales en utilisant différents noyaux SVM (la base de FEEDTUM)

La sélection des images se fait d'une manière aléatoire afin de construire le corpus d'apprentissage (1500 images) et le corpus de test (390 images).

Le taux de reconnaissance obtenu est aux alentours de 80%. Ce résultat est inférieur aux taux de reconnaissance en traitant chaque base séparément car les deux bases n'ont pas la même résolution et n'ont pas les mêmes conditions d'acquisition.

Le tableau 2.15, présente le temps nécessaire pour le calcul de chaque étape de notre approche, en utilisant un PC pentium 4 avec 2 GHz et 1 Go de mémoire, sous Windows XP. Le temps nécessaire pour localiser le visage et détecter les points d'intérêt est d'environ 0.234 secondes. Cette étape est appliquée uniquement à la première image. Le temps nécessaire pour la reconnaissance en temps réel des expressions pour chaque image est d'environ 0.03132 secondes (suivi + classification). La rapidité du calcul de notre système permet le traitement d'une séquence vidéo de 25 images par seconde, soit à partir d'une séquence vidéo enregistrée, ou bien une acquisition en temps réel.

Opération	Temps (S)
Localisation de visage	$\backsimeq$0.2
Détection des points avec le modèle anthropométrique	$\backsimeq$0.032
Détection des points avec la méthode de combinaison	$\backsimeq$0.034
Suivi des points	$\backsimeq$0.31
Classification	$\backsimeq$0.00032

TABLE 2.15 – Temps nécessaire pour chaque étape de notre système

2.5 Conclusion

Dans ce chapitre, nous avons présenté notre approche de la reconnaissance des émotions à partir des expressions faciales basée sur la variation de distances des traits caractéristiques du visage par rapport à l'état neutre. Le principe repose sur l'extraction des informations caractéristiques qui décrivent les émotions en se basant sur la norme MPEG-4. Ces informations consistent à représenter les muscles du visage par un point statique et un point dynamique qui bouge avec le mouvement du muscle. L'extraction de ces points est réalisée selon plusieurs étapes. La détection de visage dans la première image est basée sur les descripteurs de HAAR. L'extraction des caractéristiques faciales à l'intérieur du cadre qui délimite le visage est basée sur le fait que les visages humains sont construits dans la même configuration géométrique. Un

modèle anthropométrique est développé pour la détection des points d'intérêt combiné avec la méthode de Shi-Thomasi pour une meilleure localisation.
Le suivi des points caractéristiques dans les images de la séquence est assuré par l'algorithme pyramidal de Lucas et Kanade. Afin d'avoir une représentation adéquate des expressions faciales avec les points extraits, nous avons utilisé la variation des distances par rapport au cas neutre qui est plus robuste que l'utilisation directe de ces distances. Le vecteur contenant les variations des distances est l'entrée de notre classifieur séparateur à vaste marge (SVM).
Cette approche a été validée en utilisant deux bases de données. Le pourcentage de reconnaissance se situe aux alentours de 95% pour les deux bases en utilisant le noyau RBF.

Chapitre 3

Analyse des signaux physiologiques

3.1 Introduction

Le problème abordé dans ce chapitre est celui de la reconnaissance des états émotifs d'un utilisateur à partir de mesures physiologiques.

En effet, les capteurs utilisés dans notre projet se répartissent en deux catégories invasif et non invasif. Dans le premier cas, le dispositif d'acquisition est indépendant de l'utilisateur et peut être utilisé en toute occasion alors que l'utilisation de capteurs du deuxième type exige l'instrumentation de l'utilisateur. En revanche, la caméra nous permet d'accéder de manière indirecte et plus subjective aux émotions, alors que les mesures physiologiques contiennent une part d'information objective. On sait par exemple qu'une sensation de bien-être est liée à un ralentissement du rythme cardiaque et que la colère augmente la température corporelle.

Dans ce chapitre, nous nous intéressons tout d'abord à décrire les différents signaux physiologiques (conductance de la peau, volume sanguin périphérique, volume respiratoire, signal électromyographie et température cutanée) utilisés pour la prédiction émotionnelle, ainsi que leurs relations avec les processus émotionnels. Ensuite, nous détaillons les différentes étapes du traitement effectuées pour la reconnaissance des émotions à partir des signaux physiologiques.

3.2 Mesures physiologiques

Le matériel utilisé pour cette expérience est le ProComp Infiniti qui est un encodeur multimode à huit canaux servant à l'acquisition des données de biofeedback en temps réel.

Le mot « BIOFEEDBACK » (c'est-à-dire mot à mot « retour d'une donnée biologique ») caractérise une technique et par conséquent, un appareil qui permet au patient de visualiser même les plus petites variations de ses propres signaux physiologiques. Après plusieurs tentatives, le patient réussit à trouver « la clef » pour entrer dans le mécanisme de la régulation du paramètre visualisé [5].

La conception du ProComp Infiniti et de ses capteurs électroniques actifs répond à des normes de qualité élevées en termes de précision, de sensibilité, de durabilité et de convivialité. Tous les

capteurs sont parfaitement non effractifs [1] et ils n'exigent qu'un minimum de préparation avant leur utilisation. L'encodeur ProComp Infiniti permet d'acquérir :

- ❒ La conductance électrodermale (SKC) ;
- ❒ La forme d'onde du volume sanguin périphérique (BVP), rythme cardiaque et amplitude ;
- ❒ La forme d'onde, fréquence et amplitude de la respiration (VR) ;
- ❒ L'électromyographie (EMG) ;
- ❒ La température de la peau (SKT) ;
- ❒ L'électrocardiographie (ECG) ;
- ❒ L'électroencéphalographie (EEG).

Les capteurs transmettent les signaux à l'ordinateur au moyen du microprocesseur de l'encodeur ProComp Infiniti. L'encodeur échantillonne les signaux entrants, les numérise, les code et transmet ces données traitées à l'ordinateur via le connecteur logiciel TT-USB. La transmission passe par des câbles à fibres optiques donnant ainsi une pleine liberté de mouvements, une fidélité absolue des signaux et assurant l'isolement du signal. Une caractéristique unique du système permet d'inter-changer les capteurs. On peut ainsi créer toute une variété de configurations en changeant simplement le type de capteur.

FIGURE 3.1 – Matériel Procomp Infiniti [172]

Le connecteur TT-USB est branché à un des ports USB de l'ordinateur hôte. Il reçoit les données transmises par l'encodeur en format optique puis les convertit en format USB de manière à ce qu'elles puissent être comprises par l'ordinateur.

En fait, les signaux physiologiques utilisés dans notre travail, correspondants à des états émotionnels sont issus de 5 capteurs : conductance de la peau, volume sanguin périphérique, volume respiratoire, signal eléctromyographie et la température cutanée [143].

Nous avons écarté de notre choix le capteur ECG car il nécessite de déshabiller la personne et le capteur EEG à cause de ses électrodes qui se placent difficilement sur la tête.

1. Caractérise un acte médical ne comportant pas de passage à travers les tissus cutanés (peau).

3.2.1 La conductance de la peau (SKC)

C'est une mesure qui permet de déterminer le niveau de conductance électrique de la peau. Cette conductibilité est causée par la micro-sudation sécrétée par l'épiderme.

Le capteur de conductibilité de la peau est muni de deux petits fils. Au bout de chacun des électrodes, en Ag/AgCl, sont connectées à des attaches. Il possède deux électrodes amovibles à l'intérieur de sangles. Une faible tension électrique passe par les deux électrodes, généralement attachées autour de deux doigts de la main (figure 3.2), afin d'établir un circuit électrique dans lequel la personne devient la résistance variable. Ce processus permet de mesurer la variation de la conductibilité en temps réel puisqu'il s'agit de l'inverse de la résistance.

La sangle d'électrode doit être enroulée assez serrée autour du doigt, sans entraver la circulation sanguine, de manière à ce que la surface de l'électrode soit bien en contact avec la peau du doigt.

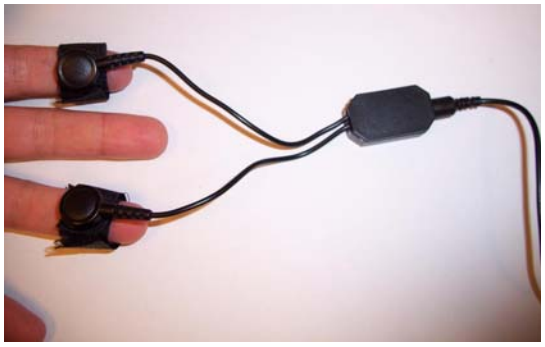


FIGURE 3.2 – Emplacement du capteur de conductance de peau

La conductibilité de la peau est mesurée en micro Siemens (µS). Une lecture typique de la conductibilité de la peau lorsque la personne est détendue devrait être d'environ 2 µS mais cette mesure peut varier considérablement en fonction de différents facteurs environnementaux et du type de peau.

Ces variations électriques de la peau se composent du niveau électrodermal qui correspond aux fluctuations électriques de base et de la réponse électrodermale.

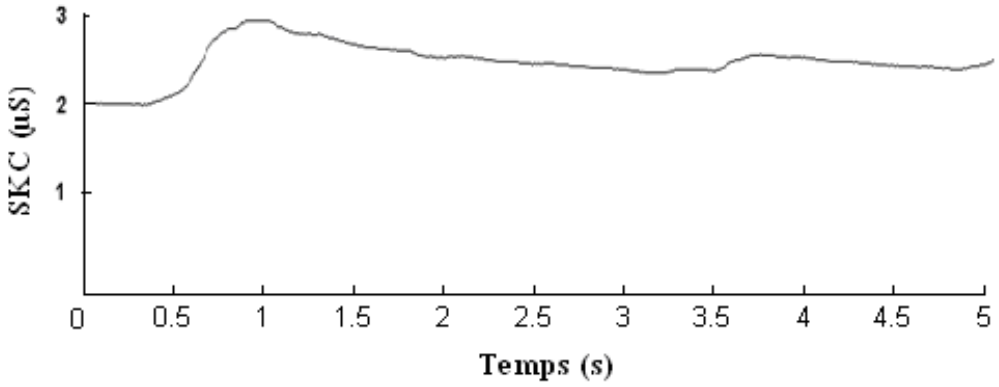


FIGURE 3.3 – Tracé obtenu suite à l'enregistrement de la conductance de la peau

Le signal de conductance de la peau représenté sur la figure 3.3, montre généralement une croissance rapide et une baisse relativement faible. La réponse électrodermale observée sur le tracé est retardée dans le temps. Elle apparaît après une certaine latence (0,5 seconde) et avec une amplitude de 2 µS. La bande passante du signal est à basse fréquence de 0 à 4 HZ.

3.2.2 L'électromyographie (EMG)

FIGURE 3.4 – Emplacement du capteur du signal EMG

La figure 3.4 montre un exemple de capteur EMG qui est muni de trois électrodes de surface, positive, négative et l'électrode de masse. Chaque électrode représente une plaque conductrice en métal recouverte avec un sel insoluble Ag/AgCl.

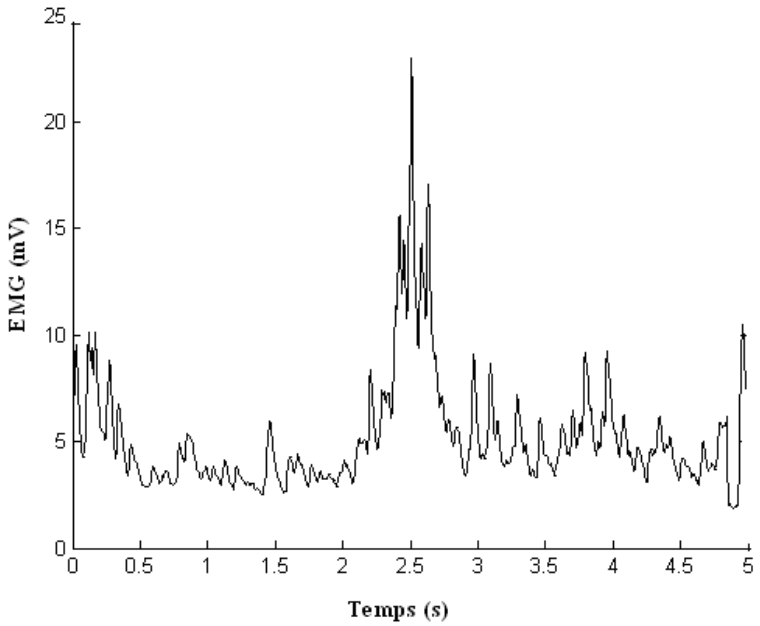


FIGURE 3.5 – Tracé obtenu suite à l'enregistrement du signal EMG

Pendant les mesures, les électrodes positives et négatives devraient être placées parallèlement

aux fibres musculaires des sourcils.

L'électrode de masse doit être placée sur une partie neutre, comme une protubérance, préférablement à une distance égale des deux autres électrodes.

Étant donné que les fibres musculaires se contractent toutes à des rythmes différents, le signal détecté par le capteur est la différence de potentiel en variation constante entre les électrodes positives et négatives. Le nombre de fibres musculaires sollicitées lors d'une contraction varie selon la force requise pour effectuer le mouvement. Par conséquent, l'intensité (l'amplitude) des signaux électriques est proportionnelle à la force de la contraction.

L'amplitude du signal EMG détecté est de 2 mV et d'énergie limitée de 20 à 500 HZ avec une énergie dominante dans la gamme de 50-150 HZ (figure 3.5). Les changements d'amplitude sont directement proportionnels à l'activité musculaire. Les valeurs normales au repos se situent habituellement entre 3 et 5 mV (elles peuvent être basses 0,3 mV voir moins si le muscle est complètement détendu).

3.2.3 Le volume sanguin périphérique (*Blood volume pulse* BVP)

FIGURE 3.6 – Emplacement du capteur du volume sanguin périphérique

Le capteur utilisé pour le volume sanguin périphérique, comprend deux dispositifs optoélectroniques :

-une source de lumière infrarouge : la LED convertit l'énergie électrique en énergie lumineuse, en émettant la lumière infrarouge ;

-un détecteur photoélectrique : c'est un dispositif assorti qui convertit l'énergie lumineuse réfléchie en énergie électrique.

Le capteur BVP se place sur le côté palmaire du doigt à l'aide de la sangle élastique (incluse avec le capteur) ou d'un bout de ruban adhésif (figure 3.6). À chaque battement de cœur le volume sanguin de la peau augmente. L'oxyhémoglobine qui est la composante principale dans le sang (90%), absorbe plus de lumière infrarouge. Donc, la lumière réfléchie diminue et si le volume sanguin diminue, alors la lumière infrarouge réfléchie augmente. Donc la lumière réfléchie varie en fonction du volume sanguin de la peau.

Le détecteur photoélectrique mesure les petites variations de l'intensité de la lumière associée aux changements du volume sanguin de la peau. Le signal BVP est une mesure relative qui n'a pas d'unité standard.

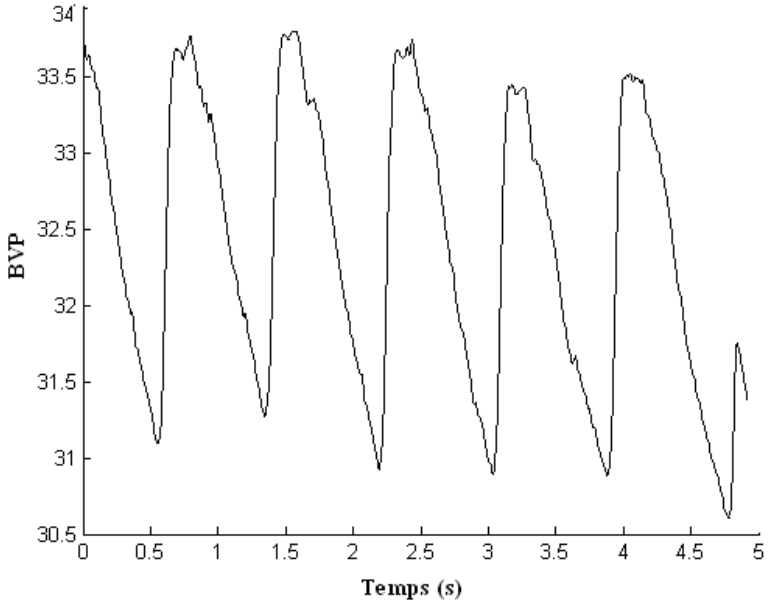


FIGURE 3.7 – Tracé obtenu suite à l'enregistrement du volume sanguin périphérique

Le tracé du volume sanguin périphérique représenté sur la figure 3.7, affiche généralement de fortes montées à cause des contractions systoliques qui sont suivies d'une descente plus lente. Pour certaines personnes, le signal affiche également un rebond sur la ligne descendante, qui correspond au pouls de la contraction diastolique. Le crête à crête du signal augmente ou diminue en fonction de l'excitation sympathique. La bande passante du signal est de 0 à 40 HZ.

3.2.4 Le volume respiratoire (VR)

FIGURE 3.8 – Capteur du volume respiratoire

Les variations du volume respiratoire sont détectées par une ceinture élastique à la hauteur du thorax ou de l'abdomen. Cette ceinture comporte deux bobines, chacune est alimentée par un faible signal radio-fréquence (RF). La ceinture élastique est sensible au gonflement de la cage thoracique. Les variations de sections entraînent des variations d'inductance. Alors, l'impédance des deux bobines augmente avec l'augmentation du volume respiratoire, cette relation dépend de plusieurs facteurs.

Ce capteur convertit les variations de volume de la cage thoracique en variations de tension électrique. La ceinture doit être placée autour de la région thoracique de la personne, juste

au-dessus de la poitrine (figure 3.8). Il devrait tenir en place lorsque la personne expire complètement. Le signal de la respiration est une mesure relative du volume de la cage thoracique (figure 3.9). La bande passante du signal est de 0,1 à 10 HZ.

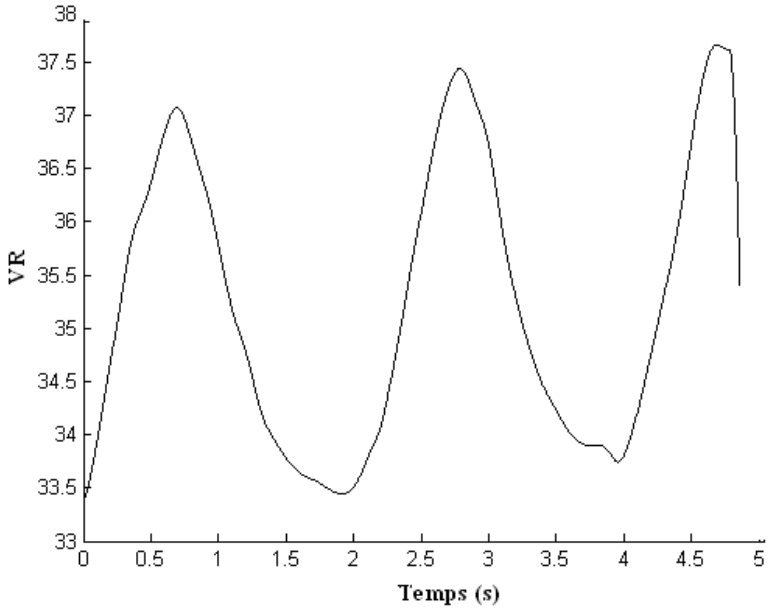


FIGURE 3.9 – Tracé obtenu suite à l'enregistrement du volume respiratoire

3.2.5 La température cutanée (*Skin Temperature* SKT)

FIGURE 3.10 – Emplacement du capteur de température cutanée

La température périphérique, telle que mesurée aux extrémités du corps, varie en fonction de l'irrigation sanguine dans la peau avec une thermistance qui convertit les changements de température en variations de courant électrique. La thermistance est attachée sur la face dorsale ou palmaire de n'importe quel doigt ou orteil à l'aide de la sangle du capteur (figure 3.10).

La figure 3.11 montre un exemple illustrant les changements de la température périphérique qui sont lents de fréquence de 0 à 1HZ et de faible amplitude autour de 21,62 C° (l'amplitude varie selon le sujet et l'environnement d'acquisition).

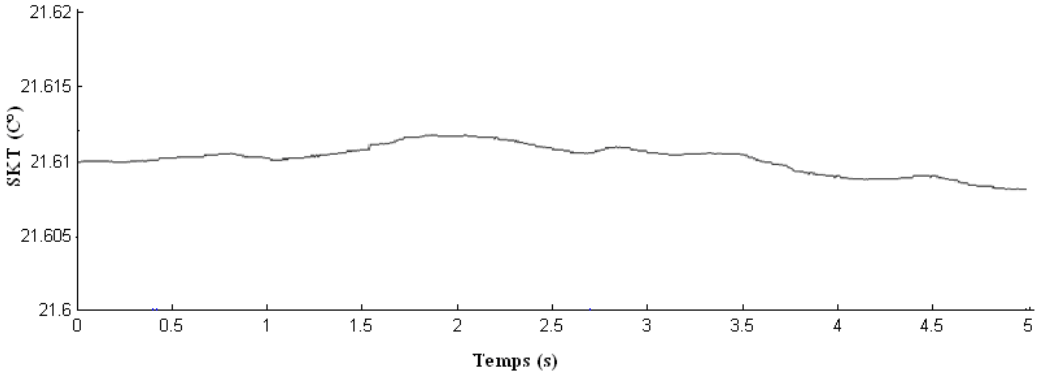


FIGURE 3.11 – Tracé obtenu suite à l'enregistrement de la température de la peau

3.3 Pré-traitement

Pour filtrer les 5 signaux physiologiques, nous avons utilisé deux types de filtres, afin de choisir le meilleur d'entre les deux :

1. Un filtre RIF optimal à phase linéaire, utilisant l'algorithme de Parks Mc Clelan ou appelé de Remez [210] ;
2. Un filtre RII de Butterworth.

Cependant, la procédure du filtrage par ces deux types de filtres nécessite la fixation de certains paramètres comme :

- ❒ fs qui est la fréquence d'échantillonnage du filtre ;
- ❒ la bande de fréquence du signal.

A des fins de choix, chaque signal parmi les cinq est filtré par les deux filtres.

Le tableau 3.1 représente les paramètres de chaque signal utilisés pour le filtrage.

Le signal	La bande fréquentielle
EMG	[50hz,150hz]
SKC	[0hz,4hz]
VR	[0hz,10hz]
SKT	[0hz,1hz]
BVP	[0hz,40hz]

TABLE 3.1 – Les paramètres des signaux utilisés

Chaque filtre est caractérisé par sa fonction de transfert $H = B/A$, sachant que les coefficients de A et de B du filtre RIF sont calculés par la fonction *firpme* de Matlab et ceux du filtre RII par la fonction *butter*. Les tableaux B.1 et B.2 représentent les coefficients des filtres utilisés pour chaque signal (annexe B).

Les figures B.1, B.2, B.3, B.4 et B.5 montrent des exemples pour les différents signaux utilisés.

D'après les résultats du filtrage, nous constatons qu'il n'y a pas de changement au niveau des signaux originaux, surtout pour le VR, SKT et BVP, ce qui confirme la remarque de Kim et al. [118] qu'aucun traitement de signal n'est nécessaire pour l'extraction des caractéristiques. Nous avons remarqué une petite amélioration au niveau du graphique du signal EMG et de la conductance de la peau. Mais pour le taux de reconnaissance, le filtrage n'a pas apporté grand chose, l'étape de la reconnaissance sera détaillée dans la prochaine section.

Nous pouvons expliquer cela par la qualité des capteurs qui répondent à des normes élevées en termes de précision, de sensibilité, de durabilité et de convivialité.

3.4 Reconnaissance des émotions

3.4.1 Extraction des caractéristiques

Après l'acquisition des signaux physiologiques, il est important de définir une méthodologie pour permettre au système de traduire les signaux acquis vers une émotion spécifique. La première étape consiste à extraire des informations caractéristiques pour le module de classification.

Pour chaque signal physiologique, nous avons calculé 6 paramètres caractéristiques définis comme suit [169] :

La moyenne temporelle

$$\mu_x = \frac{1}{T}\sum_{t=1}^{T} X(t) = \bar{X}(t) \tag{3.1}$$

L'écart type

$$\sigma_x = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(X(t)-\mu_x)^2} \tag{3.2}$$

La dérivée première

$$\delta_x = \frac{1}{T-1}\sum_{t=1}^{T-1} |X(t+1)-X(t)| \tag{3.3}$$

La dérivée première normalisée

$$\bar{\delta}_x = \frac{1}{T-1}\sum_{t=1}^{T-1} |\bar{X}(t+1)-\bar{X}(t)| = \frac{\delta_x}{\sigma_x} \tag{3.4}$$

La dérivée seconde

$$\gamma_x = \frac{1}{T-2}\sum_{t=1}^{T-2}|X(t+2)-X(t)| \tag{3.5}$$

La dérivée seconde normalisée

$$\bar{\gamma}_x = \frac{1}{T-2}\sum_{t=1}^{T-2}|\bar{X}(t+2)-\bar{X}(t)| = \frac{\gamma_x}{\sigma_x} \tag{3.6}$$

Tels que t est le numéro de l'échantillon et T est le nombre total des échantillons. En utilisant ces paramètres caractéristiques, nous obtenons un vecteur caractéristique X de 30 valeurs pour chaque échantillon :

$$\begin{aligned} X \quad = \quad & [\mu_{bvp}\ \sigma_{bvp}\ \delta_{bvp}\ \bar{\delta}_{bvp}\ \gamma_{bvp}\ \bar{\gamma}_{bvp}\ \mu_{emg}\sigma_{emg}\ \delta_{emg} \\ & \bar{\delta}_{emg}\ \ \gamma_{emg}\ \bar{\gamma}_{emg}\ \mu_{sc}\ \sigma_{sc}\ \delta_{sc}\ \bar{\delta}_{sc}\ \gamma_{sc}\ \bar{\gamma}_{sc}\ \mu_{skt}\ \sigma_{skt} \\ & \delta_{skt}\ \ \bar{\delta}_{skt}\ \gamma_{skt}\ \bar{\gamma}_{skt}\ \mu_{resp}\sigma_{resp}\ \delta_{resp}\ \bar{\delta}_{resp}\ \gamma_{resp}\ \bar{\gamma}_{resp}] \end{aligned} \tag{3.7}$$

3.4.2 Classification

Après l'extraction des paramètres caractéristiques décrite dans le paragraphe précédent, nous avons utilisé un classifieur statistique SVM pour faire la correspondance entre les paramètres calculés et les émotions correspondantes.

3.5 Induction de l'émotion

Il est indispensable d'obtenir une base de données de signaux physiologiques représentant des états émotionnels spécifiques. Afin d'acquérir une base de données dans laquelle l'influence de l'état émotionnel a été fidèlement reflétée, nous avons développé un ensemble de protocoles élaborés pour l'induction de l'émotion. Nous avons utilisé le système international de l'image affective (*international affective picture system* IAPS), développé par Lang et al. [127], et adopté par de nombreuses études de psycho-physiologie impliquant une réaction d'émotion. Notre choix s'est porté sur les images IAPS vu la simplicité de la manipulation. Nous avons testé un autre inducteur (film) [4] dont les réactions provoquées étaient trop faibles.

La base de données utilisée est constituée de 8 sujets sains (hommes) du deuxième et troisième cycle universitaire.

Les émotions étudiées dans ce chapitre sont : amusement, contentement, neutre, dégoût, peur et tristesse.

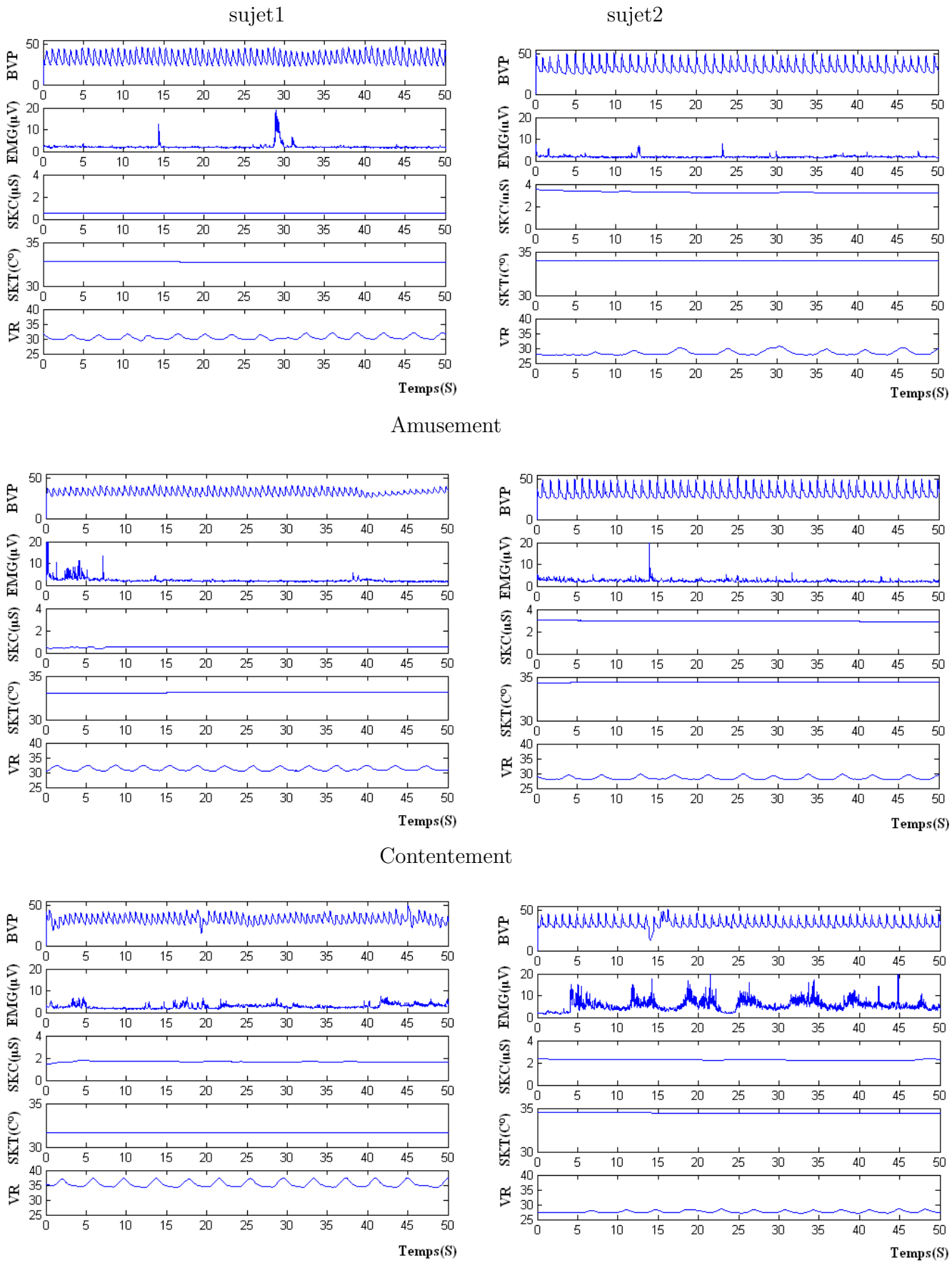


FIGURE 3.12 – Échantillons des signaux physiologiques correspondant aux six émotions

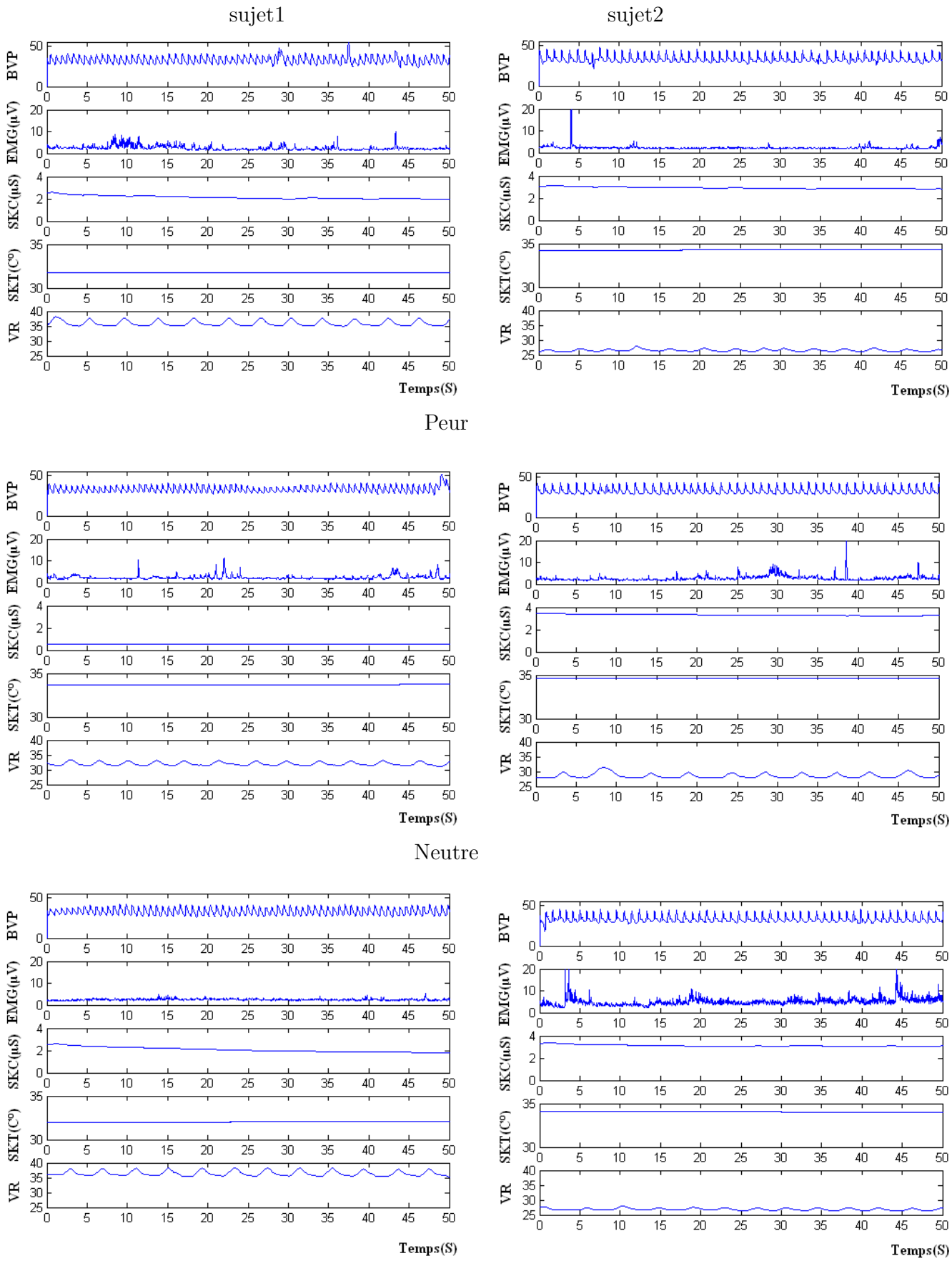


FIGURE 3.13 – Échantillons des signaux physiologiques correspondant aux six émotions

3.6 Résultats et discussion

Les figures 3.12 et 3.13 présentent un exemple de cinq signaux physiologiques acquis pendant l'induction des émotions pour deux sujets. Nous pouvons constater que le signal physiologique varie largement en fonction des émotions et des personnes.

Sujet	Taux de reconnaissance
sujet1	96.66 %
sujet2	97.57 %
sujet3	99.42 %
sujet4	90.33 %
sujet5	99.23 %
sujet6	90.76 %
sujet7	88.23 %
sujet8	97.57 %
Tous les sujets	60.45 %

TABLE 3.2 – Taux de reconnaissance des émotions en utilisant des fenêtres de taille T=17 échantillons avec un noyau RBF

Le tableau 3.2 représente le taux de reconnaissance des émotions pour huit sujets en utilisant les cinq signaux physiologiques avec la méthode des SVM. Toutefois, le mélange des données de toutes les personnes donne un taux de reconnaissance de 60%, sachant que la réponse physiologique n'est pas la même pour tout le monde. Cela dépend du sexe, l'origine, la culture et le vécu de la personne.

Notons que, pour calculer ces taux de reconnaissance (tableau 3.2), nous avons mélangé deux acquisitions afin de construire une base de données plus grande. Le choix des échantillons de test pour cette base est réalisé à l'aide d'un tirage aléatoire qui est fait de telle sorte que les deux corpus d'apprentissage et de test contiennent des échantillons de toutes les classes. Le mélange de deux bases nous permet également de vérifier l'impact de la durée des acquisitions sur le taux de reconnaissance, ce point sera détaillé dans la section 5.5.4.

Pour les tableaux 3.3, 3.4 et 3.5, nous avons utilisé deux acquisitions différentes pour calculer le taux de reconnaissance des émotions pour huit sujets et en mélangeant tous les sujets. Dans chaque tableau, nous avons présenté :

- ❒ les résultats des signaux bruités dans la première colonne ;
- ❒ les résultats des signaux filtrés avec un filtre RIF dans la deuxième colonne ;
- ❒ les résultats des signaux filtrés avec un filtre RII dans la troisième colonne.

La différence entre les trois tableaux est la taille de la fenêtre T. D'après les trois essais, nous remarquons que l'influence de la taille de la fenêtre sur le taux de reconnaissance est négligeable.

Le tableau 3.6 représente le taux de reconnaissance des émotions obtenu pour différentes

combinaisons des signaux physiologiques (un, deux, trois, quatre et cinq signaux) en mélangeant tous les sujets. La comparaison des résultats conduit aux observations suivantes :

1. La moyenne du taux de reconnaissance en utilisant un seul signal est de 20% ;
2. La moyenne du taux de reconnaissance en utilisant deux signaux est de 26% ;
3. La moyenne du taux de reconnaissance en utilisant trois signaux est de 33% ;
4. Alors que la moyenne des taux de reconnaissance en utilisant cinq signaux est de 45%.

Ces résultats montrent que l'utilisation de plusieurs signaux physiologiques est meilleure pour la reconnaissance des émotions.

Sujet	Taux sans filtrage (%)	Taux avec filtrage RIF (%)	Taux avec filtrage RII (%)
sujet1	71.33	70.52	71.33
sujet2	51.28	60.76	51.28
sujet3	51.19	52.04	51.19
sujet4	37.14	34.33	37.14
sujet5	57.57	58.90	57.57
sujet6	31.76	32.47	31.76
sujet7	43.66	44.52	43.66
sujet8	45.61	32.04	45.61
Tous les sujets	45.01	36,98	45.01

TABLE 3.3 – Taux de reconnaissance des émotions en utilisant des fenêtres de taille T=17 échantillons avec un noyau RBF

Sujet	Taux sans filtrage (%)	Taux avec filtrage RIF (%)	Taux avec filtrage RII (%)
sujet1	43.75	51.38	52.08
sujet2	52.77	52.08	50
sujet3	60.41	60.41	58.33
sujet4	36.80	38.19	34.72
sujet5	65.97	63.88	51.38
sujet6	25	25	25
sujet7	34.02	43.75	38.88
sujet8	47.22	43.75	33.33
Tous les sujets	42.62	38.36	32.55

TABLE 3.4 – Taux de reconnaissance des émotions en utilisant des fenêtres de taille T=256 échantillons avec un noyau RBF

Sujet	Taux sans filtrage (%)	Taux avec filtrage RIF (%)	Taux avec filtrage RII (%)
sujet1	30	26.66	30
sujet2	50	43.33	50
sujet3	53.33	53.33	43.33
sujet4	33.33	36.66	33.33
sujet5	60	53.33	66.66
sujet6	20	23.33	23.33
sujet7	30	40	23.33
sujet8	43.33	40	23.33
Tous les sujets	37.91	37.5	36.25

TABLE 3.5 – Taux de reconnaissance des émotions en utilisant des fenêtres de taille T=1280 échantillons (5 secondes) avec un noyau RBF

Signal physiologique	Taux (%)
BVP	19.91
EMG	20.92
SC	21.76
SKT	16.96
RESP	24.88
BVP,EMG	22.25
BVP,SC	22.37
BVP,SKT	32.39
BVP,RESP	22.41
EMG,SC	21.92
EMG,SKT	30.80
EMG,RESP	23.38
SC,SKT	34.96
SC,RESP	29.98
BVP,EMG,SC	23.48

Signal physiologique	Taux (%)
BVP,EMG,SKT	33.11
BVP,EMG,RESP	22.30
BVP,SC,SKT	40.93
BVP,SC,RESP	28.93
BVP,SKT,RESP	37.54
EMG,SC,SKT	38.41
EMG,SC,RESP	30.11
EMG,SKT,RESP	37.42
SC,SKT,RESP	47.12
BVP,EMG,SC,SKT	40.75
BVP,EMG,SC,RESP	27.41
BVP,SC,SKT,RESP	45.95
EMG,SC,SKT,RESP	44.77
BVP,EMG,SC,SKT,RESP	45.01

TABLE 3.6 – Taux de reconnaissance des émotions pour différentes combinaisons de signaux physiologiques (pour tous les sujets)

Le choix du noyau

Afin de voir l'effet du type du noyau sur l'apprentissage, nous avons refait les tests avec un noyau linéaire.

D'après les tableaux 3.7, 3.8 et 3.9, nous concluons que le noyau linéaire donne de meilleurs résultats pour un corpus d'apprentissage ayant un nombre petit d'échantillons. En revanche, le noyau RBF est préféré pour des bases de données plus élevées.

Le tableau 3.10 représente la matrice de confusion des différentes émotions pour tous les sujets. Cette matrice montre l'efficacité de la méthode de classification SVM, pour laquelle le taux de reconnaissance le plus élevé correspond toujours à la bonne émotion.

Sujet	Taux sans filtrage(%)	Taux avec un filtrage RIF(%)	Taux avec un filtrage RII(%)
sujet1	69.76	70.28	69.76
sujet2	71.80	71.90	71.80
sujet3	48.38	50.42	48.38
sujet4	38.04	35.09	38.04
sujet5	56.23	56.33	56.23
sujet6	33.33	33.33	33.33
sujet7	45.47	46.66	45.47
sujet8	45.47	46.66	46.95
Tous les sujets	31.01	35.41	31.01

TABLE 3.7 – Taux de reconnaissance des émotions en utilisant des fenêtres de taille T=17 échantillons avec un noyau linéaire

Sujet	Taux sans filtrage(%)	Taux avec filtrage RIF(%)	Taux avec filtrage RII(%)
sujet1	52.77	45.83	49.30
sujet2	57.63	56.25	33.33
sujet3	60.41	59.72	59.72
sujet4	33.33	37.5	35.41
sujet5	70.13	75.69	56.94
sujet6	31.25	33.33	31.25
sujet7	43.75	56.94	47.22
sujet8	52.77	45.13	33.33
Tous les sujets	37.67	35.50	32.03

TABLE 3.8 – Taux de reconnaissance des émotions en utilisant des fenêtres de taille T=256 échantillons avec un noyau linéaire

Sujet	Taux sans filtrage(%)	Taux avec filtrage RIF(%)	Taux avec filtrage RII(%)
sujet1	36.66	36.66	56.66
sujet2	56.66	56.66	56.66
sujet3	53.33	46.66	60
sujet4	36.66	36.66	36.66
sujet5	73.33	70	53.33
sujet6	30	26.66	40
sujet7	40	43.33	40
sujet8	60	53.33	53.33
Tous les sujets	38.75	39.16	35.41

TABLE 3.9 – Taux de reconnaissance des émotions en utilisant des fenêtres de taille T=1280 échantillons (5 secondes) avec un noyau linéaire

Émotion	1	2	3	4	5	6
1	**85.34%**	1.79%	0.31%	0.2%	11.44%	0.88%
2	19.48%	**47.37%**	20.95%	0%	11.26%	0.91%
3	13.06%	4.11%	**71.72%**	0.01%	9.97%	1.11%
4	0.31%	0.04%	0.23%	**74.96%**	12.13%	12.5%
5	0.15%	0%	0.11%	19.01%	**74.81%**	5.8%
6	0.29%	0.12%	0.27%	19.76%	23.27%	**55.56%**

TABLE 3.10 – Matrice de confusion des émotions pour la méthode des SVM (Tous les sujets) : 1- Amusement, 2- Contentement, 3- Dégoût, 4- Peur, 5- Neutre, 6- Tristesse

3.7 Conclusion

Dans ce chapitre, nous avons présenté une approche de reconnaissance automatique des émotions, fondée sur le traitement des signaux physiologiques. La démarche consiste à extraire des informations caractéristiques des signaux, suivie par une étape de classification basée sur la méthode des SVM. Les données physiologiques sont acquises pour six différents états émotifs. Les résultats expérimentaux montrent que nous avons atteint un taux de reconnaissance de 45% (pour une base globale) pour les six émotions (amusement, contentement, dégoût, peur, tristesse et la neutralité). Afin d'améliorer ce taux de reconnaissance, nous allons fusionner les informations issues des signaux physiologiques et celles issues de expressions faciales. Les méthodes de fusions font l'objet du prochain chapitre.

Chapitre 4

Système multimodal

4.1 Introduction

Les informations provenant des expressions faciales et des signaux physiologiques peuvent être regroupées en trois niveaux : bas niveau (les pixels et les signaux), moyen niveau (les caractéristiques ou bien les attributs) et haut niveau (les décisions). Dans notre étude, nous nous intéressons au niveau moyen et haut niveau. Par la suite, nous allons présenter différentes méthodes pour chacun d'entre eux.

Une étude des méthodes les plus fréquemment utilisées dans les systèmes multimodaux est présentée dans la suite à des fins de choix. Nous commençons par la fusion des caractéristiques en utilisant en premier lieu l'information mutuelle qui permet de faire une sélection des informations pertinentes et en deuxième lieu une méthode de transformation de données basée sur les composantes principales.

Concernant la fusion de décisions, nous avons opté pour une méthode simple qui est la méthode de vote. Pour bénéficier de la notion dynamique des données, nous avons développé également une méthode de fusion de décisions basée sur les réseaux Bayésiens.

La fusion d'information consiste à combiner des informations issues de plusieurs sources afin d'améliorer la prise de décision.

4.2 Fusion des caractéristiques

Le problème de la fusion des caractéristiques a été abordé sous différents angles. Cela avait comme conséquence immédiate une très grande divergence dans les points de vue. Néanmoins tous s'accordent à résoudre le problème fondamental posé par Fu et Rosenfeld [88], formulé en trois questions :

1. Quel est le nombre minimal de paramètres à utiliser ?
2. Quels sont ces paramètres ?

3. Dans quel ordre doivent-ils être pris en compte ?

Quelques chercheurs qui adoptent une approche mathématique pour résoudre le problème de sélection des paramètres pertinents, rencontrent, dans la plupart des cas, les deux sous problèmes fondamentaux suivants :

- ❒ Le choix du critère de sélection du sous-ensemble de paramètres ;
- ❒ La procédure de recherche du sous-ensemble de paramètres.

Le premier sous-ensemble traduit le besoin d'un critère qui va évaluer « l'efficacité » et la « qualité » des sous ensembles de paramètres. Les critères les plus utilisés sont :

- ❒ La probabilité de mauvaise classification ;
- ❒ La divergence ;
- ❒ La fonction d'entropie ;
- ❒ La mesure de séparabilité.

Le deuxième sous-problème est le choix de la procédure à utiliser pour la recherche du meilleur sous-ensemble de m paramètres. De ce choix va dépendre la qualité du résultat.

Nous présentons dans ce paragraphe deux méthodes, l'une basée sur les composantes principales et l'autre basée sur l'information mutuelle.

4.2.1 Méthode d'analyse en composantes principales (ACP)

En effet, la sélection de paramètres est considérée ici comme un problème de réduction de dimension de l'espace des formes E^n. Cette réduction est réalisée par la recherche de l'espace optimale E^m, où va s'effectuer la classification, mais surtout par celle de la transformation à appliquer aux vecteurs forme avant la classification.

Les transformations utilisées pour cette réduction sont linéaires. Ainsi, les nouveaux paramètres sont toujours des combinaisons linéaires des mesures initiales.

Dans l'espace des formes E^n, chaque forme est représentée par un point X_k de coordonnées :

$$\{x_{k1}, x_{k2}, ..., x_{ki}, ..., x_{kn}\} \tag{4.1}$$

La distance Euclidienne entre deux formes X_k et X_l notée $d(X_k, X_l)$ est définie par :

$$d^2(X_k, X_l) = \sum_{i=1}^{n} (x_{ki} - x_{li})^2 \tag{4.2}$$

Le principe est alors de présenter ces formes dans un espace X^m $(m < n)$ telles qu'en moyenne les distances entre tous les couples de points dans E^m soient une bonne représentation

des distances dans E^n.

Cherchons un sous-espace E^1 de E^n, de dimension 1, et remplaçons les points X_k de E^n par leurs projections orthogonales sur E^1, que l'on notera $P_1(X_k)$. Le problème est donc de trouver le sous-espace E^1 , qui maximise le terme :

$$\sum_{k=1}^{N} \sum_{l=1}^{N} ||P_1(X_k) - P_1(X_l)||^2 \tag{4.3}$$

Où le N représente le nombre de formes.

En effet, en projetant les formes sur un axes u_1, on cherche à les étaler au maximum pour les rendre séparables.

La solution est le sous-espace engendré par le vecteur propre v_{11}, associé à la plus grande valeur propre de la matrice de covariance. E est appelé premier axe factoriel de l'ensemble des points et les projetés des points sont appelés premières composantes principales.

Plus généralement, chercher un espace E^m satisfaisant au même critère revient à chercher les vecteurs propres unaires $v_1, v_2, ..., v_m$, associés aux plus grandes valeurs propres de la matrice de covariance.

4.2.2 Méthode de sélection basée sur l'information mutuelle

Dans cette section, nous allons voir comment l'information mutuelle peut être utilisée pour évaluer l'importance de chacune des variables (caractéristiques faciales et caractéristiques physiologiques) de l'émotion à reconnaître.

L'algorithme de l'information mutuelle consiste à chercher le sous-ensemble S, de dimension d inférieure à celle de l'ensemble total des descripteurs initiaux, qui maximise l'information mutuelle $IM(C, S)$ entre cet ensemble et la variable C des classes d'appartenance.

L'information mutuelle est largement utilisée pour la sélection des attributs [121, 25, 166]. En général, elle mesure la quantité d'information d'une variable contenue dans une seconde. Ainsi, lorsque cette valeur est maximale, les deux variables sont dites « identiques ». Sélectionner le descripteur étant le plus lié à la classe C peut se faire en maximisant leur information mutuelle. Généralement, cette information est basée sur la notion d'entropie. Néanmoins, on trouve 3 définitions équivalentes dont chacune permet d'expliquer différemment cette information (équations 4.5, 4.6 et 4.7).

4.2.2.1 Entropie

L'entropie est la quantité d'information contenue dans une source qui peut être un système physique, une image ou autre. Cette source d'information est l'ensemble des réalisations d'un descripteur $X = (x_1, ..., x_n)$ ayant toutes une probabilité p_i d'être réalisée. Plus les p_i sont équiprobables, plus son entropie $H(X)$ est grande. Shannon [192] propose une définition de l'entropie

telle que :

$$H(X) = -\sum_i p_i.log(p_i) \tag{4.4}$$

Cette expression montre bien que plus un élément est rare, plus il a de signification. Par exemple, si une image est constituée de pixels de plusieurs niveaux de gris, alors elle contient une information plus importante que celle d'un seul niveau de gris. La définition de Shannon de l'entropie indique l'information moyenne que l'on peut tirer de chaque élément de l'image.

4.2.2.2 Information mutuelle

Soit X une variable à n réalisations ($X = x_1, ..., x_n$) et C une variable discrète à k classes ($C = c_1, ..., c_k$).

La première définition de l'information mutuelle $IM(C,X)$ utilise la différence de l'entropie de X et de l'entropie de la même variable X sachant la classe C :

$$IM(C,X) = H(X) - H(X|C) = H(C) - H(C|X) \tag{4.5}$$

$H(X|C)$ mesure la quantité d'information contenue dans X lorsque la classe C est fournie. L'information mutuelle correspond à la quantité d'information que C possède sur X, ou inversement, la quantité d'information que X possède sur C.

La seconde définition évoque la distance de Kullback-Leibler qui mesure la distance entre deux distributions :

$$IM(X,C) = \sum_{x \in X} \sum_{c \in C} p_{xc} log \frac{p_{xc}}{p_x p_c} \tag{4.6}$$

Cette définition de l'information mutuelle mesure la dépendance entre X et C. L'attribut X sera sélectionné s'il est fortement corrélé à C.

La troisième définition de l'information mutuelle est une combinaison des entropies séparées et jointes des deux variables :

$$IM(C,X) = H(X) + H(C) - H(X,C) \tag{4.7}$$

L'entropie jointe $H(X,C)$ mesure la quantité d'information que les deux variables apportent en même temps. S'ils sont proches, on dira qu'une variable explique bien la seconde et l'entropie jointe est minimale.

La procédure de sélection des variables qui seront utilisées dans le système de reconnaissance des émotions, parmi les p variables disponibles, est alors la suivante :

1. Minimiser la redondance :

$$min\, W_I, \qquad W_I = \frac{1}{|S|^2} \sum_{i,j \in S} IM(i,j)$$

S est l'ensemble des caractéristiques, $IM(i,j)$ est l'information mutuelle entre les caractéristiques i et j.

2. Maximiser la pertinence :

$$max\, V_I, \qquad V_I = \frac{1}{|S|^2} \sum_{i \in S} IM(h,i)$$

h est la classe correspondante (émotion).

3. Combiner entre la minimisation de la redondance et la maximisation de la pertinence revient à maximiser :

$$max\, (V - W)$$

4.3 Fusion de décisions

La combinaison des résultats de plusieurs systèmes uni-modaux constitue une voie prometteuse pour améliorer les performances globales de la reconnaissance d'émotions. En effet, une décision fondée sur un grand nombre d'informations d'origines et de natures variées est généralement plus robuste que toute décision prise individuellement par chaque source d'information.

4.3.1 Méthode non paramétrique : Vote

Le principe du vote est la méthode de fusion d'informations la plus simple à mettre en œuvre. Plus qu'une approche de fusion, le principe du vote est une méthode de combinaison particulièrement adaptée aux décisions de type symbolique [146]. Notons $\mathrm{S}_j(\mathrm{x}) = \mathrm{i}$ le fait que la source Sj attribue la classe Ci à l'observation x. Nous supposons ici que les classes Ci sont exclusives. A chaque source nous associons la fonction indicatrice :

$$M_i^j = \begin{cases} 1 & si \quad S_j(x) = i, \\ 0 & sinon. \end{cases} \tag{4.8}$$

La combinaison des sources s'écrit par :

$$M_k^E(x) = \sum_{j=1}^{m} M_k^j(x), \tag{4.9}$$

Pour tout k, l'opérateur de combinaison est associatif et commutatif. La règle du vote majoritaire consiste à choisir la décision prise par le maximum de sources, c'est-à-dire le maximum de M_k^E. Cependant, cette règle simple n'admet pas toujours de solutions dans l'ensemble des classes $D = C1, ..., Cn$. En effet, si le nombre de sources m est pair et que $m/2$ sources décident C_{i1} et $m/2$ autres sources décident C_{i2}. Nous sommes obligés d'ajouter une classe C_{n+1} qui représente l'incertitude totale liée au conflit des sources sous l'hypothèse de l'exhaustivité des classes $Cn + 1 = C_1, ..., C_n$. La décision finale de l'expert prise par cette règle s'écrit donc par :

$$E(x) = \begin{cases} k & si \quad max_k M_k^E(x), \\ n+1 & sinon. \end{cases} \tag{4.10}$$

Cette règle est cependant peu satisfaisante dans les cas où deux sources donnent le maximum pour des classes différentes. La règle la plus employée est la règle du vote majoritaire absolu qui s'écrit :

$$E(x) = \begin{cases} k & si \quad max_k M_k^E(x) > \frac{m}{2}, \\ n+1 & sinon. \end{cases} \tag{4.11}$$

A partir de cette règle il a été démontré [126] plusieurs résultats prouvant que la méthode du vote permet d'obtenir de meilleures performances que toutes les sources prises séparément, sous des hypothèses d'indépendance statistique des sources et de même probabilité, et ceci est d'autant plus vrai que m est impaire. Il est possible de généraliser le principe du vote majoritaire afin de supprimer le conflit. L'idée est d'employer une somme pondérée [223] au lieu de combiner les réponses des sources par une somme simple comme dans l'équation 4.9 :

$$M_k^E(x) = \sum_{j=1}^{m} \alpha_{jk} M_k^j(x), \tag{4.12}$$

où $\sum_{j=1}^{m} \sum_{k=1}^{n} \alpha_{jk} = 1$. Les poids α_{jk} représentent la fiabilité d'une source pour une décision donnée et l'estimation de ces poids peut se faire à partir des taux normalisés de réussite pour chaque classe et chaque classifieur. Notons qu'alors nous introduisons une connaissance a priori non nécessaire précédemment. Les différentes règles de décision possibles peuvent se résumer par la formule suivante :

$$E(x) = \begin{cases} k & si \quad max_i M_i^E(x) > cm + b(x), \\ n+1 & sinon. \end{cases} \tag{4.13}$$

où c est une constante de $[0,1]$ et $b(x)$ est une fonction de $M_k^E(x)$.

4.3.2 Méthodes paramétriques

Parmi les approches probabilistes, les réseaux Bayésiens (RB) parfois appelés aussi modèles graphiques sont classiquement 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 [106, 107, 150].

Les avantages de l'utilisation d'un tel (RB) dans le cadre de notre application sont :

- ❒ Traitement des imprécisions : comme les capteurs utilisés (caméra ou bien capteurs physiologiques) sont généralement affectés par une imprécision qui varie avec le temps et po-

tentiellement défaillante, l'utilisation de la théorie probabiliste (Bayes) est justifiée pour modéliser les erreurs des capteurs ;

- ❒ Généralisation du filtrage de Kalman : un filtre de Kalman n'est qu'un cas particulier d'un (RB) [150]. En plus, cet outil nous offre la possibilité de manipuler les variables discrètes et continues dans le même réseau ce que ne permet pas de faire un filtre de Kalman classique ;

- ❒ Manipulation de la multi-hypothèse : le RB permet de gérer l'ambiguïté comme le fait le filtre particulaire sans se soucier du nombre de particules à utiliser ;

- ❒ Fusion de données reconfigurable : Les capteurs utilisés sont amenés à fournir des informations imparfaites (incertaines), le (RB) permet la fusion de plusieurs modalités et ainsi d'accéder à une information globale plus fiable et plus complète parce qu'il permet d'exploiter au mieux les avantages de chacune des sources d'information, tout en essayant de pallier leurs limitations individuelles. La complémentarité et la redondance des informations sont alors deux facteurs essentiels pour obtenir un tel effet.

4.3.2.1 Modélisation par réseaux Bayésiens (RB)

La modélisation d'un processus stochastique par un RB consiste d'abord à définir les variables X^i du modèle. L'identification de ces variables se fait généralement de façon naturelle et intuitive par un expert qui attribue à chaque variable l'ensemble des états qu'elle peut prendre [26]. Soit $val(X^i)$ l'ensemble des états du nœud X^i :

$$val(X^i) = \left\{x_1^i, x_2^i, ..., x_{ni}^i\right\} \tag{4.14}$$

La seconde phase de la modélisation consiste à trouver les liens de dépendance entre les nœuds : c'est la définition de la structure qui peut être construite directement par avis d'experts, ou par apprentissage. Une fois la structure du modèle graphique définie, l'apprentissage des paramètres sert à établir les Tables de Probabilités Conditionnelles (TPC). A partir de ces dernières, les probabilités conditionnelles $P(X^i|X^j = x_k^j)$ sont calculées par inférence en se basant sur le calcul des lois jointes.

4.3.2.2 Les Réseaux Bayésiens Dynamiques (RBD)

Les Réseaux Bayésiens Dynamiques (RBD), introduits par [61], sont une extension des RB modélisant des processus stochastiques variant dans le temps. En plus des nœuds statistiques introduits par les RB classiques, les RBD introduisent un nouveau type de nœuds dits temporels pour modéliser des variables aléatoires discrètes dépendant du temps. Les RBD généralisent les systèmes dynamiques linéaires (LDS) et les modèles de Markov cachés (MMC) en représentant les états cachés (et observés) en tant que variables d'états, possédant des interdépendances complexes. La structure graphique dans un RBD fournit une manière simple de détailler ces

indépendances conditionnelles et fournit ainsi une paramétrisation réduite du modèle. En effet, la propagation de la probabilité jointe d'un RBD est obtenue par une équation équivalente au calcul des probabilités jointes dans un RB statique (équation 4.15) :

$$P(X^1, X^2, ..., X^n) = \prod_{i=1}^{n} P(X^i | pa(X^i)) \tag{4.15}$$

$$P(X^1_{t+1}, X^2_{t+1}, ..., X^n_{t+1}) = \prod_{i=1}^{n} P(X^i_{t+1} | pa(X^i_{t+1})) \tag{4.16}$$

L'équation 4.16 est à l'origine d'algorithmes d'inférence qui explicitent clairement l'avantage des réseaux Bayésiens dynamiques par rapport aux méthodes classiques de modélisation (Diagramme d'état). En effet, ces équations lient la structure du RBD et les informations issues des tables de probabilités conditionnelles TPC et des TP pour calculer les probabilités jointes à chaque pas de temps. Ainsi, les informations d'indépendance données par le graphe servent à factoriser la loi jointe, et par conséquence, de réduire de façon très considérable le nombre de paramètres à définir.

Si nous considérons une variable aléatoire X_t dépendant du temps, il existe deux façons différentes de représenter son évolution. La première approche consiste à dérouler un RB statique sur une période de temps T (Voir Figure 4.1(a)) ce qui permet de surveiller la dynamique du nœud sur cette fenêtre de temps. La deuxième façon de représenter cette variable aléatoire est le modèle compact [32] pour lequel X_t est représenté sur deux tranches de temps t et $t+1$. La mise à jour de la table de probabilités du nœud X_t est effectuée par une itération successive [220], c'est à dire la distribution de probabilités calculée à l'instant t du nœud X_{t+1} est transmise au nœud X_t à l'itération suivante. Cette procédure est représentée par l'arc en pointillé sur la figure 4.1(b).

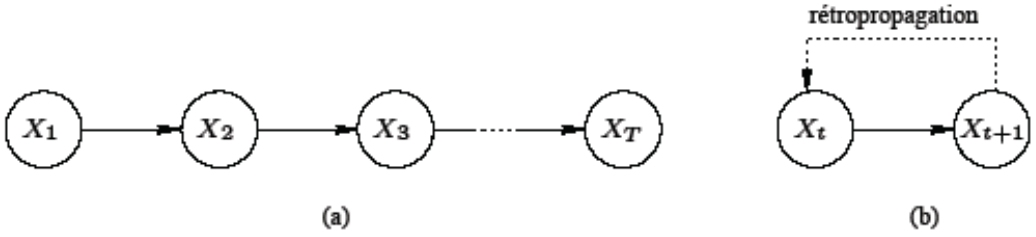


FIGURE 4.1 – Représentation d'une variable aléatoire dynamique : a- Représentation déroulée, b- Représentation compacte

Le problème de la fusion de décisions pour la reconnaissance des émotions relève d'une question d'estimation d'une variable cachée (émotion) d'un système dynamique (les décisions des expressions faciales et des signaux physiologiques) étant donné un historique d'observation (dans un intervalle de temps). Par ailleurs, les variables mesurées pour les raisons énoncées ci-dessus ne sont pas précises. Elles peuvent même être manquantes ou non cohérentes temporellement. C'est pourquoi, nous avons choisi dans ce travail de recherche, une approche fondée sur la modélisation probabiliste pour sa robustesse à l'incertitude tant des modèles que des données mesurées.

4.4 Protocole d'induction des émotions

La méthode courante pour induire les émotions consiste à demander à un acteur de sentir ou d'exprimer une émotion particulière. Cette stratégie a été largement utilisée pour l'évaluation de l'émotion à partir des expressions faciales et, dans une certaine mesure, à partir de signaux physiologiques [98]. Toutefois, même si les acteurs font un effort pour sentir profondément l'émotion qu'ils cherchent à exprimer, il est difficile d'assurer les réponses physiologiques qui sont cohérentes et reproductibles par des non-acteurs. En outre, les bases de données des acteurs de jeux sont souvent loin des émotions présentes dans la vie quotidienne.

Une autre approche pour induire des émotions est de présenter des stimuli particuliers à un participant ordinaire. De nombreux stimuli peuvent être utilisés comme des images, des sons, des vidéos [137] ou des jeux vidéo. Cette approche présente l'avantage de ne pas avoir besoin d'un acteur professionnel et que les réponses sont plus proches de celles observées dans la vie quotidienne. Il est indispensable d'obtenir une base de données de signaux physiologiques et des expressions faciales émotionnelles représentant des états émotionnels spécifiques. Afin d'acquérir une base de données dans laquelle l'influence de l'état émotionnel a été fidèlement reflétée, nous avons développé un ensemble de protocoles élaborés pour l'induction de l'émotion. Nous utilisons le système international des images affectives (*international affective picture system* IAPS), développé par Lang et al. [127], et adopté par de nombreuses études psycho physiologiques impliquant une réaction d'émotion.

Nous avons construit notre base de données avec 10 sujets sains (8 hommes et 2 femmes) du deuxième et troisième cycle universitaire. Les mesures ont été effectuées durant quatre jours différents pour chaque participant. Les images IAPS utilisées pour l'induction ont été choisies par un psychologue. La figure 4.2 montre un participant portant les capteurs physiologiques et regardant les images IAPS sur un PC portable avec une caméra intégrée.

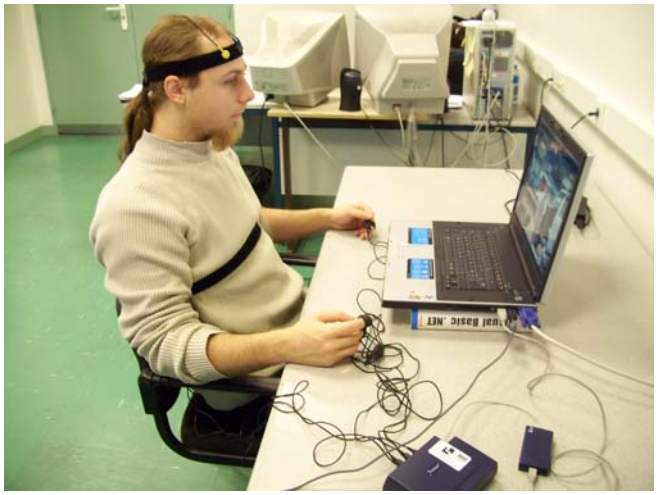
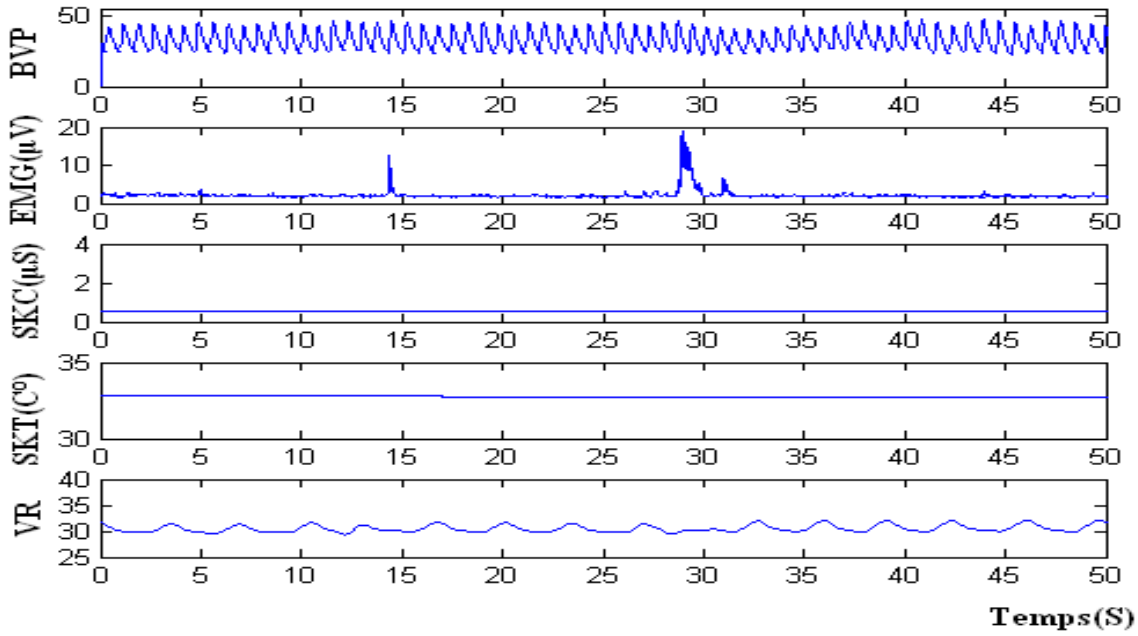


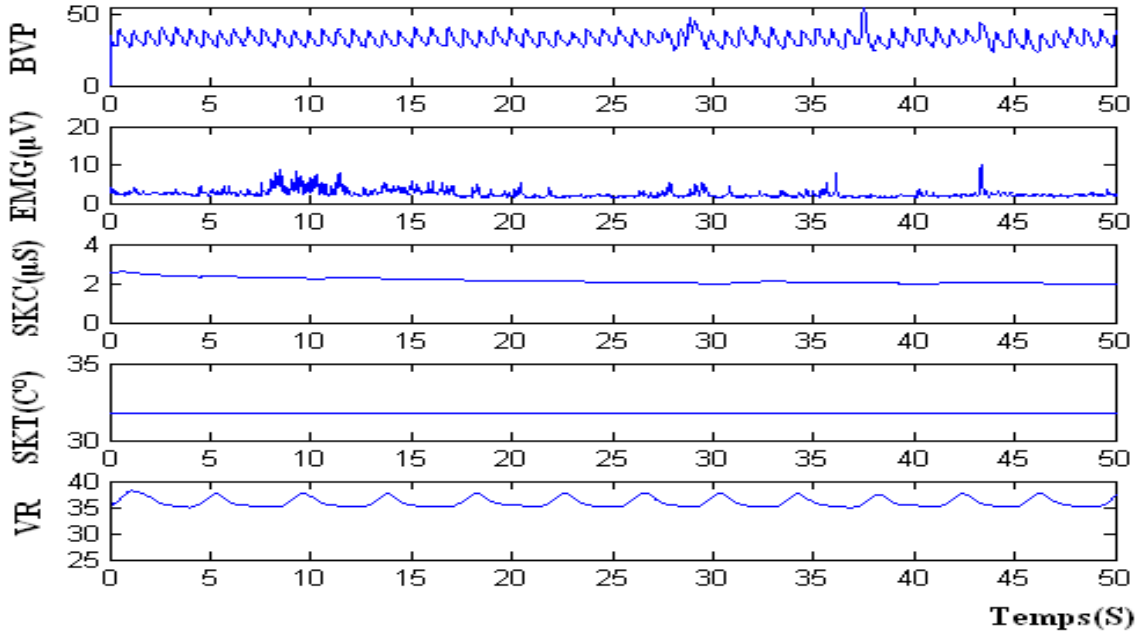
FIGURE 4.2 – Environnement d'acquisition

Notons que dans ce travail, deux états émotionnels ont été étudiés (positif et négatif).

Des échantillons de signaux physiologiques sont présentés dans la figure 4.3. La figure 4.4 montre des exemples d'expressions faciales pour 3 cas.



Positive



Négative

FIGURE 4.3 – Exemple des signaux physiologiques pour deux états émotionnels : positif et négatif

Neutre Positive Négative

FIGURE 4.4 – Exemples des expressions faciales : neutre, positive et négative

Nous avons testé notre approche en utilisant une webcam standard avec des conditions d'éclairage et de fond réelles. Nous avons utilisé cinq signaux physiologiques (la pression sanguine volumique (BVP), l'électromyographie (EMG), l'activité électrodermale (SC), la température cutanée (SKT) et la respiration (Resp)).

Les résultats obtenus dans le troisième chapitre, nous ont amené à étudier deux états émotionnels (positif et négatif). Vu l'effet de l'inducteur, ce n'est pas possible d'induire les six émotions. Pour cette raison, nous avons présenté 60 images durant 300 secondes (5 secondes/image) pour induire deux émotions avec valence positive et négative pour chaque sujet.

Le participant assis devant l'ordinateur doit être calme et relaxé.

Les signaux physiologiques ont été acquis en utilisant le système PROCOMP Infiniti [172]. Le taux d'échantillonnage a été fixé à 256 échantillons par seconde pour tous les canaux. Le traitement des données faciales et physiologiques s'effectue en parallèle à l'aide d'un système multitâche qui permet la synchronisation entre les deux modalités.

La durée d'une session d'expériences est d'environ 5 min. Les sujets ont été priés d'être aussi détendus que possible pendant cette période. Par la suite, un stimulus émotionnel a été appliqué pendant l'acquisition des deux modalités signaux physiologiques et expressions faciales.

Figure 4.5 montre le schéma de notre système bimodal avec les différents niveaux de fusion utilisés dans notre étude.

4.5 Implémentation et discussion

Nous avons utilisé les SVM pour la classification dans les systèmes uni-modaux et au niveau de la fusion des caractéristiques.

Les paramètres du SVM sont calculés en utilisant la méthode de la validation croisée qui

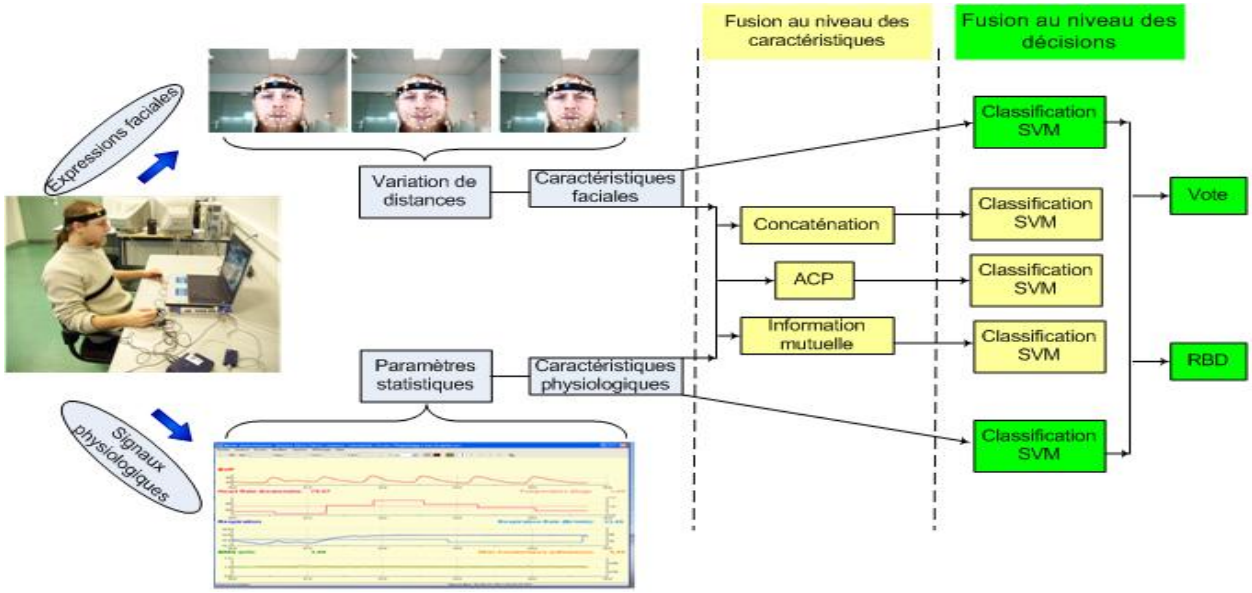


FIGURE 4.5 – Architecture de notre système bimodal avec différents niveaux de fusion

permet d'acquérir une certaine robustesse dans le choix des paramètres.

Nous avons procédé à la classification des données en considérant deux facteurs :

1. **Les données à classer :** selon sujets-dépendants ou sujets-indépendants :

   (a) Dans le cas **dépendant**, la reconnaissance des émotions est évaluée sur des corpus de test pour des sujets qui appartiennent au corpus d'apprentissage.

   (b) Dans le cas **indépendant**, les résultats de la reconnaissance des émotions sont évalués sur des données de test pour des sujets qui ne participent pas au corpus d'apprentissage.

2. **L'acquisition :** les données acquises pendant plusieurs jours permettent de construire un ensemble plus vaste par rapport à celles acquises pendant un seul jour. Pour voir l'impact de la durée des acquisitions nous avons construit deux bases de données :

   (a) Base1 : l'acquisition du corpus d'apprentissage et du corpus test est réalisée lors de sessions de 5 minutes répétées à divers moments pendant 4 jours en utilisant les mêmes images.

   (b) Base2 : l'acquisition du corpus d'apprentissage et du corpus test est réalisée le même jour.

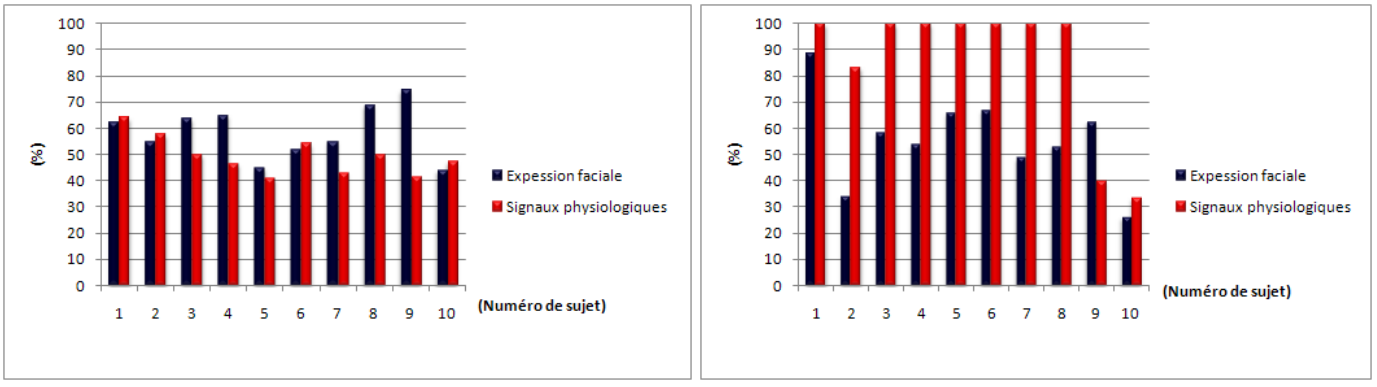
4.5.1 Résultats de la reconnaissance des émotions du système uni-modal

Les figures 4.6-a et 4.6-b représentent les taux de la classification des expressions faciales et des signaux physiologiques séparément pour la base 1 et la base 2.

Nous remarquons que les meilleurs taux de reconnaissance pour les sujets 1, 2, 6 et 10 de la base 1 sont obtenus en utilisant les signaux physiologiques. Par contre, les meilleurs taux de reconnaissance pour les sujets 3, 4, *5,* 7, 8 et 9 sont obtenus en utilisant les expressions faciales. Pour la base 2, les meilleurs résultats sont obtenus avec les signaux physiologiques pour tous les sujets sauf sujet 9.

Ces résultats ne permettent pas de généraliser la meilleure modalité à utiliser pour la reconnaissance des émotions, pour cela nous avons fusionné les deux modalités.

Il est raisonnable d'espérer que certaines caractéristiques des émotions peuvent être obtenues par l'utilisation de l'une des caractéristiques physiologiques ou faciales, par exemple (EMG des caractéristiques physiologiques, les sourcils des caractéristiques faciales). Cette information redondante est très précieuse pour améliorer la performance du système de reconnaissance des émotions lorsque les caractéristiques de l'une des deux modalités sont perdues ou de mauvaise qualité. Par exemple dans le cas d'une personne avec des cheveux sur le front ou des lunettes, ses expressions du visage seront extraites avec un niveau d'erreur élevé. Dans ce cas, les caractéristiques physiologiques (EMG) peuvent être utilisées pour surmonter la limitation de l'information visuelle.



a : Acquisition durant 4 jours b : Acquisition pendant un seul jour

FIGURE 4.6 – Résultats de la reconnaissance uni-modale des émotions

4.5.2 Résultats de la reconnaissance des émotions du système bimodal

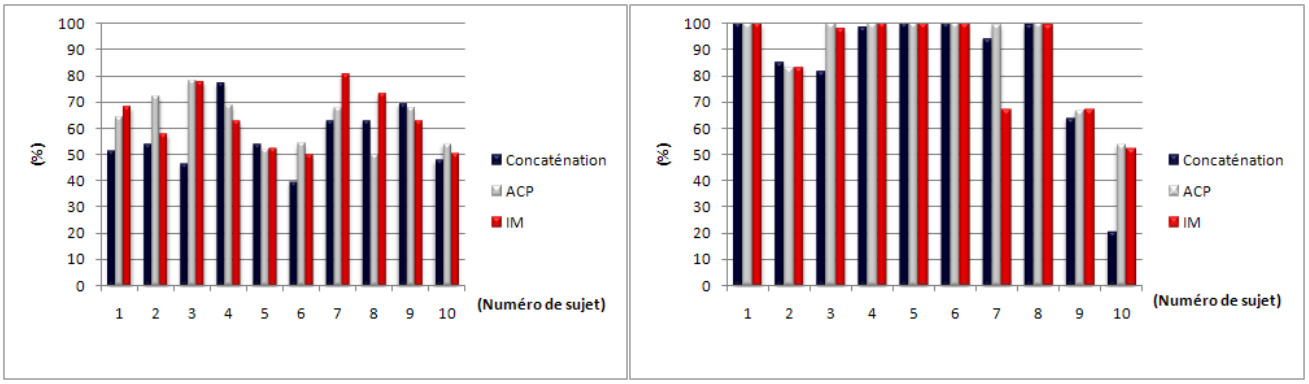
Notre système bimodal est basé sur la fusion des données faciales et les données physiolo-

giques. Chaque modalité a ses avantages et ses inconvénients, cependant quelques inconvénients peuvent être surmontés en utilisant les deux modalités.

Dans cette section, nous présentons les résultats expérimentaux des différentes méthodes utilisées pour la fusion des caractéristiques et de décisions.

4.5.2.1 Fusion des caractéristiques

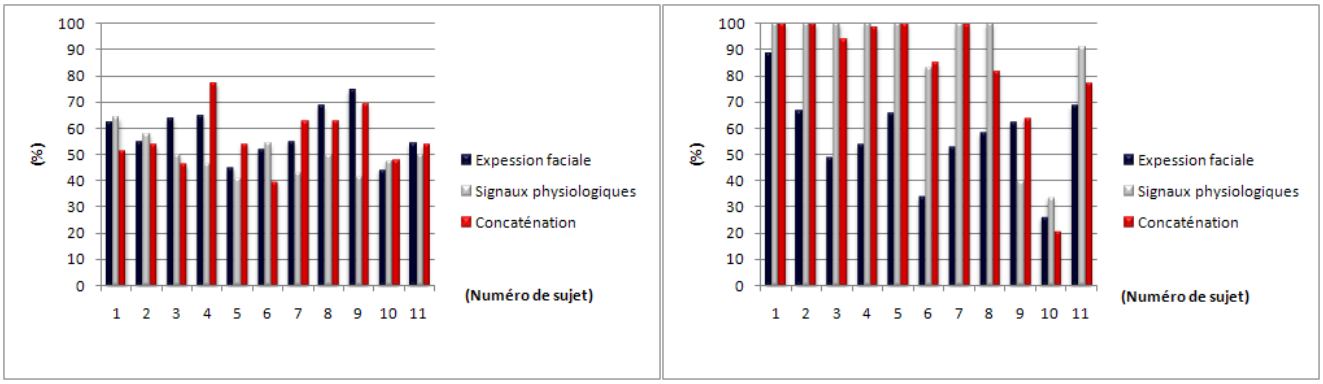
La fusion des caractéristiques pour la reconnaissance des émotions est une phase primordiale qui permet de sélectionner un ensemble pertinent répondant à la cohérence entre les deux modalités et assurant une homogénéisation entre les données. Les taux de la classification des méthodes de fusions de caractéristiques utilisées sont présentés dans les figures 4.7-a et 4.7-b pour la base 1 et la base 2 relativement.



a : Acquisition durant 4 jours b : Acquisition pendant un seul jour

FIGURE 4.7 – Résultats de la reconnaissance des émotions avec la fusion des caractéristiques

Résultats de la concaténation (Con)



a : Acquisition durant 4 jours b : Acquisition pendant un seul jour

FIGURE 4.8 – Comparaison entre les résultats des systèmes unimodaux et la concaténation

La concaténation des données est une méthode simple. On prend toutes les caractéristiques des expressions faciales et des signaux physiologiques sans aucun tri et aucune transformation. D'après les résultats obtenus (figure 4.8), les taux de reconnaissance de la concaténation sont inférieurs par rapport aux résultats de chaque modalité pour les deux bases.

La concaténation des caractéristiques des deux modalités a abaissé les performances du système.

Résultats de l'information mutuelle (IM)

Parmi les 21 caractéristiques du visage et les 30 caractéristiques physiologiques, nous souhaitons identifier celles qui contribuent le plus dans la classification des émotions. Nous appliquons l'information mutuelle IM pour identifier le sous-ensemble le plus important qui permet de faire la distinction entre les différentes émotions.

L'algorithme de l'IM trie les éléments selon leur pertinence. Les résultats de l'algorithme de sélection suggèrent que les six caractéristiques importantes pour la base 1 sont :

1. La dérivée première normalisée du signal BVP ;
2. La dérivée seconde normalisée du signal BVP ;
3. La dérivée première normalisée du signal EMG ;
4. La dérivée seconde normalisée du signal EMG ;
5. La dérivée première normalisée du signal SC ;
6. La dérivée seconde normalisée du signal SC.

En se basant sur les résultats de la sélection des caractéristiques bimodales, nous avons choisi les six premières caractéristiques. Concernant la base 2, la méthode de l'IM a donnée une sélection différente pour chaque personne (voir annexe B).

Globalement, d'après les résultats obtenus (figure 4.7), les performances obtenues avec l'IM sont meilleures par rapport aux résultats obtenus avec les systèmes uni-modaux (figure 4.6) et par rapport à la concaténation. Cette amélioration est due à la pertinence des caractéristiques sélectionnées avec cette méthode.

Résultats de l'ACP

Nous avons appliqué l'ACP pour trouver une transformation des 51 caractéristiques dans un autre espace plus réduit. Les résultats obtenus avec la base 2 sont meilleurs par rapport à la base 1 dû au fait que les données acquises durant plusieurs jours ont plus de variation par rapport aux données acquises durant un seul jour (figure 4.11 et 4.12). Globalement, l'ACP a amélioré les performances du système par rapport à l'IM (figure 4.7).

Les tableaux 4.1, 4.2 montrent le taux de reconnaissance des émotions par rapport au nombre de caractéristiques (ou composantes pour ACP) pour la base 1. Nous ne pouvons pas généraliser une conclusion en fonction du nombre de composantes ou le nombre de caractéristiques qui donnent les meilleurs résultats. Par exemple, le troisième sujet a de bons résultats avec 5 composantes (5 caractéristiques) pour l'ACP et l'IM. Pour le cas de tous les sujets, le meilleur résultat est obtenu avec 6 composantes de l'ACP.

Sujet	5 caract (%)	10 caract (%)	12 caract (%)	15 caract (%)	20 caract (%)
Sujet 1	51.5	51.5	60.61	60.61	68.21
Sujet 2	50.09	50.09	57.8	57.8	38.62
Sujet 3	77.71	77.71	54.28	73.98	62.86
Sujet 4	50.31	50.31	54.28	50.29	62.59
Sujet 5	52.39	52.39	40.95	40.95	43.12
Sujet 6	49.92	49.92	49.92	49.92	21.58
Sujet 7	50.49	50.49	36.35	80.68	56.74
Sujet 8	49.46	49.46	49.8	53.39	73.46
Sujet 9	49.93	49.93	41.39	62.79	59.81
Sujet 10	47.31	47.31	47.32	47.32	50.32
Tous les sujets	50.22	50.22	48.13	50.89	52.81

TABLE 4.1 – Résultats de la fusion avec l'information mutuelle

Sujet	2 comp	3 comp	4 comp	5 comp	6 comp	7 comp	ACP à 100%
Sujet 1	58 %	64.1 %	55.82 %	56.1 %	56.28 %	54.95 %	54.95 %
Sujet 2	57.8 %	57.7 %	50.68 %	50.09 %	72.1 %	49.77 %	49.77 %
Sujet 3	49.95 %	49.95 %	76.57 %	78.3 %	49.77 %	48.75 %	48.75 %
Sujet 4	46.25 %	50.06 %	45.28 %	42.11 %	51.04 %	68.82 %	68.82 %
Sujet 5	40.95 %	40.94 %	43.87 %	40.58 %	51.03 %	51.71 %	51.71 %
Sujet 6	46.07 %	49.95 %	54.15 %	54.13 %	54.12 %	54.13 %	39.22 %
Sujet 7	48,64 %	42.48 %	66.63 %	67.96 %	64.83 %	57.38 %	48.6 %
Sujet 8	49,8 %	48.08 %	48.6 %	41.42 %	45.19 %	36.71 %	32.24 %
Sujet 9	42,23 %	40.44 %	59.85 %	65.11 %	66 %	67.79 %	67.79 %
Sujet 10	47,31 %	47.31 %	47.31 %	47.31 %	54.1 %	31.14 %	52.08 %
Tous les sujets	49,42 %	49 %	50.03 %	49.80 %	55.22 %	49.93 %	52.91 %

TABLE 4.2 – Résultats de la fusion avec ACP

4.5.2.2 Fusion des décisions

La fusion de décisions vise à déterminer l'émotion la plus probable du sujet compte tenu de l'émotion déterminée dans un intervalle de temps précis. Une fenêtre de données contient les décisions en cours et les précédentes pour l'analyse. Dans ce travail, nous avons pris un intervalle (fenêtre) $N = 15$, parce que nous avons 15 images par seconde pour l'expression faciale et nous avons besoin d'une décision par seconde.

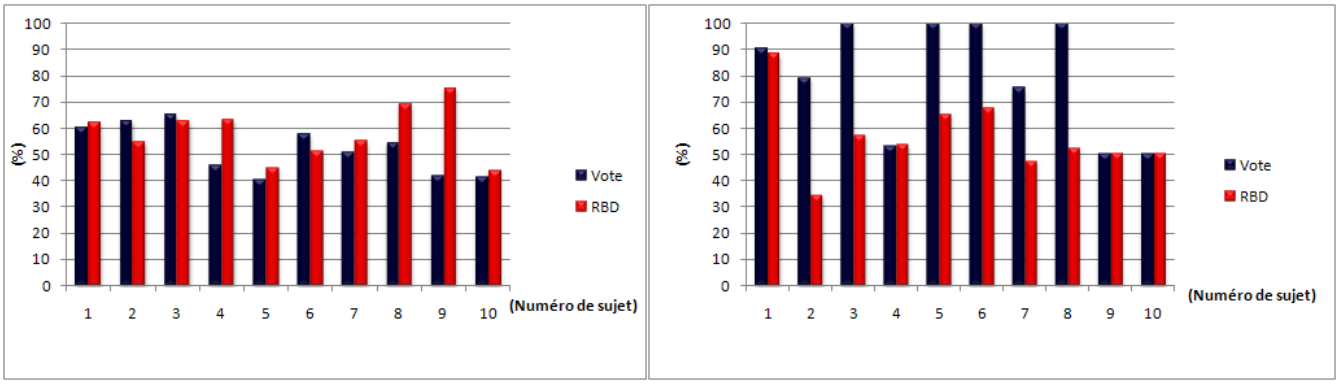
Les taux de la classification des méthodes de fusions de décisions utilisées sont présentés dans les figures 4.9-a et 4.9-b pour la base 1 et la base 2 relativement.

Résultats de la méthode de vote

Le processus de vote utilisé pour la fusion de décisions est une méthode simple qui délivre un résultat par seconde sans aucune pondération des deux modalités. Les résultats obtenus avec la deuxième base de données sont bien meilleurs par rapport à la première dû à la qualité des décisions unimodales (figure 4.9).

Résultats du RBD

La figure 4.10 illustre la structure de notre réseau Bayésien dynamique utilisée pour la fusion



a : Acquisition durant 4 jours b : Acquisition pendant un seul jour

FIGURE 4.9 – Résultats de la reconnaissance des émotions avec la fusion des décisions

de décisions pour la reconnaissance des émotions. Le réseau utilisé est constitué de trois nœuds, deux pour les observations (la décision de l'expression faciale et la décision des signaux physiologique) et un troisième pour la prise de décision. L'estimation des probabilités conditionnelles est basée sur l'apprentissage des paramètres avec l'algorithme EM (*Expectation-Maximisation*) [151]. L'algorithme de l'inférence utilisé est basé sur l'arbre de jonction détaillé dans [150]. L'implémentation des RBD est réalisée avec la librairie PNL (*Probabilistic Network Library*) [7].

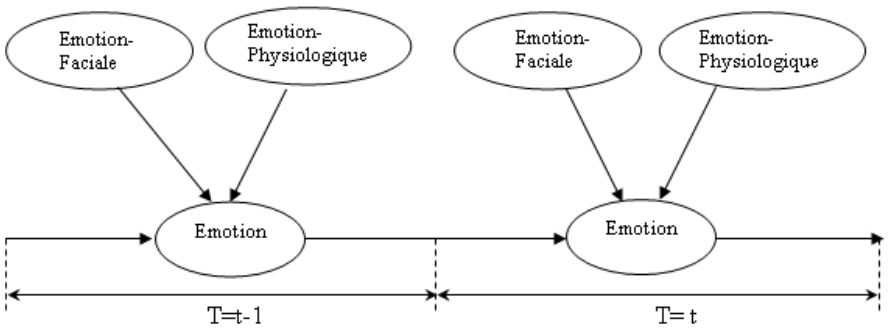


FIGURE 4.10 – Représentation compacte du RBD utilisé pour la reconnaissance bimodale des émotions

Les résultats obtenus avec la base 1 sont bien meilleurs que ceux obtenus avec la base 2, parce que les données acquises durant plusieurs jours permettent de collecter plus d'informations sur les décisions des deux modalités, autrement dit, la variation temporelle ou dynamique des décisions unimodales.

En comparant entre les deux méthodes de fusion des décisions, on remarque que les RBD ont donné des résultats meilleurs par rapport au vote pour la base 1, contrairement à la base 2 où les meilleurs résultats sont obtenus avec le vote. Cela peut être expliqué par le fait que la méthode de vote ne considère que les bonnes décisions de chaque classifieur dans un intervalle de temps, tandis que les RBD prennent en considération l'erreur de chaque classieur [116].

	positif	négatif
positif	78.9%	21.1%
négatif	21.1%	78.9%

(a)

	positif	négatif
positif	11.6%	88.4%
négatif	4.2%	95.8

(b)

TABLE 4.3 – Matrice de confusion pour Base 1 avec l'ACP (5 composantes) : a- cas dépendant, b-cas Indépendant

4.5.3 Matrice de confusion

Les tableaux 4.3-a et 4.3-b représentent les matrices de confusion pour sujet 3 (cas dépendant) et les 10 sujets (cas indépendant) respectivement pour la base 1.

Pour une seule personne (cas dépendant), la distinction entre les classes est très bonne avec un taux qui dépasse 78%, mais dans le cas de tous les sujets (indépendant) il y a une mauvaise distinction entre les classes, nous pouvons expliquer cela par la domination d'une seule émotion sur le participant.

4.5.4 L'effet de l'acquisition pendant un seul jour sur l'état émotionnel

Les données acquises pendant plusieurs jours permettent de construire un ensemble plus vaste par rapport à celles acquises pendant un seul jour. Les données sont potentiellement utiles pour l'analyse de l'état émotionnel.

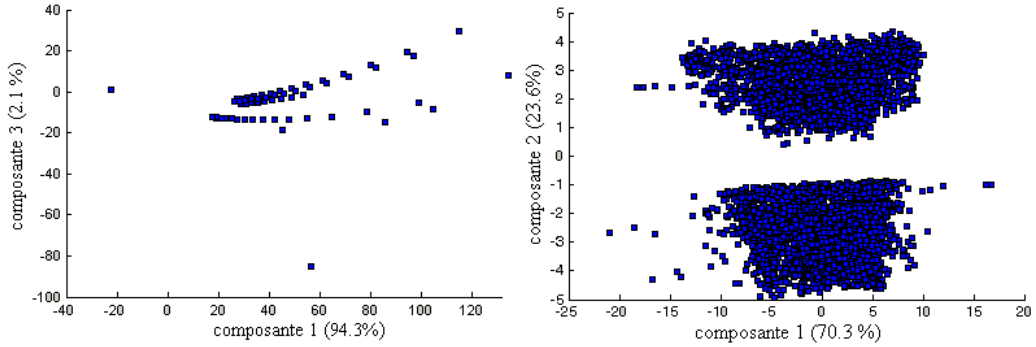
Pour voir l'impact de la durée des acquisitions, nous avons construit les deux bases : une pendant un seul jour et l'autre durant plusieurs jours.

Les figures 4.11 et 4.12 montrent la projection des données dans l'espace de l'ACP, les données acquises durant plusieurs jours (les figures 4.11-a et 4.12-a) présentent une dispersion par rapport aux données acquises pendant un seul jour (les figures 4.11-b et 4.12-b). La comparaison entre les résultats obtenus avec les deux bases permet de voir l'effet de la périodicité de l'acquisition des mesures sur la reconnaissance de l'état émotionnel. Cette variation est due au fait que les réactions des participants dépendent de leurs humeurs, de leurs états d'esprit le jour de l'acquisition et de l'effet de l'accoutumance [78, 108]. Nous remarquons que plus nous augmentons l'intervalle de temps plus il est difficile de distinguer entre les émotions (figure 4.11 et 4.12).

La dépendance au jour de l'acquisition est liée probablement aux variations de la physiologie qui est due à deux facteurs [169] :

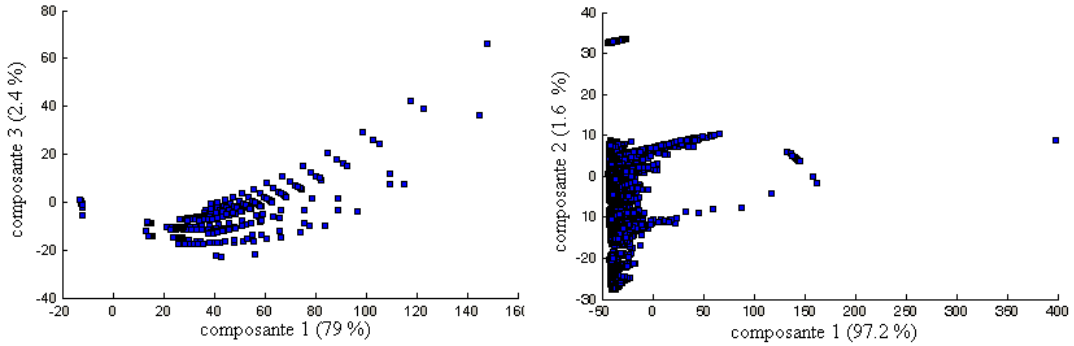
1. Le sommeil, les hormones et autres facteurs non-émotionnels,
2. L'humeur ou une incapacité à construire une intense expérience de la joie, si le sujet a senti une forte tristesse ce jour-là.

Plusieurs de ces sources de variation sont naturelles et ne peuvent pas être contrôlées dans la réalité à long terme.



a : Acquisition durant 4 jours b : Acquisition pendant un seul jour

FIGURE 4.11 – La projection des données dans l'espace de l'ACP pour 1 seul sujet



a : Acquisition durant 4 jours b : Acquisition pendant un seul jour

FIGURE 4.12 – La projection des données dans l'espace de l'ACP pour tous les sujets

4.5.5 Résultats de l'auto-évaluation

Les images IAPS utilisées comme inducteur dans notre système ont un pré-classement selon les travaux de Lang et al. [127] qui ont attribué à chaque image une valence et une arousal. Une étape d'auto évaluation a été demandée à tous les sujets pour avoir un meilleur classement des images IAPS. Nous avons recalculé les taux de la reconnaissance de la base 1 avec le nouveau classement des images.

Les figure 4.13 et 4.14 montrent les résultats dans le cas dépendant avec toutes les méthodes citées précédemment. Globalement, les taux obtenus sont meilleurs par rapport aux résultats du premier classement en comparant entre les deux, les résultats de l'auto-évaluation sont bien meilleurs par rapport à ceux du classement IAPS surtout pour le sujet 5.

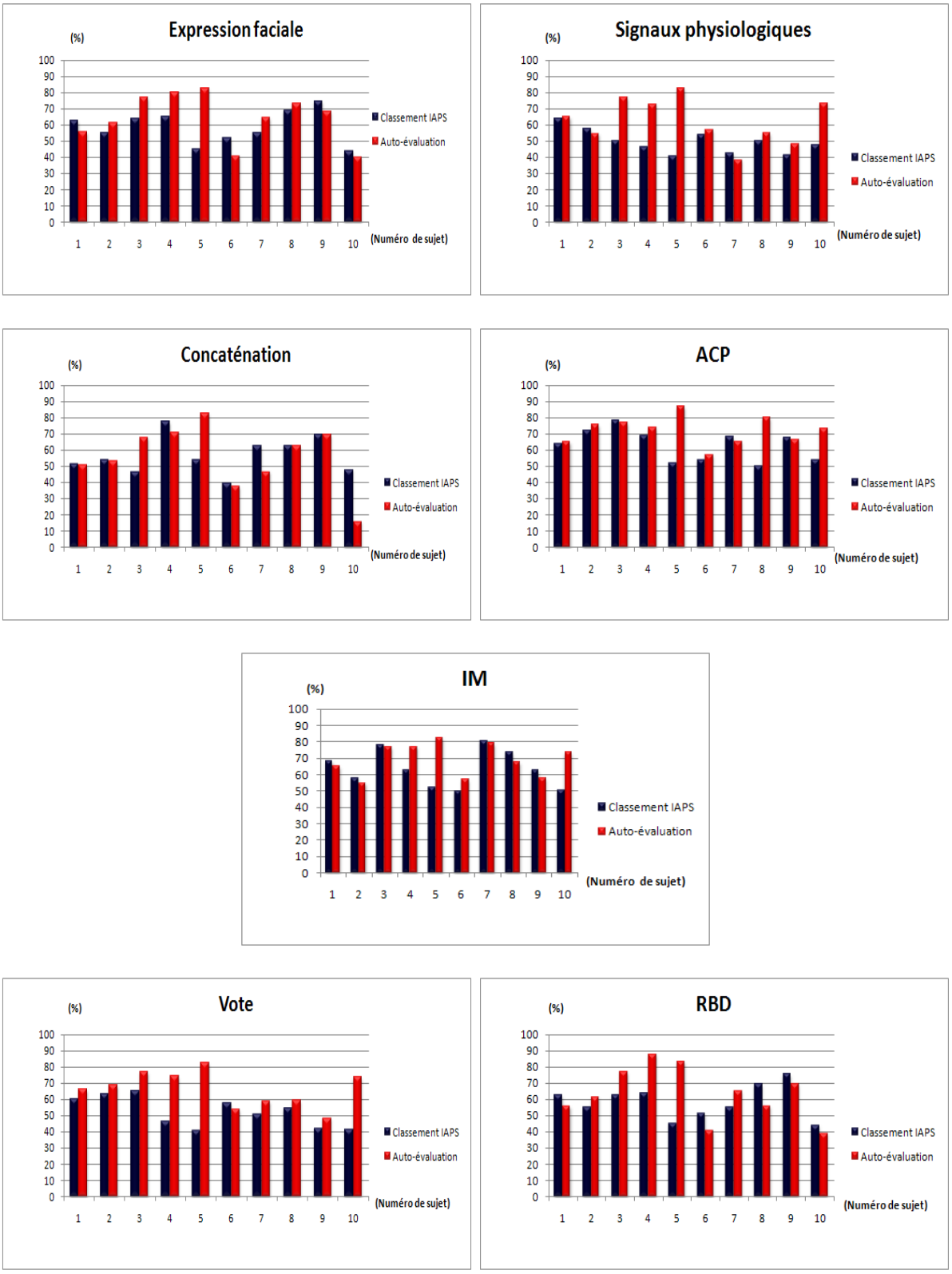


FIGURE 4.13 – Comparaison entre les résultats de l'auto-évaluation et le classement IAPS pour chaque méthode

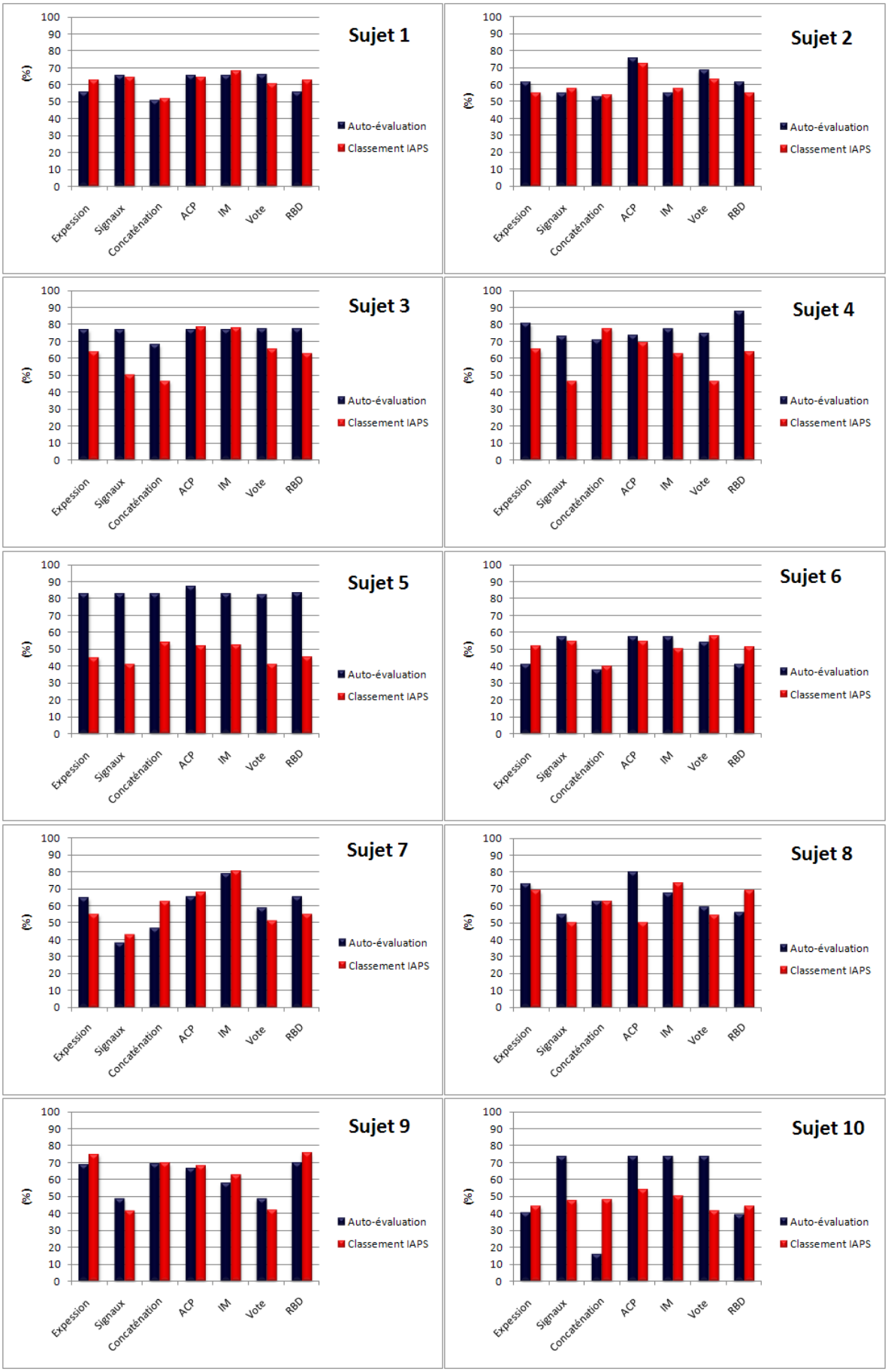


FIGURE 4.14 – Comparaison entre les résultats de l'auto-évaluation et le classement IAPS pour chaque sujet

4.5.6 La différence entre une base individuelle et une base globale

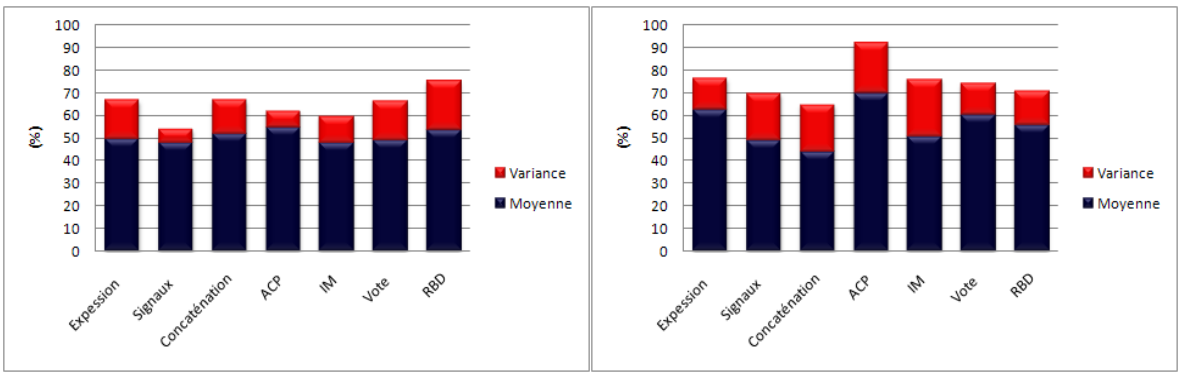
Tous les résultats présentés dans ce chapitre montrent la supériorité des modèles individuels par rapport à un modèle global. En général, le sujet a son propre contrôle. La figure 4.15 montre les résultats de la reconnaissance dans le cas indépendant pour les deux bases où l'ACP a donné les meilleurs performances en termes de moyenne (entre les sujets).

Il existe plusieurs raisons de se concentrer sur la reconnaissance des émotions dans le cas dépendant. Ekman et ses collègues reconnaissent qu'un étiquetage simple des émotions comme la joie et la colère peut avoir des interprétations différentes entre les individus de la même culture, ce qui complique la recherche pour voir si les personnes suscitent des modèles physiologiques semblables de la même émotion [169].

Pour certaines applications de reconnaissance de la voix à utilisation personnelle, le système nécessite une étape d'apprentissage des caractéristiques de l'utilisateur et pas certaines réponses moyennes formées a partir d'un groupe qui ne peuvent pas s'appliquer à l'individu correctement [169].

Lorsque les données de nombreux sujets ont été traitées, il était difficile de trouver des modèles physiologiques importants parce que la physiologie peut varier avec la façon dont chacun interprète chaque émotion. Travailler avec un panel de personnes là où chacun a ses propres attitudes rend l'étude plus difficile.

La figure 4.16 montre les résultats de la reconnaissance dans le cas dépendant du modèle global pour les deux bases. On remarque que, quelque soit la méthode utilisée, les résultats obtenus avec la base 2 sont meilleurs par rapport à la base 1.



a : Acquisition durant 4 jours b : Acquisition pendant un seul jour

FIGURE 4.15 – Résultats de la reconnaissance des émotions du cas indépendant

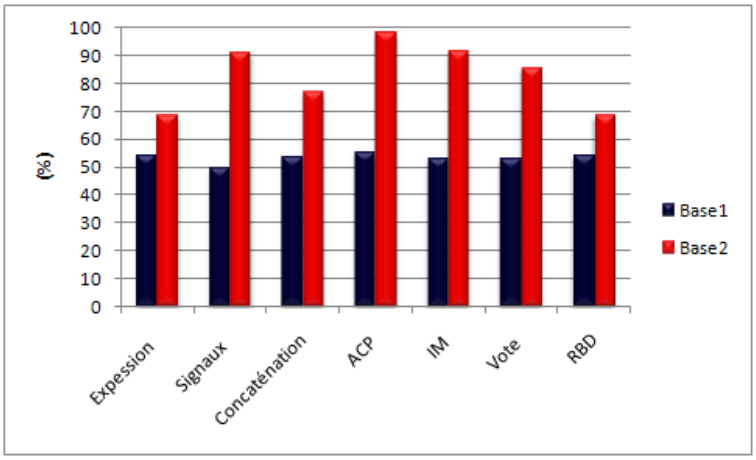


FIGURE 4.16 – Résultats de la reconnaissance des émotions du modèle global (cas dépendant)

4.6 Conclusion

Dans ce chapitre, nous avons présenté un système bimodal pour la reconnaissance des émotions à partir des expressions faciales et des signaux physiologiques.

Le vecteur de variation des distances par rapport à l'état neutre a été utilisé comme un descripteur de l'expression faciale et 30 valeurs statistiques pour les signaux physiologiques.

La reconnaissance des émotions à partir des expressions faciales ou à partir des signaux physiologiques donne des performances proches.

Pour améliorer les résultats du système uni-modal, nous avons utilisé un système bimodal. Pour la fusion au niveau des caractéristiques, nous avons utilisé la concaténation des caractéristiques des deux modalités qui n'a pas amélioré les résultats des deux bases utilisées. Pour une meilleure fusion des caractéristiques, nous avons utilisé l'ACP afin de transformer les caractéristiques et l'information mutuelle pour sélectionner le meilleur sous-ensemble. L'analyse en composantes principales a donné de meilleurs résultats par rapport à une sélection avec l'information mutuelle car la sélection des caractéristiques néglige certains paramètres comme les expressions faciales, par contre l'ACP réduit la dimensionnalité en faisant une combinaison linéaire de toutes les données initiales.

Pour la fusion au niveau de la décision, nous avons testé deux méthodes : le processus de vote et un réseau Bayésien dynamique. Globalement, nous avons montré que la fusion au niveau des caractéristiques et la fusion au niveau de décision des deux modalités donnent des améliorations significatives pour la reconnaissance des émotions.

Les résultats obtenus en fusionnant les deux modalités sont meilleurs par rapport à l'utilisation séparée de chaque modalité. Le traitement séparé des deux modalités et la combinaison des différentes décisions donne moins de performances. Notre conclusion confirme l'expérience de Chen Hung [40].

Le facteur principal qui a influencé sur les résultats de la reconnaissance des émotions est bien l'induction. Les participants ont constaté que l'utilisation des images IAPS n'est pas un bon inducteur. L'accoutumance est plus rapide lorsqu'un stimulus est présenté fréquemment. Elle a ses effets les plus marquants vers la fin de toute étude et doit être prise en considération dans l'évaluation de toutes les données acquises lors d'une étude psychophysiologique [169].

Il est nécessaire de trouver un autre inducteur afin de générer des données réelles pour le traitement. L'évocation d'un état émotionnel peut être réalisée en utilisant les 3 systèmes sensoriels principaux : visuel, auditif et kinesthésique pour augmenter le sentiment de présence. Richard Bandler dans son livre « Un cerveau pour changer » [21] nous montre l'impact des 3 sous-modalités sur nos croyances et sur nos états internes. En modifiant la structure de notre expérience, nous modifions nos réactions émotionnelles. Le tableau 4.4 montre quelques sous-modalités des trois systèmes sensoriels.

Sous-modalités visuelles	Sous-modalités auditives	Sous-modalités kinesthésiques
Taille/ format	Volume	Interne/ externe
Associé/ dissocié	Tonalité	Localisation
Film/ diapositive	Mono/ stéreo	Intensité
Intensité	Provenance interne ou externe	Durée
Luminosité	Rythme	Fréquence
Cadre/ pas de cadre	Durée	Température
Couleurs/ noir et blanc	Continu/ discontinu	Tension
Netteté/ flou	Éloignement/ proximité	Texture (toucher)
Distance (au sujet)	Vitesse	Mouvement
Son/ muet		Poids

TABLE 4.4 – Les sous-modalités des 3 systèmes sensoriels [52]

Conclusion et perspectives

I. Conclusion

Notre travail visait à concevoir une application de reconnaissance des émotions. Trois volets ont été traités : la reconnaissance des émotions à partir des expressions faciales, des signaux physiologiques et d'un système bimodal basé sur les deux premières modalités.

Dans le cadre de la reconnaissance des émotions à partir des expressions faciales, nous nous sommes intéressés aux déformations du visage qui caractérisent chacune d'elles. La première étape de ce système a été conçue dans le but de détecter le visage sans contrôler l'environnement utilisé. Pour une localisation grossière, nous avons déterminé les axes principaux des caractéristiques faciales en se basant sur une analyse bas niveau. Pour une localisation plus précise, nous avons proposé un modèle anthropométrique qui permet la localisation des points faciaux. Afin d'améliorer la précision, nous avons utilisé la technique de Shi-Tomasi qui permet d'avoir des points sillons qui assurent la stabilité de l'étape de suivi basée sur le flux optique. Chaque muscle facial est représenté par une distance entre deux points. La distinction entre les différentes expressions faciales est basée sur la variation de ces distances par rapport à l'état neutre. Les séparateurs à vastes marges ont été choisis pour la classification des données et pour valider notre approche, nous avons utilisé deux bases de données des expressions actées qui ont aboutit à de bons résultats.

L'utilisation des expressions faciales permet d'avoir une vision externe des émotions. Afin de prendre en considération le facteur interne du mécanisme émotionnel, un deuxième volet a porté sur la présentation d'une approche de reconnaissance des émotions à partir des signaux physiologiques. Nous avons utilisé la conductance de la peau, le volume sanguin périphérique, le volume respiratoire, le signal électromyographie et la température cutanée pour la reconnaissance de l'état émotionnel et nous avons établi leurs relations avec les processus émotionnels. Afin d'avoir une représentation adéquate des signaux, nous avons calculé un ensemble de paramètres statistiques classés en utilisant les SVM. Pour valider l'approche, nous avons enregistré une base de données dans notre laboratoire dans des conditions réelles avec une induction reposant sur les images IAPS. Les signaux physiologiques sont très intéressants pour le traitement des émotions mais les capteurs sont intrusifs et peuvent perturber les utilisateurs.

Nous avons fusionné les deux modalités étudiées afin de privilégier le caractère de robustesse des signaux physiologiques et la simplicité de l'acquisition des expressions faciales, en testant plusieurs méthodes. Au niveau de la fusion des caractéristiques, l'analyse en composantes principales a été utilisée et a donné de meilleurs résultats que la sélection par l'information mutuelle, ainsi qu'aux méthodes de fusion de décisions basées sur le vote ou sur les réseaux Bayésiens. Pour valider les résultats du système bimodal, une deuxième base de données a été réalisée sur quatre jours pour obtenir la variation réelle de l'émotion de chaque sujet. La qualité des résultats obtenus est liée à l'inducteur utilisé. Selon les participants, l'induction des émotions avec les images IAPS n'est pas suffisante. Le deuxième facteur qui a influé les résultats est le délai entre les tests et le corpus d'apprentissage. Ce dernier a été conçu pendant trois jours et le test à partir d'un quatrième. L'étape de l'auto-évaluation nous a montré l'effet de l'annotation des images et notamment l'efficacité des méthodes présentées dans ce mémoire.

Durant cette étude, nous nous sommes basés sur le principe qu'une émotion peut être décrite comme un processus multidimensionnel déterminé par deux composantes, physiologique et comportementale qui varient d'une personne à une autre et qui rendent l'étude difficile.

L'implication des émotions dans le comportement humain va plus loin et concerne la perception, la cognition et l'action. Les émotions sont un moteur qui guide et oriente les processus mentaux qui interviennent au niveau de la construction de la pensée. Au niveau perceptif, elles sélectionnent les informations dans l'environnement, au niveau cognitif, elles les traitent par les inférences et les raisonnements et au niveau comportemental, elles prennent les décisions pour les actions. Le dernier niveau nous intéresse plus particulièrement car les émotions ont un effet très important dans la vie sociale. Elles interviennent directement par des tendances à l'action involontaire et automatique. Dans le cas d'une personne phobique, une émotion peut produire une situation de peur irrationnelle et générer des comportements de fuite inappropriés. En effet, le contrôle d'une situation en mesurant l'état émotionnel est crucial pour une application de thérapie de la phobie sociale basée sur la réalité virtuelle.

Dans une application thérapeutique, il est nécessaire de provoquer une émotion afin d'avoir des réactions qui permettent une acquisition de données réelles correspondant à des classes bien déterminées d'émotion. Une méthodologie qui s'appuie sur des savoir-faire issus de disciplines connexes telles que l'informatique, la psychologie et la linguistique assure l'extraction des informations utiles.

II. Perspectives

Un certain nombre de problèmes restent ouverts. Nous avons pu dégager quelques perspectives.

II.1 Expressions faciales

Dans notre application, la personne est face à la caméra, les points faciaux qui sont détectés à partir d'une vue de face décrivent la forme des caractéristiques (bouche, sourcil et yeux) et du contour du visage. L'ensemble de ces points a une disposition uniforme qui peut être déformée légèrement d'une expression à une autre à cause des mimiques faciales. Dans certaines situations, il est possible que la personne change de position ce qui perturbe le suivi des points faciaux et déforme complètement la disposition des points. La vérification de la disposition peut être effectuée en faisant une comparaison entre la localisation actuelle et l'initiale. L'utilisation d'un critère de dispersion des points permet d'avoir toujours une disposition admissible au niveau de la localisation. Selon la valeur d'un critère choisi, on peut proposer une étape d'initialisation. On peut citer le test du CHI2 qui permet de rejeter une hypothèse de départ en fonction de la distance jugée excessive entre deux ensembles d'informations. Pour une application en temps réel, on peut effectuer la comparaison que 5 fois/seconde par exemple au lieu de 15/seconde.

Notre système de reconnaissance des expressions faciales exige une vue de face pour une meilleure localisation. Il n'est pas évident de demander à la personne de maintenir une telle position sur une longue durée au risque de la fatiguer et de la perturber. Afin de donner plus de tolérance à notre système, l'utilisation de plusieurs caméras permet le contrôle avec différentes poses. Par exemple, pour une application de thérapie de la phobie sociale, la personne risque de réagir en fuyant la situation actuelle, en tournant la tête pour ne pas regarder l'image qui l'a excité. On peut proposer à ce niveau d'utiliser la méthode d'analyse de mouvement afin de résoudre ce problème.

L'utilisation d'une base de données des expressions réelles est un autre point très important, dans une étude sur la reconnaissance des émotions à partir des expressions faciales. La plupart des corpus des expressions réelles existants présentent des émotions de la vie de tous les jours [232], plus modérées et plus contenues que les émotions recherchées pour des applications bien spécifiques comme la thérapie de la phobie sociale. Ceci est dû au fait que les émotions intenses spontanées sont plus rares et plus imprévisibles que les émotions de la vie de tous les jours. Certains corpus spontanés, cependant, fournissent un contenu émotionnel plus intense avec des émotions de type peur [48].

II.2 Signaux physiologiques

Les signaux physiologiques sont la modalité la plus robuste pour la reconnaissance des émotions mais elle présente un inconvénient principal qui réside dans l'intrusion du système d'acquisition. Le port des capteurs et la liaison avec l'ordinateur via des fils peuvent stresser l'utilisateur ce qui peut influencer les résultats. A ce stade, on peut proposer des perspectives afin de résoudre ce problème :

1. L'utilisation des réseaux de capteurs sans fil permet théoriquement une surveillance permanente des patients et une possibilité de collecter des informations de meilleure qualité ;

2. La miniaturisation est un élément majeur pour l'acceptation du capteur par l'utilisateur, celui-ci doit être très discret (l'intégration dans un vêtement étant une bonne option) et capable d'enregistrer ou d'envoyer les informations acquises à une fréquence de l'ordre de 20 Hz [23] ;
3. Une meilleure solution est l'utilisation des capteurs à distance comme, par exemple, une caméra infra rouge pour détecter la température.

D'un autre coté, au niveau du traitement de données, on propose d'analyser les signaux physiologiques par la transformée d'ondelettes qui décompose le signal à différentes bandes fréquentielles et à différentes résolutions. Le choix du type de l'ondelette ou de son ordre dépend du signal traité, par exemple, un signal à hautes fréquences (EMG) sera traité différemment d'un signal à basses fréquences (la conductance de la peau).

Dans un but de remédiation, le psychiatre demande au patient de réaliser des exercices chez lui. Les capteurs physiologiques ne sont pas à disposition de chacun ce qui les exclut dans une telle application. Par contre l'utilisation de l'image est facilement réalisable par l'utilisation de webcam.

II.3 L'induction

Au niveau de l'induction, nous aurions aimé procéder autrement au lieu de faire défiler les images d'une façon monotone, nous aurions voulu interrompre cette routine par l'envoi d'éléments perturbateurs inattendus comme le son d'une alarme afin de créer une situation de peur par exemple, et à ce moment là, recueillir les données, car l'émotion est caractérisée par une courte durée.

On peut également modifier le déroulement de la séance des mesures afin d'arriver à une situation où le sujet cesse de se rendre compte de son environnement. Il faut arriver à l'accompagner dans une intense concentration, vers une notion perturbée du temps et de la réalité. L'utilisation d'un simple écran d'ordinateur ne permet pas de plonger l'utilisateur dans un autre environnement pour lui inculquer des réflexes qui serviront ensuite dans un cas réel. On peut adapter l'environnement en jouant sur l'installation de la personne, son confort, sur le décor de la salle, sur l'éclairage, sur la musique, sur la taille de l'écran et sur la qualité des images utilisées.

II.4 Système multimodal

Nous constatons que les traits du visage et de la physiologie ne sont pas les seules informations pour la détection d'émotion. On peut avantageusement utiliser d'autres informations telles que la parole et les gestes.

En effet, la parole ne peut pas se ramener à un simple message nécessaire à l'action. Elle a d'autres fonctions :

1. Une est liée à la nature d'être social de l'homme qui permet d'instaurer un climat de convivialité garant d'une cohésion du groupe par l'échange d'informations concernant la personne elle-même, ses idées, ses goûts ou sa famille.
2. Une autre est de maintenir un niveau de vigilance donné.

L'ajout d'autres modalités comme la parole et les gestes peut enrichir la reconnaissance des émotions en analysant un comportement de tous les sens mais l'utilisation de la parole avec les expressions faciales peut poser un problème. L'expression par la parole a une influence sur les caractéristiques du visage. Dans ce cas, on peut surmonter ce problème en utilisant uniquement la partie supérieure du visage (yeux et sourcil) avec l'information acoustique.

II.5 Fusion hybride

Pendant ce travail, nous avons effectué une fusion au niveau des caractéristiques et au niveau de la décision. On peut proposer, à cette étape de traitement, un autre niveau de fusion, celui des décisions issues de l'étape de la fusion des caractéristiques (transformation ou bien sélection effectuées au niveau de chaque modalité) (figure 4.17).

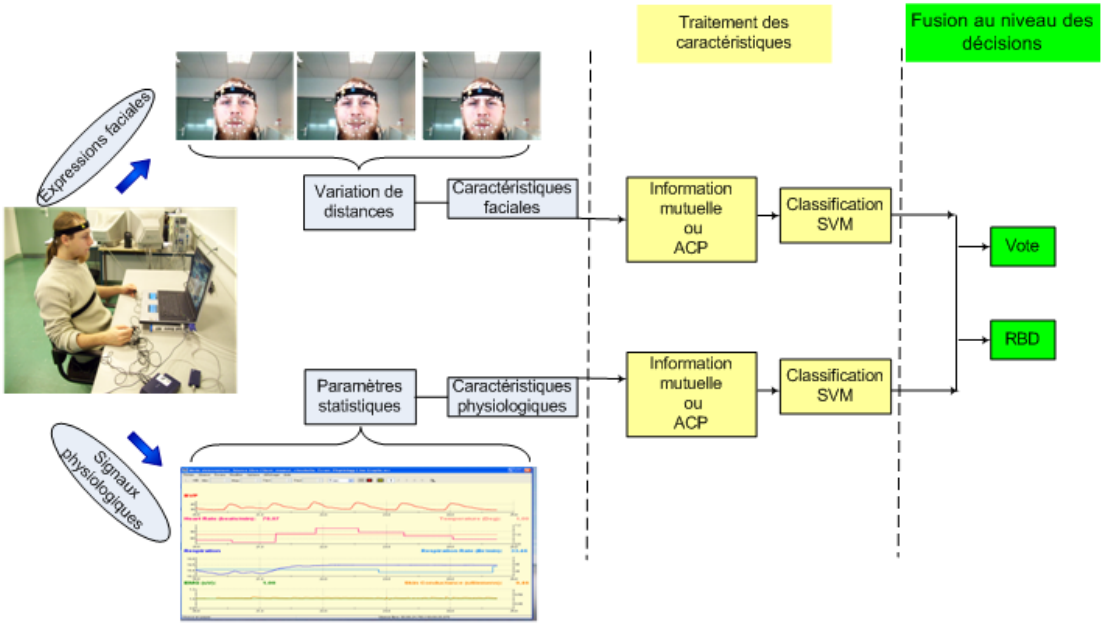


FIGURE 4.17 – Nouvelle architecture proposée

Au cours de cette thèse, nous n'avons pas eu la chance de faire des tests avec des personnes phobiques comme était prévu au départ. Le travail avec des sujets sains a rendu l'étude difficile, surtout pour généraliser des conclusions car l'effet de l'inducteur varie d'une personne à une autre, d'un jour à un autre et de l'environnement. Le travail réalisé peut faire l'objet de plusieurs applications et pas uniquement le traitement de la phobie sociale.

Annexe A

Séparateur à vastes marges

Les SVM constituent la forme la plus connue parmi les méthodes à noyaux, inspirées de la théorie statistique de l'apprentissage de Vladimir Vapnik [211]. Ce sont des classifieurs qui reposent sur deux idées clés :

1. La notion de marge maximale qui est la distance entre la frontière de séparation et les échantillons les plus proches, ces derniers sont appelés vecteurs supports.
2. La transformation de l'espace de représentation des données d'entrées en un espace de plus grande dimension $\mathbb{F}$, dans lequel il est probable qu'il existe une séparatrice linéaire, afin de pouvoir traiter des cas où les données ne sont pas linéairement séparables.

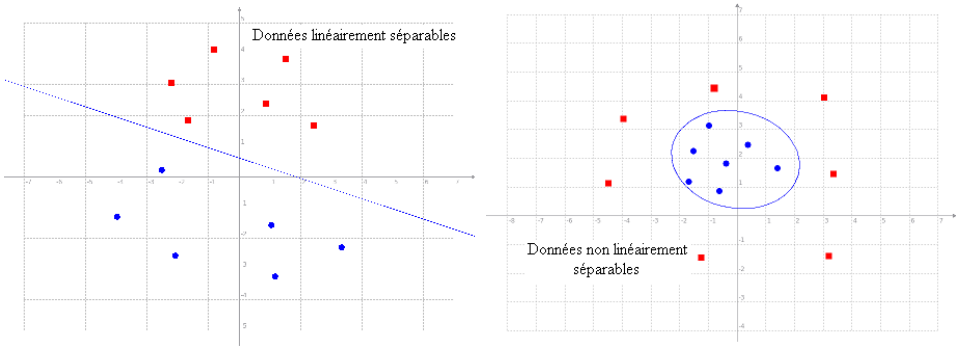


FIGURE A.1 – Séparateurs linéaires et non-linéaires

$\mathbb{F}$ peut être vu comme l'espace vectoriel généré par une famille de fonctions Φ_k. Sous certaines conditions très générales, il se trouve que la série $\Phi_k(x)\Phi_k(y)$ converge vers la fonction noyau de :

$$K(x,y) = \sum_{k\in\mathbb{N}} \Phi_k(x)\Phi_k(y) \tag{A.1}$$

Dans ce cas, la fonction de décision est donnée par le signe de la fonction de discrimination suivante qui ne dépend plus que du noyau K :

$$f(x) = \sum_{i=1}^{\ell} y_i \alpha_i K(x, x_i) + b \tag{A.2}$$

Tel que l'ensemble d'apprentissage est présenté par $\{x_i, y_i\}_{i=1...\ell}$ où $x_i \in \Re^n$ et $x_i \in \{-1, 1\}$. où les α_i et b sont des coefficients à déterminer, en maximisant la distance, appelée marge, entre la frontière de décision $f(x) = 0$ et le nuage de points dans $\mathbb{F}$. Le problème à résoudre s'écrit alors :

$$max_{\alpha_i} \sum_{i=1}^{\ell} \alpha_i - \frac{1}{2} \sum_{j=1}^{\ell} \alpha_i \alpha_j y_i y_j K(x_i, x_j) \tag{A.3}$$

avec $0 < \alpha_i < C, i = 1, \ell$ et $\sum \alpha_i y_i = 0$

Où C est un paramètre qui permet de régler le taux d'erreur admissible dans la solution.

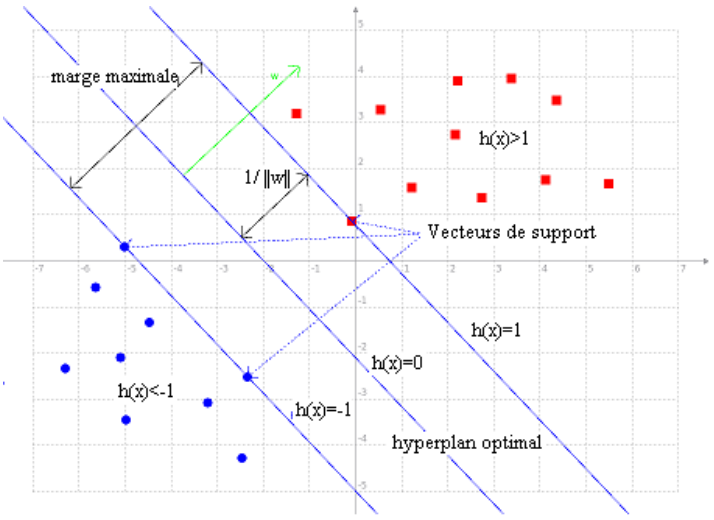


FIGURE A.2 – Illustration de la recherche de l'hyperplan optimal

Annexe B

Résultats du filtrage des signaux physiologiques

Le signal	B	A	Ordre du filtre
EMG	[0,32977 0 -1,9786 0 4,9466 0 -6,5954 0 4,9466 0 -1,9786 0 0,32977]	[1 1,618 -2,5831 -4,1824 3,8631 5,1175 -3,6484 -3,3974 2,1755 1,1967 -0,73937 -0,17595 0,10875]	6
SKC	[1,1628e-008 6,9768e-008 1,7442e-007 2,3256e-007 1,7442e-007 6,9768e-008 1,1628e-008]	[1 -5,6207 13,175 -16,484 11,61 -4,3647 0,6842]	6
VR	[2,1878e-006 1,3127e-005 3,2817e-005 4,3756e-005 3,2817e-005 1,3127e-005 2,1878e-006]	[1 -5,0522 10,7 -12,151 7,8013 -2,6834 0,38623]	6
SKT	[3,2583e-012 1,955e-011 4,8875e-011 6,5167e-011 4,8875e-011 1,955e-011 3,2583e-012]	[1 -5,9052 14,53 -19,07 14,078 -5,5434 0,90953]	6
BVP	[0,0031536 0,018922 0,047304 0,063072 0,047304 0,018922 0,0031536]	[1 -2,2303 2,649 -1,8176 0,76097 -0,17851 0,018297]	6

Table B.1 – Les paramètres des filtres RII

Le signal	B	A	Ordre du filtre
EMG	[-0,046457 0,014726 0,039312 0,074577 0,11301 0,14557 0,16419 0,16419 0,14557 0,11301 0,074577 0,039312 0,014726 -0,046457]	[1]	13
SKC	[0,017852 0,007396 0,0088124 0,010323 0,011925 0,013605 0,015346 0,017114 0,018887 0,020654 0,02241 0,024125 0,025742 0,02731 0,028737 0,030044 0,031199 0,03219 0,032995 0,033616 0,034029 0,034241 0,034241 0,034029 0,033616 0,032995 0,03219 0,031199 0,030044 0,028737 0,02731 0,025742 0,024125 0,02241 0,020654 0,018887 0,017114 0,015346 0,013605 0,011925 0,010323 0,0088124 0,007396 0,017852]	[1]	43
VR	[0,017498 0,0091942 0,011392 0,013772 0,016309 0,018968 0,021699 0,024441 0,027152 0,029806 0,032336 0,034661 0,036787 0,038616 0,040133 0,041302 0,042097 0,042495 0,042495 0,042097 0,041302 0,040133 0,038616 0,036787 0,034661 0,032336 0,029806 0,027152 0,024441 0,021699 0,018968 0,016309 0,013772 0,011392 0,0091942 0,017498]	[1]	35
SKT	[0,015437 0,0032395 0,0035646 0,0039031 0,0042544 0,0046205 0,0050008 0,0053912 0,0057892 0,0061933 0,0066042 0,0070283 0,0074701 0,0079133 0,0083266 0,0087861 0,0092211 0,0096626 0,010105 0,010544 0,010979 0,011409 0,011831 0,012246 0,012655 0,013054 0,013436 0,013804 0,014165 0,0145 0,014824 0,01513 0,015414 0,015679 0,015924 0,016146 0,016346 0,016524 0,016677 0,016805 0,01691 0,016989 0,01704 0,017068 0,017068 0,01704 0,016989 0,01691 0,016805 0,016677 0,016524 0,016346 0,016146 0,015924 0,015679 0,015414 0,01513 0,014824 0,0145 0,014165 0,013804 0,013436 0,013054 0,012655 0,012246 0,011831 0,011409 0,010979 0,010544 0,010105 0,0096626 0,0092211 0,0087861 0,0083266 0,0079133 0,0074701 0,0070283 0,0066042 0,0061933 0,0057892 0,0053912 0,0050008 0,0046205 0,0042544 0,0039031 0,0035646 0,0032395 0,015437]	[1]	87
BVP	[0,024677 0,024798 0,035045 0,046089 0,057133 0,067374 0,075963 0,082106 0,085347 0,085347 0,082106 0,075963 0,067374 0,057133 0,046089 0,035045 0,024798 0,024677]	[1]	17

TABLE B.2 – Les paramètres des filtres RIF

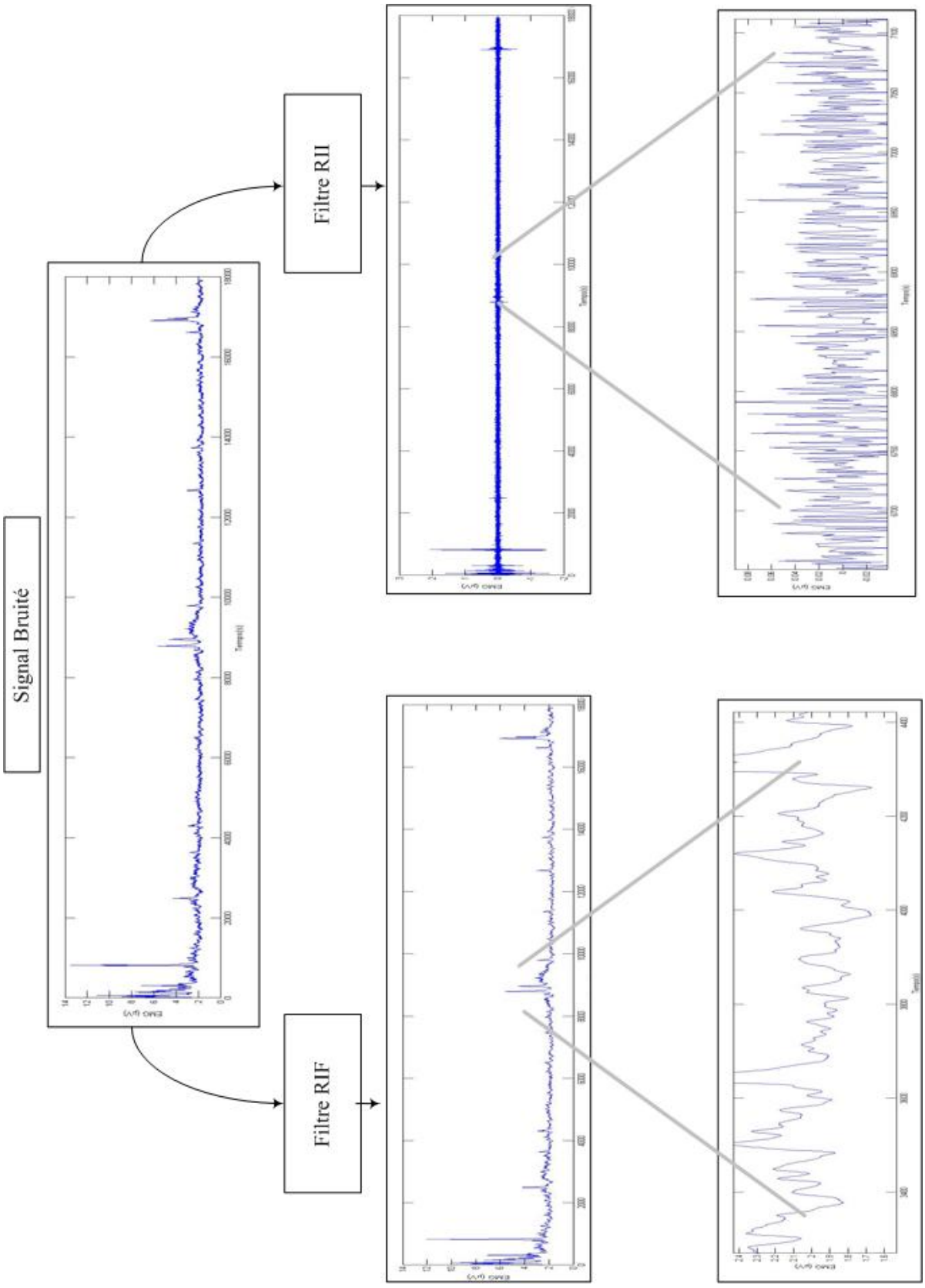


FIGURE B.1 – Filtrage du signal EMG

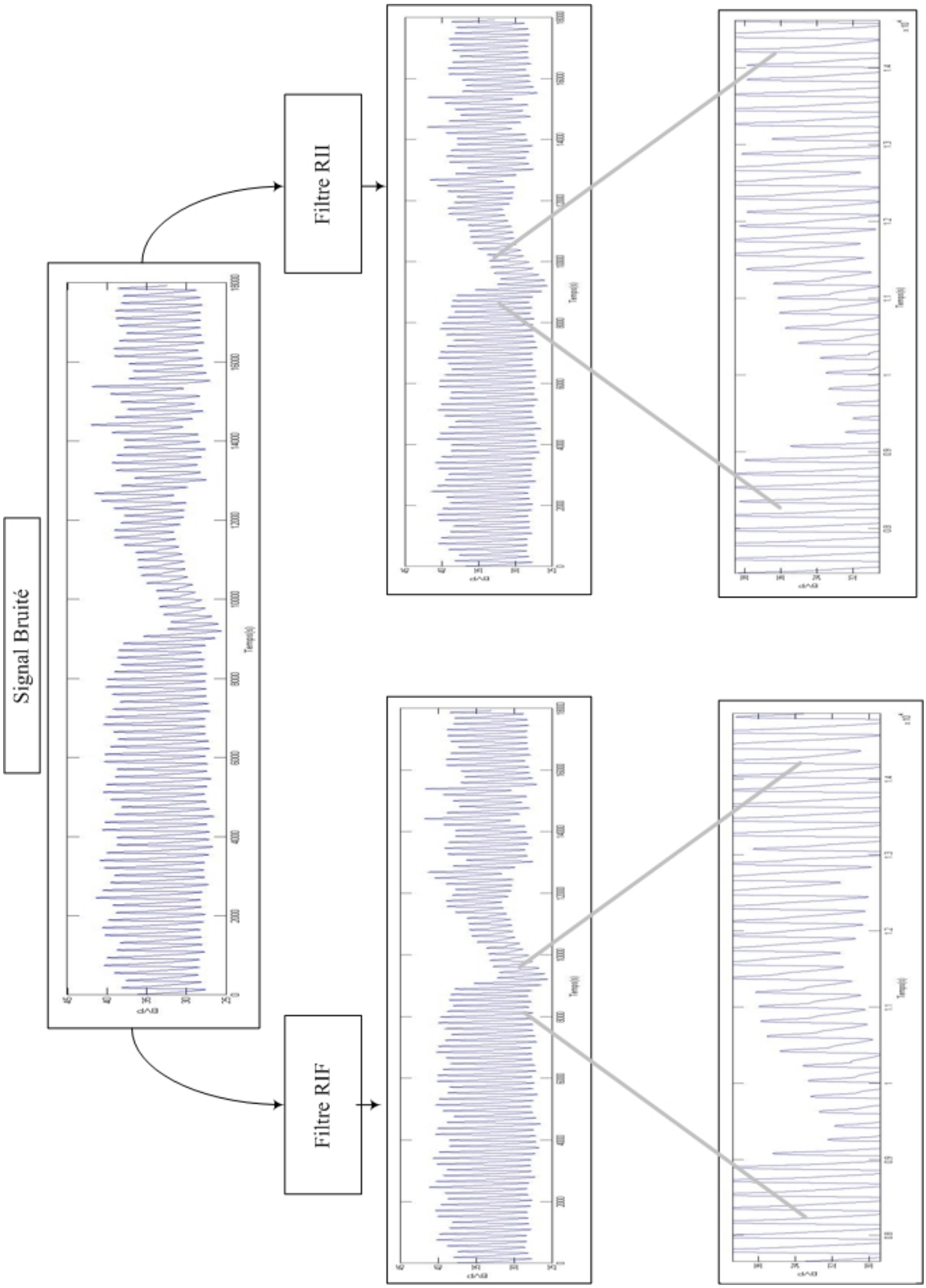


FIGURE B.2 – Filtrage du signal BVP

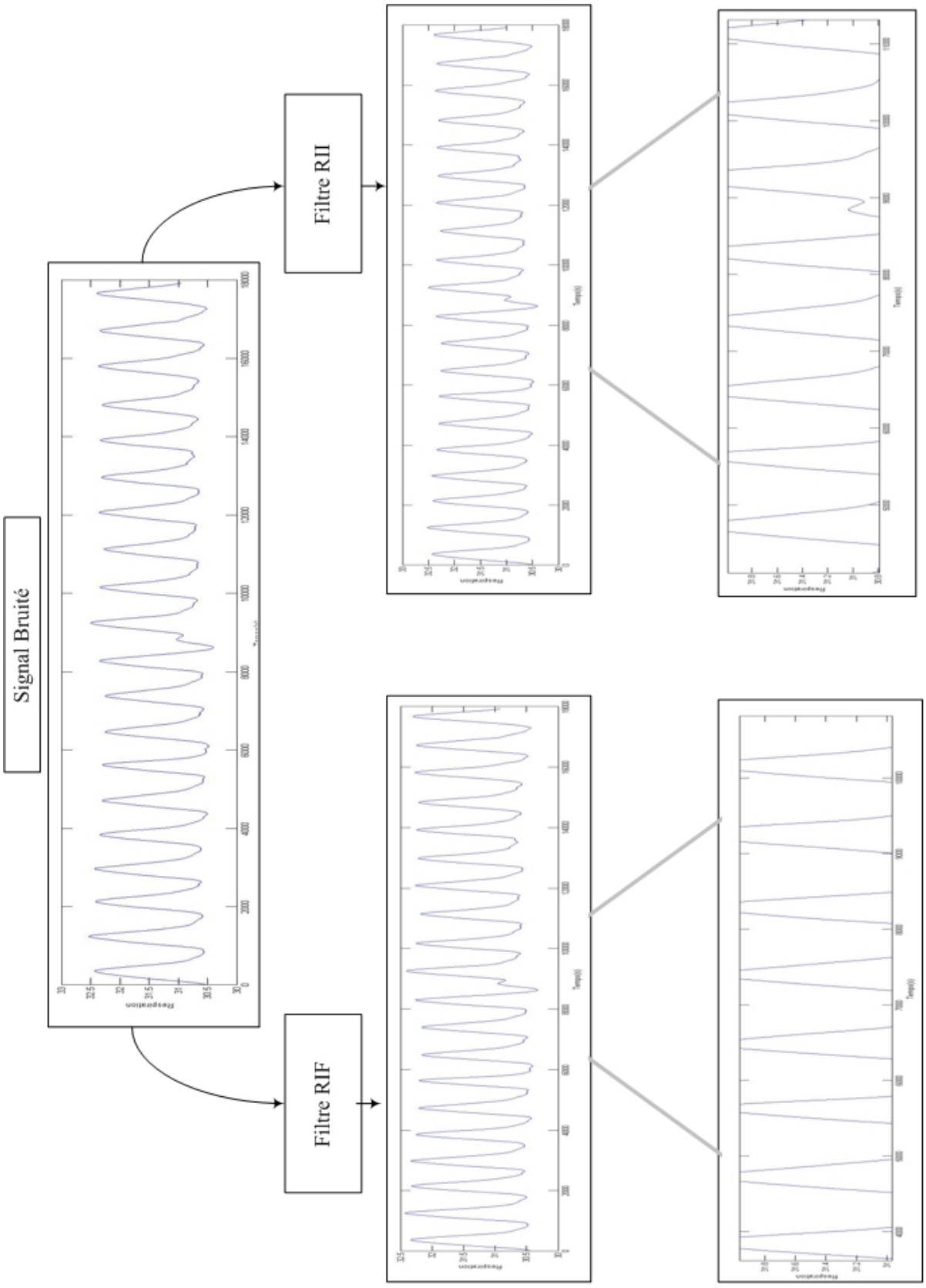


FIGURE B.3 – Filtrage du signal VR

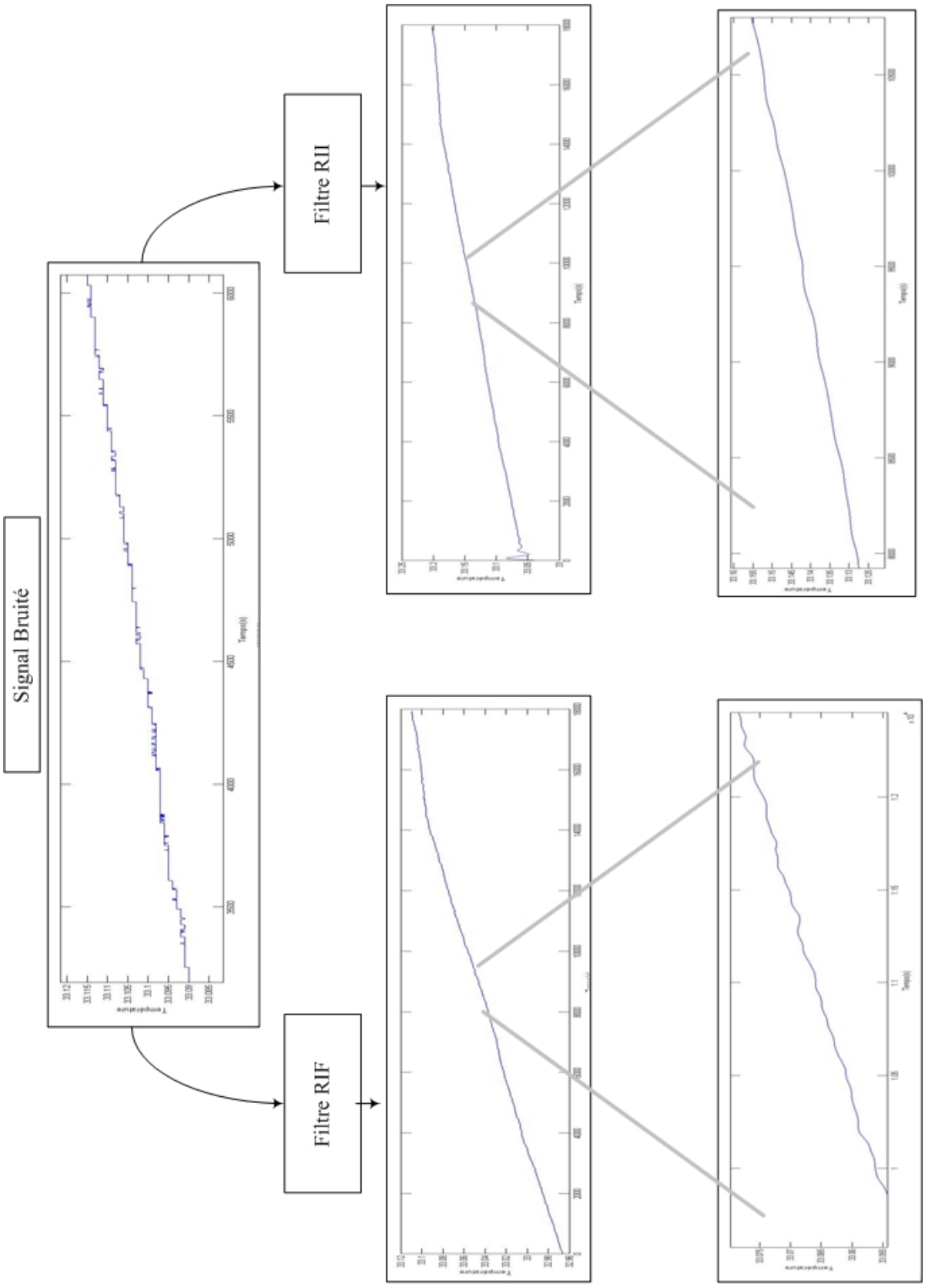


FIGURE B.4 – Filtrage du signal SKT

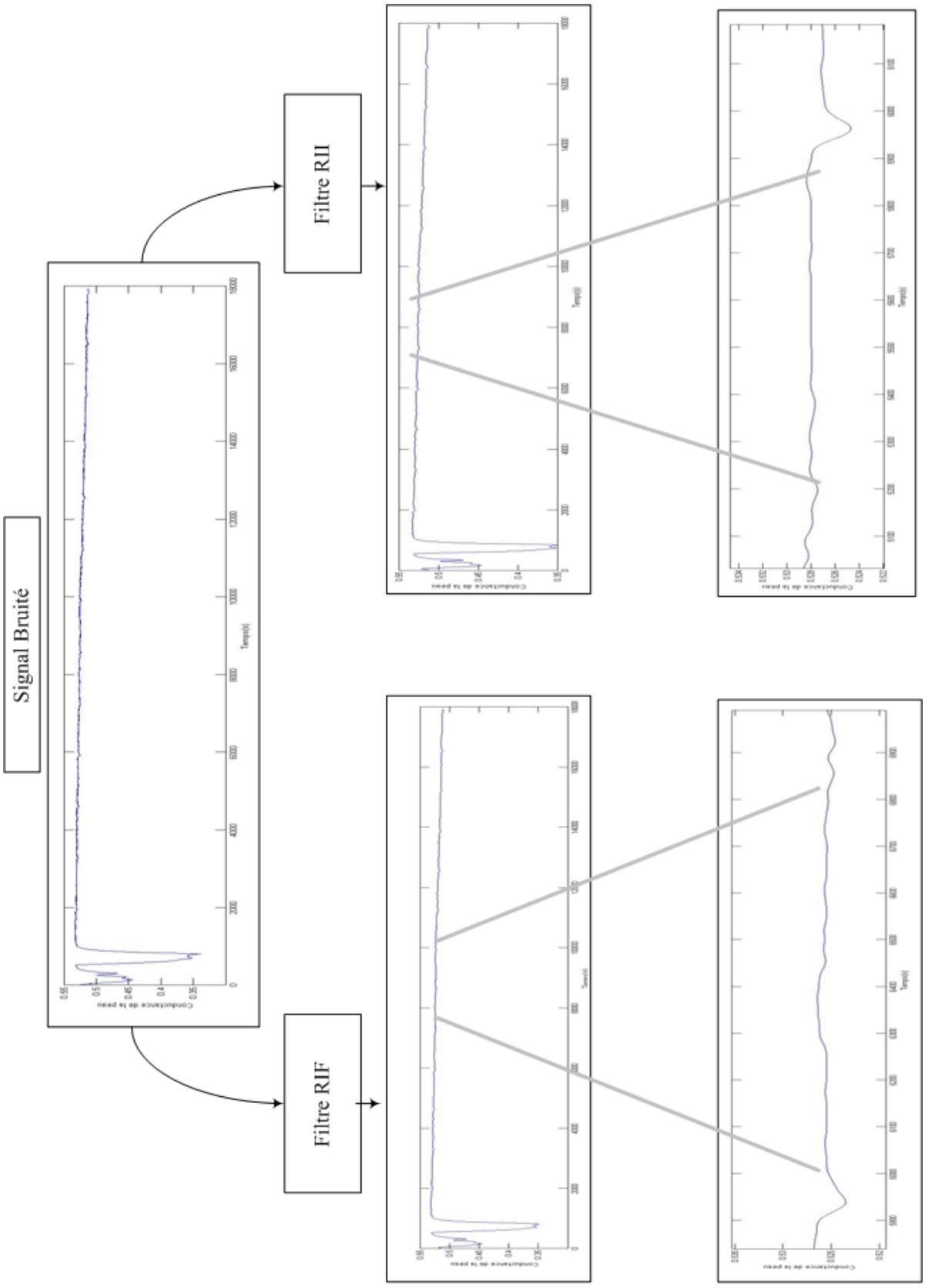


FIGURE B.5 – Filtrage du signal SC

Annexe C

Le tri des caractéristiques utilisées avec la sélection de l'information mutuelle

Les caractéristiques faciales = [$D1$, $D2$,...,$D21$].

Les caractéristiques physiologiques = [μ_{bvp}, σ_{bvp}, δ_{bvp}, $\bar{\delta}_{bvp}$, γ_{bvp}, $\bar{\gamma}_{bvp}$, μ_{emg}, σ_{emg}, δ_{emg}, $\bar{\delta}_{emg}$, γ_{emg}, $\bar{\gamma}_{emg}$, μ_{sc}, σ_{sc}, δ_{sc}, $\bar{\delta}_{sc}$, γ_{sc}, $\bar{\gamma}_{sc}$, μ_{skt}, σ_{skt}, δ_{skt}, $\bar{\delta}_{skt}$, γ_{skt}, $\bar{\gamma}_{skt}$, μ_{resp}, σ_{resp}, δ_{resp}, $\bar{\delta}_{resp}$, γ_{resp}, $\bar{\gamma}_{resp}$].

	Sujet1	Sujet2	Sujet3	Sujet4	Sujet5	Sujet6	Sujet7	Sujet8	Sujet9	Sujet10	Tout
P1	μ_{bvp}	μ_{bvp}	μ_{bvp}	μ_{bvp}	μ_{bvp}	μ_{bvp}	μ_{emg}	μ_{bvp}	μ_{emg}	μ_{resp}	μ_{emg}
P2	μ_{emg}	μ_{emg}	μ_{emg}	μ_{emg}	μ_{emg}	μ_{emg}	μ_{sc}	μ_{emg}	μ_{bvp}	μ_{skt}	μ_{skt}
P3	γ_{sc}	μ_{sc}	μ_{sc}	μ_{sc}	μ_{sc}	μ_{sc}	γ_{sc}	μ_{sc}	μ_{resp}	μ_{bvp}	δ_{skt}
P4	μ_{skt}	γ_{sc}	μ_{skt}	μ_{skt}	γ_{sc}	μ_{skt}	μ_{skt}	δ_{sc}	μ_{skt}	μ_{emg}	γ_{skt}
P5	σ_{skt}	μ_{skt}	σ_{skt}	δ_{skt}	μ_{skt}	σ_{skt}	σ_{skt}	γ_{sc}	μ_{sc}	μ_{sc}	μ_{resp}
P6	δ_{skt}	σ_{skt}	δ_{skt}	γ_{skt}	σ_{skt}	δ_{skt}	δ_{skt}	μ_{skt}	δ_{emg}	σ_{emg}	σ_{skt}
P7	γ_{skt}	δ_{skt}	γ_{skt}	μ_{resp}	δ_{skt}	γ_{skt}	γ_{skt}	σ_{skt}	σ_{emg}	σ_{bvp}	γ_{resp}
P8	μ_{resp}	γ_{skt}	μ_{resp}	γ_{sc}	γ_{skt}	μ_{resp}	μ_{resp}	δ_{skt}	δ_{bvp}	$D8$	μ_{bvp}
P9	γ_{resp}	μ_{resp}	γ_{resp}	σ_{skt}	μ_{resp}	γ_{resp}	γ_{resp}	γ_{skt}	δ_{resp}	$D14$	γ_{sc}
P10	δ_{sc}	γ_{resp}	δ_{sc}	γ_{resp}	γ_{resp}	γ_{sc}	μ_{bvp}	μ_{resp}	σ_{bvp}	$D18$	δ_{sc}
P11	σ_{sc}	δ_{sc}	γ_{sc}	δ_{sc}	δ_{sc}	σ_{bvp}	δ_{sc}	γ_{resp}	σ_{resp}	$D9$	σ_{bvp}
P12	σ_{bvp}	σ_{sc}	σ_{sc}	σ_{emg}	$D6$	$D15$	δ_{bvp}	σ_{sc}	δ_{skt}	$D17$	σ_{emg}
P13	δ_{bvp}	σ_{emg}	$D21$	σ_{bvp}	σ_{bvp}	δ_{bvp}	$D9$	σ_{bvp}	δ_{sc}	$D2$	δ_{bvp}
P14	σ_{emg}	σ_{bvp}	σ_{emg}	δ_{emg}	σ_{emg}	$D16$	σ_{bvp}	δ_{bvp}	σ_{skt}	$D19$	$D16$
P15	$D17$	$D9$	δ_{bvp}	δ_{bvp}	$D9$	σ_{emg}	σ_{emg}	σ_{emg}	σ_{sc}	$D21$	δ_{emg}
P16	$D16$	δ_{bvp}	σ_{bvp}	γ_{emg}	$D17$	δ_{emg}	δ_{emg}	$D16$	$D1$	$D5$	$D15$
P17	$D15$	$D16$	$D15$	$D14$	δ_{bvp}	$D20$	γ_{emg}	δ_{emg}	$D6$	$D6$	$D9$
P18	$D9$	$D17$	$D16$	$D16$	$D18$	γ_{emg}	$D16$	γ_{emg}	$D4$	δ_{skt}	γ_{emg}
P19	$D14$	$D14$	$D8$	$D10$	$D16$	$D8$	$D8$	$D15$	$D3$	γ_{skt}	$D17$
P20	$D18$	$D8$	δ_{emg}	$D7$	δ_{emg}	$D21$	$D2$	$D12$	$D5$	δ_{sc}	$D10$
P21	$D8$	δ_{emg}	γ_{emg}	$D17$	$D8$	$D14$	$D21$	$D7$	$D13$	σ_{skt}	$D8$
P22	γ_{emg}	$D18$	$D13$	$D20$	γ_{emg}	$D17$	$D10$	$D10$	$D12$	γ_{sc}	$D14$
P23	δ_{emg}	$D15$	$D9$	σ_{resp}	$D15$	$D13$	$D17$	$D20$	$D2$	$D4$	$D7$
P24	σ_{resp}	$D1$	$D14$	$D11$	$D11$	$\bar{\gamma}_{bvp}$	$D11$	$D11$	δ_{bvp}	$D5$	$D18$
P25	$D20$	γ_{emg}	$D17$	$D9$	$D10$	$\bar{\gamma}_{emg}$	$D1$	σ_{resp}	$\bar{\gamma}_{bvp}$	σ_{sc}	σ_{resp}

TABLE C.1 – Le tri des caractéristiques de la base 1 (acquisition durant 4 jours) de P1 à P25

	Sujet1	Sujet2	Sujet3	Sujet4	Sujet5	Sujet6	Sujet7	Sujet8	Sujet9	Sujet10	Tout
P26	$D19$	$D21$	$D10$	$D1$	$D21$	$D20$	$D19$	$D1$	$D4$	$D20$	$D19$
P27	$D15$	$D17$	$D18$	$D7$	$D13$	$D21$	γ_{bvp}	$D15$	$D19$	$D18$	$D12$
P28	$D12$	$D18$	$D9$	γ_{bvp}	$D1$	γ_{sc}	γ_{emg}	$D2$	$D5$	$D5$	$D2$
P29	$D5$	$D16$	$D8$	γ_{emg}	$D6$	$D3$	γ_{sc}	$D4$	$D20$	$D6$	$D5$
P30	$D4$	$D7$	$D15$	γ_{sc}	$D5$	$D5$	γ_{skt}	$D12$	$D3$	$D1$	$D7$
P31	$D3$	$D1$	$D12$	γ_{skt}	$D16$	$D6$	γ_{resp}	$D16$	γ_{bvp}	$D12$	$D20$
P32	γ_{bvp}	$D2$	γ_{bvp}	γ_{resp}	$D2$	$D17$	$D20$	$D3$	γ_{emg}	$D2$	γ_{bvp}
P33	γ_{emg}	γ_{bvp}	γ_{emg}	$D2$	$D3$	$D2$	$D7$	$D7$	γ_{sc}	γ_{bvp}	γ_{emg}
P34	γ_{sc}	γ_{emg}	γ_{sc}	$D5$	$D15$	γ_{bvp}	σ_{bvp}	$D20$	γ_{skt}	γ_{emg}	γ_{skt}
P35	γ_{skt}	γ_{sc}	γ_{skt}	$D4$	γ_{bvp}	γ_{emg}	δ_{bvp}	γ_{bvp}	γ_{resp}	γ_{skt}	γ_{resp}
P36	γ_{resp}	γ_{skt}	γ_{resp}	$D3$	γ_{emg}	γ_{skt}	σ_{emg}	γ_{emg}	$D15$	γ_{resp}	γ_{sc}
P37	$D7$	γ_{resp}	$D11$	$D20$	γ_{sc}	γ_{resp}	δ_{emg}	γ_{sc}	$D12$	$D19$	$D1$
P38	$D13$	$D4$	$D19$	$D6$	γ_{skt}	$D8$	σ_{sc}	γ_{skt}	σ_{sc}	$D7$	$D4$
P39	$D21$	$D3$	$D16$	σ_{bvp}	γ_{resp}	$D1$	δ_{sc}	γ_{resp}	σ_{bvp}	$D4$	$D6$
P40	$D6$	$D6$	σ_{bvp}	δ_{bvp}	$D4$	$D12$	σ_{skt}	σ_{bvp}	δ_{bvp}	$D3$	$D3$
P41	σ_{bvp}	$D20$	δ_{bvp}	σ_{emg}	$D7$	σ_{sc}	δ_{skt}	δ_{bvp}	σ_{emg}	μ_{sc}	σ_{sc}
P42	δ_{bvp}	$D5$	σ_{emg}	δ_{emg}	σ_{bvp}	σ_{skt}	σ_{resp}	σ_{emg}	δ_{emg}	γ_{sc}	σ_{skt}
P43	σ_{emg}	δ_{bvp}	δ_{emg}	σ_{sc}	δ_{bvp}	σ_{resp}	δ_{resp}	δ_{emg}	δ_{sc}	σ_{skt}	σ_{resp}
P44	δ_{emg}	γ_{bvp}	σ_{sc}	δ_{sc}	σ_{emg}	δ_{skt}	$D2$	σ_{sc}	σ_{skt}	σ_{sc}	δ_{skt}
P45	σ_{sc}	δ_{emg}	δ_{sc}	σ_{skt}	δ_{emg}	δ_{sc}	$D5$	δ_{sc}	δ_{skt}	σ_{bvp}	σ_{bvp}
P46	δ_{sc}	σ_{sc}	σ_{skt}	δ_{skt}	σ_{sc}	$D11$	$D3$	σ_{skt}	σ_{resp}	δ_{bvp}	δ_{bvp}
P47	σ_{skt}	δ_{sc}	δ_{skt}	σ_{resp}	δ_{sc}	σ_{bvp}	$D4$	δ_{skt}	δ_{resp}	σ_{emg}	σ_{emg}
P48	δ_{skt}	σ_{skt}	σ_{resp}	δ_{resp}	σ_{skt}	δ_{bvp}	$D6$	σ_{resp}	$D6$	δ_{emg}	δ_{emg}
P49	σ_{resp}	δ_{skt}	δ_{resp}	μ_{sc}	δ_{skt}	σ_{emg}	$D1$	δ_{resp}	μ_{sc}	δ_{skt}	δ_{resp}
P50	δ_{resp}	σ_{resp}	$D7$	μ_{skt}	σ_{resp}	δ_{emg}	μ_{skt}	μ_{skt}	μ_{skt}	σ_{resp}	δ_{sc}
P51	$D1$	δ_{resp}	$D20$	μ_{resp}	δ_{resp}	δ_{resp}	μ_{resp}	μ_{resp}	μ_{resp}	δ_{resp}	μ_{sc}

TABLE C.2 – Le tri des caractéristiques de la base 1 (acquisition durant 4 jours) de P26 à P51

	Sujet1	Sujet2	Sujet3	Sujet4	Sujet5	Sujet6	Sujet7	Sujet8	Sujet9	Sujet10	Tout
P1	μ_{bvp}	μ_{bvp}	μ_{bvp}	μ_{bvp}	μ_{bvp}	μ_{bvp}	μ_{emg}	μ_{bvp}	μ_{emg}	μ_{resp}	μ_{emg}
P2	μ_{emg}	μ_{emg}	μ_{emg}	μ_{emg}	μ_{emg}	μ_{emg}	μ_{sc}	μ_{emg}	μ_{bvp}	μ_{skt}	μ_{skt}
P3	γ_{sc}	μ_{sc}	μ_{sc}	μ_{sc}	μ_{sc}	μ_{sc}	γ_{sc}	μ_{sc}	μ_{resp}	μ_{bvp}	δ_{skt}
P4	μ_{skt}	γ_{sc}	μ_{skt}	μ_{skt}	γ_{sc}	μ_{skt}	μ_{skt}	δ_{sc}	μ_{skt}	μ_{emg}	γ_{skt}
P5	σ_{skt}	μ_{skt}	σ_{skt}	δ_{skt}	μ_{skt}	σ_{skt}	σ_{skt}	γ_{sc}	μ_{sc}	μ_{sc}	μ_{resp}
P6	δ_{skt}	σ_{skt}	δ_{skt}	γ_{skt}	σ_{skt}	δ_{skt}	δ_{skt}	μ_{skt}	δ_{emg}	σ_{emg}	σ_{skt}
P7	γ_{skt}	δ_{skt}	γ_{skt}	μ_{resp}	δ_{skt}	γ_{skt}	γ_{skt}	σ_{skt}	σ_{emg}	σ_{bvp}	γ_{resp}
P8	μ_{resp}	γ_{skt}	μ_{resp}	γ_{sc}	γ_{skt}	μ_{resp}	μ_{resp}	δ_{skt}	δ_{bvp}	$D8$	μ_{bvp}
P9	γ_{resp}	μ_{resp}	γ_{resp}	σ_{skt}	μ_{resp}	γ_{resp}	γ_{resp}	γ_{skt}	δ_{resp}	$D14$	γ_{sc}
P10	δ_{sc}	γ_{resp}	δ_{sc}	γ_{resp}	γ_{resp}	γ_{sc}	μ_{bvp}	μ_{resp}	σ_{bvp}	$D18$	δ_{sc}
P11	σ_{sc}	δ_{sc}	γ_{sc}	δ_{sc}	δ_{sc}	σ_{bvp}	δ_{sc}	γ_{resp}	σ_{resp}	$D9$	σ_{bvp}
P12	σ_{bvp}	σ_{sc}	σ_{sc}	σ_{emg}	$D6$	$D15$	δ_{bvp}	σ_{sc}	δ_{skt}	$D17$	σ_{emg}
P13	δ_{bvp}	σ_{emg}	$D21$	σ_{bvp}	σ_{bvp}	δ_{bvp}	$D9$	σ_{bvp}	δ_{sc}	$D2$	δ_{bvp}
P14	σ_{emg}	σ_{bvp}	σ_{emg}	δ_{emg}	σ_{emg}	$D16$	σ_{bvp}	δ_{bvp}	σ_{skt}	$D19$	$D16$
P15	$D17$	$D9$	δ_{bvp}	δ_{bvp}	$D9$	σ_{emg}	σ_{emg}	σ_{emg}	σ_{sc}	$D21$	δ_{emg}
P16	$D16$	δ_{bvp}	σ_{bvp}	γ_{emg}	$D17$	δ_{emg}	δ_{emg}	$D16$	$D1$	$D5$	$D15$
P17	$D15$	$D16$	$D15$	$D14$	δ_{bvp}	$D20$	γ_{emg}	δ_{emg}	$D6$	$D6$	$D9$
P18	$D9$	$D17$	$D16$	$D16$	$D18$	γ_{emg}	$D16$	γ_{emg}	$D4$	δ_{skt}	γ_{emg}
P19	$D14$	$D14$	$D8$	$D10$	$D16$	$D8$	$D8$	$D15$	$D3$	γ_{skt}	$D17$
P20	$D18$	$D8$	δ_{emg}	$D7$	δ_{emg}	$D21$	$D2$	$D12$	$D5$	δ_{sc}	$D10$
P21	$D8$	δ_{emg}	γ_{emg}	$D17$	$D8$	$D14$	$D21$	$D7$	$D13$	σ_{skt}	$D8$
P22	γ_{emg}	$D18$	$D13$	$D20$	γ_{emg}	$D17$	$D10$	$D10$	$D12$	γ_{sc}	$D14$
P23	δ_{emg}	$D15$	$D9$	σ_{resp}	$D15$	$D13$	$D17$	$D20$	$D2$	$D4$	$D7$
P24	σ_{resp}	$D1$	$D14$	$D11$	$D11$	$\bar{\gamma}_{bvp}$	$D11$	$D11$	$\bar{\delta}_{bvp}$	$D5$	$D18$
P25	$D20$	γ_{emg}	$D17$	$D9$	$D10$	$\bar{\gamma}_{emg}$	$D1$	σ_{resp}	$\bar{\gamma}_{bvp}$	σ_{sc}	σ_{resp}

TABLE C.3 – Le tri des caractéristiques de la base 2 (acquisition pendant un seul jour) de P1 à P25

	Sujet1	Sujet2	Sujet3	Sujet4	Sujet5	Sujet6	Sujet7	Sujet8	Sujet9	Sujet10	Tout
P26	$D6$	σ_{resp}	$D12$	$D18$	$D14$	$\bar{\gamma}_{sc}$	$D7$	$D13$	$\bar{\delta}_{emg}$	$D3$	$D11$
P27	$D10$	$D13$	$\bar{\delta}_{bvp}$	$D15$	$\bar{\gamma}_{bvp}$	$\bar{\gamma}_{skt}$	σ_{resp}	$D1$	$\bar{\gamma}_{emg}$	$D20$	$D1$
P28	$D7$	$D10$	$\bar{\delta}_{emg}$	$D19$	$\bar{\gamma}_{emg}$	$\bar{\gamma}_{resp}$	$D15$	$\bar{\gamma}_{bvp}$	$\bar{\delta}_{sc}$	δ_{resp}	$D13$
P29	$\bar{\gamma}_{bvp}$	$\bar{\delta}_{bvp}$	$\bar{\delta}_{sc}$	$\bar{\gamma}_{bvp}$	$\bar{\gamma}_{sc}$	σ_{resp}	$\bar{\gamma}_{bvp}$	$\bar{\gamma}_{emg}$	$\bar{\gamma}_{sc}$	$D11$	$D20$
P30	$\bar{\gamma}_{emg}$	$\bar{\delta}_{emg}$	$\bar{\delta}_{skt}$	$\bar{\gamma}_{emg}$	$\bar{\gamma}_{skt}$	$D10$	$\bar{\gamma}_{emg}$	$\bar{\gamma}_{sc}$	$\bar{\delta}_{skt}$	$D13$	$\bar{\gamma}_{bvp}$
P31	$\bar{\gamma}_{sc}$	σ_{sc}	$\bar{\delta}_{resp}$	$\bar{\gamma}_{sc}$	$\bar{\gamma}_{resp}$	$\bar{\delta}_{bvp}$	$\bar{\gamma}_{sc}$	$\bar{\gamma}_{skt}$	$\bar{\gamma}_{skt}$	$D7$	$\bar{\gamma}_{emg}$
P32	$\bar{\gamma}_{skt}$	$\bar{\delta}_{skt}$	$D7$	$\bar{\gamma}_{skt}$	$D19$	$\bar{\delta}_{emg}$	$\bar{\gamma}_{skt}$	$\bar{\gamma}_{resp}$	$\bar{\delta}_{resp}$	γ_{bvp}	$\bar{\gamma}_{sc}$
P33	$\bar{\gamma}_{resp}$	$\bar{\delta}_{resp}$	$\bar{\gamma}_{bvp}$	$\bar{\gamma}_{resp}$	$\bar{\delta}_{bvp}$	$\bar{\delta}_{sc}$	$\bar{\gamma}_{resp}$	$D21$	$\bar{\gamma}_{resp}$	$D12$	$\bar{\gamma}_{skt}$
P34	$D21$	$\bar{\gamma}_{bvp}$	$\bar{\gamma}_{emg}$	$D4$	$\bar{\delta}_{emg}$	$\bar{\delta}_{skt}$	$\bar{\delta}_{bvp}$	$D3$	$D7$	$D10$	$\bar{\gamma}_{resp}$
P35	$\bar{\delta}_{bvp}$	$\bar{\gamma}_{emg}$	$\bar{\gamma}_{sc}$	$\bar{\delta}_{bvp}$	$\bar{\delta}_{sc}$	$\bar{\delta}_{resp}$	$\bar{\delta}_{emg}$	$\bar{\delta}_{bvp}$	$D14$	$D15$	$\bar{\delta}_{bvp}$
P36	$\bar{\delta}_{emg}$	$\bar{\gamma}_{sc}$	$\bar{\gamma}_{skt}$	$\bar{\delta}_{emg}$	$\bar{\delta}_{skt}$	$D2$	$\bar{\delta}_{sc}$	$\bar{\delta}_{emg}$	$D15$	$\bar{\gamma}_{bvp}$	$\bar{\delta}_{emg}$
P37	$\bar{\delta}_{sc}$	$\bar{\gamma}_{skt}$	$\bar{\gamma}_{resp}$	$\bar{\delta}_{sc}$	$\bar{\delta}_{resp}$	$D9$	$\bar{\delta}_{skt}$	$\bar{\delta}_{sc}$	$D19$	$\bar{\gamma}_{emg}$	$\bar{\delta}_{sc}$
P38	δ_{skt}	$\bar{\gamma}_{resp}$	$D10$	$\bar{\delta}_{skt}$	$D7$	$D12$	δ_{resp}	$\bar{\delta}_{skt}$	γ_{bvp}	$\bar{\gamma}_{sc}$	$\bar{\delta}_{skt}$
P39	$\bar{\delta}_{resp}$	$D21$	$D1$	$\bar{\delta}_{resp}$	σ_{resp}	$D11$	$D12$	$\bar{\delta}_{resp}$	γ_{emg}	$\bar{\gamma}_{skt}$	$\bar{\delta}_{resp}$
P40	μ_{sc}	$D6$	σ_{resp}	$D3$	$D13$	$D3$	$D20$	$D14$	γ_{sc}	$\bar{\gamma}_{resp}$	$D12$
P41	γ_{bvp}	$D2$	$D11$	$D6$	$D1$	$D7$	$D18$	γ_{bvp}	γ_{resp}	$D1$	$D21$
P42	$D2$	$D7$	γ_{bvp}	γ_{bvp}	γ_{bvp}	$D6$	$D19$	$D6$	γ_{skt}	σ_{resp}	$D2$
P43	$D1$	γ_{bvp}	$D20$	$D13$	$D20$	γ_{bvp}	γ_{bvp}	$D4$	$D16$	δ_{emg}	$D19$
P44	$D13$	$D3$	$D18$	$D12$	$D4$	$D5$	$D6$	δ_{resp}	$D20$	$D16$	γ_{bvp}
P45	$D11$	$D12$	$D2$	$D8$	δ_{resp}	$D4$	$D13$	$D19$	$D17$	γ_{emg}	$D6$
P46	$D3$	$D19$	$D19$	δ_{resp}	$D5$	$D19$	$D4$	$D5$	$D18$	δ_{bvp}	$D3$
P47	$D19$	$D11$	$D4$	$D1$	$D12$	$D1$	$D14$	$D9$	$D11$	$\bar{\delta}_{bvp}$	$D4$
P48	δ_{resp}	$D4$	δ_{resp}	$D2$	$D3$	δ_{resp}	$D3$	$D2$	$D10$	$\bar{\delta}_{emg}$	δ_{resp}
P49	$D4$	δ_{resp}	$D6$	$D21$	σ_{sc}	$D18$	δ_{resp}	$D17$	$D21$	δ_{sc}	$D5$
P50	$D12$	$D5$	$D3$	σ_{sc}	$D21$	σ_{sc}	σ_{sc}	$D8$	$D9$	$\bar{\delta}_{skt}$	σ_{sc}
P51	$D5$	$D20$	$D5$	$D5$	$D2$	δ_{sc}	$D5$	$D18$	$D8$	$\bar{\delta}_{resp}$	μ_{sc}

TABLE C.4 – Le tri des caractéristiques de la base 2 (acquisition pendant un seul jour) de P26 à P51

Bibliographie

[1] http ://en.wikipedia.org/wiki/interbeat-interval.

[2] http ://wassil.free.fr.

[3] http ://www.em-consulte.com/article/146215.

[4] http ://www.ipsp.ucl.ac.be/recherche/filmstim.

[5] http ://www.teaergo.com/.

[6] http ://www.u-picardie.fr/labo/ugbm.

[7] http ://sourceforge.net/projects/openpnl/, 2003.

[8] B. Abboud, F. Davoine et M. Dang : Facial expression recognition and synthesis based on appearance model. *Signal Processing : Image Communication*, 19:723–740, 2004.

[9] F. Abdat, C.Maaoui et A.Pruski : Gradient based method for static facial features localization. *International Conference Visualization, Imaging and Image Processig , VIIP07*, 2007.

[10] F. Abdat, C.Maaoui et A.Pruski : Suivi du gradient pour la localisation des caractéristiques faciales. *Colloque de GRETSI ,Troyes France*, 2007.

[11] F. Abdat, C.Maaoui et A.Pruski : Le suivi des caractéristiques faciales en temps réel avec l'algorithme pyramidal de lucas-kanade. *Colloque Handicap, Paris*, 2008.

[12] F. Abdat, C.Maaoui et A.Pruski : Overview of automatic facial expressions analysis. *IEEE International Conference on System, Signals and Devices*, 2009.

[13] F. Abdat, C.Maaoui et A.Pruski : Reconnaissance d'expressions faciales en temps réel à partir d'une séquence vidéo. *Sciences et Technologies pour le Handicap, Edition Hermés*, 2009.

[14] F. Abdat, C.Maaoui et A.Pruski : *Human-Robot Interaction*, chapitre Real Facial Feature Points Tracking with Pyramidal Lucas-Kanade Algorithm. I-Tech Education , Vienna, Austria, 2010.

[15] F. Abdat, C. Maaoui et A. Pruski : Facial feature extraction for emotion recognition. *IEEE International Conference Human Interaction, HUMAN07 Algeria*, 2007.

[16] F. Abdat, C. Maaoui et A. Pruski : Real facial feature points tracking with pyramidal lucas-kanade algorithm. *IEEE RO-MAN08, The 17th International Symposium on Robot and Human Interactive Communication, Germany*, 2008.

[17] F. Abdat, C. Maaoui et A. Pruski : Tracking of points detected using wavelet transform for facial expression recognition. *IEEE International Conference on Systems Signals and Devices SSD09*, 2009.

[18] K. Anderson et P. McOwn : A real-time automated system for the recognition of human facial expressions. *IEEE Trans. on Systems, Man, And Cybernetics Part B : Cybernetics*, 36:96–105, 2006.

[19] C. André et F. Lelord : *La force des émotions*. Odile Jacob, 2003.

[20] J. Bailonson, E. Pontikakisb, I. Maussc, J. Grossd, M. Jabone, C. Hutchersond, C. Nassa et O. Johnf : Real-time classification of evoked emotions using facial feature tracking and physiological responses. *International Human Computer Studies*, 66:303–317, 2008.

[21] R. Bandler : *Un cerveau pour changer*. InterEdition, 1990.

[22] T. Banziger, D. Grandjean, P. Bernard, G. Klasmeyer et K. Scherer : Prosodie de l'émotion : étude de l'encodage et du décodage. *Cahiers de Linguistique française*, 23, 2001.

[23] P. Barralon : *Classification et fusion de données actimétriques pour la télévigilance médicale*. Thèse en traitement du signal et des images, L'université de Joseph Fourier, 2005.

[24] J. N. Bassili : Facial motion in the perception of faces and of emotional expression. *Experimental Psychology - Human Perception and Performance*, 4:373–379, 1978.

[25] R. Battiti : Using mutual information for selecting features in supervised neural net learning. *IEEE Transactions on Neural Networks*, 5:537–550, 1994.

[26] A. Bensalem : *Modèles Probabilistes de Séquences Temporelles et Fusion de Décisions : Application à la Classification de Défauts de Rails et à leur Maintenance*. Thèse, université Henri Poincaré Nancy, 2008.

[27] T. W. Bickmore et R. W. Picard : Towards caring machines. *Conference on Human Factors in Computing Systems CHI '04 extended abstracts on Human factors in computing systems, Vienna,Austria*, 2004.

[28] M.J Black et Y. Yacoob : Recognizing facial expressions in image sequences using local parametrized models of image motion. *International conf. on Computer Vision*, pages 374–381, 1995.

[29] A. Botino : Real time head and facial features tracking from uncalibrated monocular views. *Proc. 5th Asian Conference on Computer Vision ACCV Australia*, pages 23–25, 2002.

[30] J.Y. Bouguet : Pyramidal implementation of the lucas kanade feature tracker. *Intel Corporation*, Microprocessor Research Labs, 2000.

[31] J.M. Bouroche et G. Saporta : L'analyse des données. que sais-je ? *Presses Universitaires de France, Paris, France*, 1980.

[32] X. Boyen et D. Koller : Tractable inference for complex stochastic processes. *Proceedings of the 14th Uncertainty in Artificial Intelligence (UAI)*, 1998.

[33] G. Bradski, T. Darrell, I. Essa, J. Malik, P. Perona, S. Sclaroff et C. Tomasi : http ://sourceforge.net/projects/opencvlibrary/, 2006.

[34] C. Busso, Z. Deng, S. Yildirim, M. Bulut, C.M. Lee, A. Kazemzadeh, S. Lee, U. Neumann et S. Narayanan : Analysis of emotion recognition using facial expressions, speech and multimodal information. *Proc. 6th International Conference on Multimodal Interfaces (ICMI)*, pages 205–211, 2004.

[35] J. Cacioppo et L. Tassimary : Inferring psychological significance from physiological signals. *American Psychologist*, 45:16–28, 1990.

[36] R. Caloz, F.J. Bonn, C. Collet et G. Rochon : *Précis de télédétection.* PUQ 2001 ISBN 2760511456 9782760511453, 2001.

[37] G. Caridakis, L. Malatesta, L. Kessous, N. Amir, A. Paouzaiou et K. Karpouzis : Modeling naturalistic affective states via facial and vocal expression recognition. *Proc. Eighth ACM Int'l Conf. Multimodal Interfaces*, 2006.

[38] G. Chanel, K. Ansari-Asl et T. Pun : Valence-arousal evaluation using physiological signals in an emotion recall paradigm. *IEEE International Conference on Systems, Man and Cybernetics, 2007. ISIC*, 2007.

[39] C.C. Chang et C.J. Lin : Libsvm : library for support vector machines. 2001.

[40] L. S. Chen et T. S. Huang : Emotional expressions in audiovisual human computer interaction. *ICME-2000*, pages 423–426, 2000.

[41] L.S. Chen, T.S. Huang, T. Miyasato et R. Nakatsu : Multimodal human emotion/expression recognition. *Third IEEE International Conference on Automatic Face and Gesture Recognition*, 1998.

[42] Y. Cheon et D. Kim : A natural facial expression recognition using differential-aam and k-nns. *Multimedia, International Symposium on*, 0:220–227, 2008.

[43] P. Chou : Optimal partitioning for classification and regression trees. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 13:340–354, 1991.

[44] V. Christophe : *Les émotions : tour d'horizon des principales théories.* Presses universitaires du Septentrion, 1998.

[45] C. Chuang et F. Y. Shih : Rapid and brief communication : Recognizing facial action units using independent component analysis and support vector machine. *Pattern Recogn.*, 39(9):1795–1798, 2006.

[46] C. Chuang, J. Tsai, C. Wang et P. Chung : Emotion recognition with consideration of facial expression and physiological signals. *IEEE*, 2009.

[47] Z. Chuang et C. Wu : Multi-modal emotion recognition from speech and text. *Computational Linguistics and Chinese Language Processing*, 9(2):45–62, 2004.

[48] C. Clavel : *Analyse et reconnaissance des manifestations acoustiques des émotions de type peur en situations anormales.* Thèse en signal et images, l'École Nationale Supérieure des Télécommunications, 2007.

[49] I. Cohen, F. Cozman, N. Sebe, M. Cirelo et T.S. Huang : Semi-supervised learning of classifiers : Theory, algorithms, and their applications to human-computer interaction. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 26(12):1553–1567, 2004.

[50] I. Cohen, N. Sebe, L. Chen, A. Garg et T.S. Huang : Facial expression recognition from video sequences : temporal and static modelling. *Computer Vision and Image Inderstanding : Special issue on face recognition*, 91:160–187, 2003.

[51] J. Colletta et A. Tcherkassof : *Les émotions : cognition, langage et développement.* Pierre Mardaga, 2003.

[52] E. Couzon et F. Dorn : *Les émotions : développer son intelligence émotionnelle.* Issy-les-Moulineaux : ESF éditeur, 2009.

[53] J. Novovi Cova, P. Somol, M. Haindl et P. Pudil : Conditional mutual information based feature selection for classification task. *LNCS, Springer-Verlag*, 4756, 2007.

[54] R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz et J. G. Taylor : Emotion recognition in human computer interaction. *IEEE Signal Processing Magazine*, 18, 2001.

[55] R. Damasio : Le sentiment même de soi, corps, émotions, conscience. *Editions Odile Jacob, Paris*, 1999.

[56] C. Darwin : *The expression of the emotions in man and animals.* Londres : John Murray, 1872.

[57] M. Dash, K. Choi, P. Scheuermann et H. Liu. : Feature selection for clustering-a filter solution. *In 2nd International Conference on Data Mining*, 2002.

[58] D. Datcu et L. Rothkrantz : Facial expression recognition in still pictures and videos using active appearance models : a comparison approach. *In CompSysTech '07 : Proceedings of the 2007 international conference on Computer systems and technologies*, pages 1–6, New York, NY, USA, 2007. ACM.

[59] D. Datcu et L.J.M. Rothkrantz : Automatic bi-modal emotion recognition system based on fusion of facial expressions and emotion extraction from speech. *8th IEEE International Conference on Automatic Face and Gesture Recognition, 2008. FG '08*, 2008.

[60] R. Davidson, P. Ekman, P. Saron, C. Senulis et J. Friesen : Approach/withdrawal and cerebral asymmetry : Emotional expression and brain physiology. *Journal of personality and social Psychology*, 58, 1990.

[61] T. Dean et K. Kanasawa : A model for reasoning about persistence and causation. *Computational Intelligence*, 5, 1989.

[62] F. Dellaert, T. Polzin et A. Waibel : Recognizing emotion in speech. *Fourth International Conference on Spoken Language ICSLP 96*, 3:1970–1973, 1996.

[63] P. Demartines : *Analyse de données par réseau de neurones auto-organisés*. Thèse de doctorat, Institut National Polytechnique de Grenoble, Laboratoire TIRF, 1994.

[64] P. Demartines et J. Herault : Curvilinear component analysis : A selforganizing neural network for nonlinear mapping of data sets. *IEEE Transactions on Neural Networks*, 8(1):148–154, 1997.

[65] X. Deng, C. H. Chang et E. Brandle : A new method for eye extraction from facial image. *Proc. 2nd IEEE international workshop on electronic dessign, test and applications (DELTA) Australia*, 2:29–34, 2004.

[66] S. Dubuisson, F. Davoine et M. Masson : A solution for facial expression representation and recognition. *Signal Processing : Image Communication*, 17:657–673, 2002.

[67] R.P.W. Duin et M. Loog : Linear dimensionality reduction via a heteroscedastic extension of lda : the chernoff criterion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26:732–739, 2004.

[68] V. Durin : Evaluation indirecte de la qualité vocale perçue. *Mémoire pour le master Science et technologie de l'UPMC*, 2005.

[69] P. Ekman : The argument and evidence about universals in facial expressions of emotion. *H. Wagner and A. Manstead (Eds.), Handbook of social psychophysiology*, 1989.

[70] P. Ekman : *Basic emotions*. Handbook of cognition and emotion, 1999.

[71] P. Ekman : *Facial Expression, The Handbook of Cognition and Emotion*. John wiley et sons édition, 1999.

[72] P. Ekman : Darwin, deception, and facial expression. *Annals of the New York Academy of Sciences*, 1000, 2003.

[73] P. Ekman et R. J. Davidson : Voluntary smiling changes regional brain activity. *Psychological Science*, 4(5):342–345, 1993.

[74] P. Ekman, R. J. Davidson et W. V. Friesen : The duchenne smile : emotional expression and brain physiology. *II. Journal of personality and social psychology*, 58(2):342–353, 1990.

[75] P. Ekman et W. Friesen : Facial action coding system : A technique for the measurement of facial movement palo alto calif. *Consulting psychologists press*, 1978.

[76] P. Ekman et W. V. Friesen : Facial action coding system. *Consulting Psychologist Press*, 18:881–905, 1978.

[77] P. Ekman, W. V. Friesen et P. Ellsworth : What emotion categories or dimensions can observers judge from facial behavior ? *In P. Ekman (Ed.), Emotion in the human face . New York : Cambridge University Press*, 1982.

[78] D. G. Elmes, B. H. Kantowitz et H. L. Roediger : Research miethods in psychology. *Eighth ed. Belmont, CA : Thomson- Wadsworth*, 2006.

[79] I. Essa et A. Pentland : Coding, analysis, interpretation, recognition of facial expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):757–763, 1997.

[80] J.M. Bouroche et G. Saporta : L'analyse des données, pour la science. *Les mathématiques sociales*, 1999.

[81] N. Eveno : *Segmentation des l'evres par un modèle dŽeformable analytique.* spécialité signal image parole télécoms, LIS laboratory Grenoble France, 2003.

[82] C. Fan, H. Sarrafzadeh, F. Dadgostar1 et H. Gholamhosseini : Facial expression analysis by support vector regression. 2005.

[83] B. Fasel, F. Monay et D. Gatica-Perez : Latent semantic analysis of facial action codes for automatic facial expression recognition. *Proc. Sixth ACM Int'l Workshop Multimedia Information Retrieval (MIR '04)*, 2004.

[84] R.A. Fisher : The use of multiple measurements in taxonomic problems. *In Annals of Eugenics*, 7:179–188, 1936.

[85] F. Fleuret : Fast binary feature selection with conditional mutual information. *Journal of Machine Learning Research*, 5:1531–1555, 2004.

[86] F. Fragopanagos et J.G. Taylor : Emotion recognition in human-computer interaction. *Neural Networks*, 18, 2005.

[87] N. H. Frijda : The emotions. *Cambridge : Cambridge University Press*, 1986.

[88] K. S. Fu et A. Rosenfeld : Pattern recognition and image processing. *IEEE trans. On Computers*, 25(12):1336–1346, 1976.

[89] Y. Gao et M.K.H. Leung : Human face recognition using line edge maps. *Proc. IEEE 2nd Workshop Automatic Identification Advanced Technology*, pages 173–176, 1999.

[90] Y. Gao, M.K.H Leung, S.C. Hui et M.W. Tananda : Facial expression recognition from line-based caricatures. *IEEE Transaction on System Man and Cybernetics -PART A : System and Humans*, 33, 2003.

[91] S. S. Ge, C. Wang et C. C. Hang : Facial expression imitation in human robot interaction. *IEEE RO-MAN08, The 17th International Symposium on Robot and Human Interactive Communication, Germany*, 2008.

[92] H.J. Go, K.C. Kwak, D.J. Lee et M.G. Chun : Emotion recognition from facial image and speech signal. *Int. conf. of the society of instrument and control ingineers*, 2003.

[93] J. Godefroid : Psychologie : Science humaine et science cognitive. *de boeck*, 2008. ISBN 978-2-8041-5901-6.

[94] H. Gunes et M. Piccardi : Affect recognition from face and body : Early fusion versus late fusion. *Proc. IEEE Int'l Conf. Systems, Man, and Cybernetics (SMC'05)*, 2005.

[95] M. Gurban : *Multimodal Feature Extraction and Fusion for Audio-visuel Speech Recognition.* thèse doctorale en informatique, communications et information, Ecole polythechnique fédérale de lausanne Suisse, 2009.

[96] Z. Hammal : *Segmentation des traits du visages, analyse et reconnaissance des expressions faciales par le modèle de croyance transférable.* Thèse de doctorat, Université Joseph Fourier de Grenoble, 2006.

[97] D.J. Hand : Discrimination and classification. *John Wiley and Sons*, 1981.

[98] B. Herbelin, P. Benzaki, F. Riquier, O. Renault et D. Thalmann : Using physiological measures for emotional assessment : a computer-aided tool for cognitive and behavioural therapy. *5th Int. Conf on Disability*, 2004.

[99] S. Hoch, F. Althoff, G. McGlaun et G. Rigoll : Bimodal fusion of emotional data in an automotive environement. *ICASSP*, 2, 2005.

[100] G. Holmes et C. G. Nevill-Manning : Feature selection via the discovery of simple classification rules. *In Proceedings of the International Symposium on Intelligent Data Analysis (IDA'95)*, 1995.

[101] C.L. Huang et Y.M. Huang : Facial expression recognition using model-based feature extraction and action parameters classification. *Visual Communication and Image Representation*, 8:278–290, 1997.

[102] T. S. Huang, L. S. Chen, H. Tao, T. Miyasato et R. Nakatsu : Bimodal emotion recognition by man and machine. *In ATR Workshop on Virtual Communication Environments*, 1998.

[103] N. Idrissi : *La navigation dans les bases d'images : prise en compte des attributs de texture*. Thèse de doctorat, Ecole Polytechnique De L'université De Nantes, 2008.

[104] S. Ioannou, A. Raouzaiou, V. Tzouvaras, T. Mailis, K. Karpouzis et S. Kollias : Emotion recognition through facial expression analysis based on a neurofuzzy method. *Neural Networks*, 18, 2005.

[105] C. E. Izard : Human emotions. *New York : Plenum Press*, 1977.

[106] F. Jensen, S. Lauritzen et K. Olsen : Bayesian updating in recursive graphical models by local computations. *In Computational Statisticals Quarterly*, 4, 1990.

[107] F. V. Jensen : Bayesian networks and decision graphs. *Statistics for Engineering and Information Science, Springer, Berlin, Heidelberg*, 2001.

[108] B. Johnson et L. Christensen : Educational research quantitative, qualitative, and mixed approaches. *2nd ed. Boston, MA : Pearson Education, Inc.*, 2004.

[109] R. El Kaliouby et P. Robinson : Real-time inference of complex mental states from facial expression and head gestures. *IEEE Int'l Conf. Computer Vision and Pattern Recognition (CVPR '04)*, 3, 2004.

[110] T. Kanade, J. F. Cohn et Y. Tian : Comprehensive database for facial expression analysis. *Fourth IEEE International Conference on Automatic Face and Gesture Recognition Grenoble France*, FG'00:46–53, 2000.

[111] M. Kapmann et L. Zhang : Estimation of eye, eyebrow and nose features in videophone sequences. *International Workshop on Very Low Bitrate Video Coding (VLBV 98)*, pages 101–104, 1998.

[112] K. Karpouzis, G. Caridakis, L. Kessous, N. Amir, A. Raouzaiou, L. Malatesta et S. Kollias : Modeling naturalistic affective states via facial, vocal, and bodily expression recognition. *LNAI 4451*, 2007.

[113] M. Kass, A. Withins et D. Tersopolos : Snakes : Actives contours models. *International Journal of computer vision*, pages 321–331, 1987.

[114] K. Kenji et A. R. Larry : A practical approach to feature selection. *In Proceedings of the 9th international workshop on Machine learning (ML'92),Aberdeen, Scotland,Morgan Kaufmann Publishers*, 1992.

[115] C. D. Kidd et C. Breazeal : Human-robot interaction experiments : Lessons learned. *AISB'05 Symposium Robot Companions : HardProblems and Open Challenges in Robot-Human Interaction, Hatfield, Hertfordshire, UK*, 2005.

[116] Eunju Kim, Wooju Kim et Yillbyung Lee : Combination of multiple classifiers for the customer's purchase behavior prediction. *Decis. Support Syst.*, 34(2):167–175, 2003.

[117] J. Kim, E. André, M. Rehm, T. Vogt et J. Wagner : Integrating information from speech and physiological signals to achieve emotional sensitivity. *In Interspeech 2005-Eurospeech*, pages 809–812, Lisbon, Portugal, 4-8 September, 2005.

[118] K. H. Kim, S. W. Bang et S. R. Kim : Emotion recognition system using short-term monitoring of physiological signals. *Medical and Biological Engineering and Computing*, 42, 2004.

[119] J-G Ko, K-N Kim et R. S. Ramakrishma : Facial feature tracking for eye-head controlled human computer interface. *IEEE TENCON Korea*, 1999.

[120] R. Kohavi et G. John. : Wrappers for feature subset selection. *Artificial Intelligence*, 1997.

[121] C. Krier, D. Francois, V. Wertz et M. Verleysen : Feature scoring by mutual information for classification of mass spectra. *In 7th International FLINS Conference on Applied Artificial Intelligence (FLINS'06), Genova, Italy*, 2006.

[122] D. Kulic et E. A. Croft : Estimating intent for human-robot interaction. *11th International Conference on Advanced Robotics (ICAR2003), Coimbra, Portugal*, 2003.

[123] N. Kumar et A.G. Andreou : Heteroscedastic discriminant analysis and reduced rank hmms for improved speech recognition. *Speech Communication*, 26:283–297, 1998.

[124] N. Kwak et C.H. Choi : Input feature selection by mutual information based on parzen window. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(12):1667–1671, 2002.

[125] N. Kwak et C.H. Choi : Input feature selection for classification problems. *IEEE Transactions on Neural Networks*, 13(1):143–159, 2002.

[126] L. Lam et C.Y. Suen : Application of majority voting to pattern recognition : An analysis of its behavior and performance. *IEEE Transactions on Systems, Man Cybernetics*, 27:553–568, 1997.

[127] P. Lang, M. Bradley et B. Cuthbert : International affective picture system (iaps) : Digitized photographs, instruction manual and affective ratings. *Technical report A-6 University of Florida*, 2005.

[128] P. J. Lang, M. M. Bradley et B. N. Cuthbert : Emotion, arousal, valence, and the startle reflex. *The structure of emotion : psychophysiological,cognitive, and clinical aspects*, 1993.

[129] A. Lanitis, C.J. Taylor et T.F. Cootes : An Automatic Face Identification System Using Flexible Appearance Models. *Image and Vision Computing*, 13(5):393–401, 1995.

[130] J.T. Lanzetta et S.P. Orr : Excitatory strength of expressive faces : Effects of happy and fear expressions and context on the extinction of a conditioned fear response. *J Pers Soc Psychol*, 50:190–194, 1986.

[131] S. H. Launois : Exploration des voies aériennes supérieures :avancées récentes (1999-2004). *SPLF*, 2004.

[132] C. Lee et A. Elgammal : Facial expression analysis using nonlinear decomposable generative models. *Proc. Second IEEE Int'l Workshop Analysis and Modeling of Faces and Gestures (AMFG)*, 2005.

[133] C. M. Lee, S. Yildirim, M. Bulut, A. Kazemzadeh, C. Busso, Z. Deng, S. Lee et S.S. Narayanan : Emotion recognition based on phoneme classes. *ICSLP'04*, 2004.

[134] S. Lespinats, M. Verleysen, A. Giron et B. Fertil : Dd-hds : a tool for visualization and exploration of high-dimensional data. *IEEE Transactions on Neural Networks*, 18(5): 1264–1279, 2007.

[135] M.C. Lichtlé et V. Plichon : *La diversité des émotions ressenties dans un point de vente.* cahiers de recherche centre de recherche en marketing de bourgogne, 2003.

[136] R. Lienhart et J. Maydt : An Extended Set of Haar-like Features for Rapid Object Detection. *IEEE ICIP*, 1:900–903, September 2002.

[137] C.L. Lisetti et F. Nasoz : Using noninvasive wearable computers to recognize human emotions from physiological signals. *Journal on applied Signal Processing*, pages 1672–1687, 2004.

[138] G.C. Littlewort, M.S. Bartlett et K. Lee : Faces of pain :automated measurement of spontaneous facial expressions of genuine and posed pain. *Proc. Ninth ACM Int'l Conf. Multimodal Interfaces (ICMI '07)*, 2007.

[139] C. Liu, P. Rani et N. Sarkar : Affective state recognition and adaptation in human-robot interaction : A design approach. *in International Conference on Intelligent Robots and Systems, Beijing, China*, 2006.

[140] H. Liu et H. Motoda : Feature selection for knowledge discovery and data mining. 1998.

[141] B. D. Lucas et T. Kanade : An iterative image registration technique with an application to stereo vision. *Proc. of International Joint Conference on Artificial Intelligence*, 18:674–680, 1981.

[142] M.J. Lyons et S. Akamatsu : Coding facial expressions with gabor wavelets. *Third IEEE International Conference on Automatic Face and Gesture Recognition*, pages 200–205, 1998.

[143] C. Maaoui, A. Pruski et F. Abdat : Emotion recognition for human-machine communication. *IEEE International Conference on Intelligent RObots and Systems*, 2008.

[144] M. Malciu : *Approche orientées modèle pour la capture des mouvements du visage en vision par ordinateur*. Thèse de doctorat, l'université René Descartes Paris V, 2001.

[145] M. Malciu et F. Preteux : Mpeg-4 compliant tracking of facial features in video sequences. *Proc. of International Conference on Augmented, Virtual Environments and 3D Imaging Greece*, 108–111, 2001.

[146] A. Martin : Fusion de classifieurs pour la classification d'images sonar. *Revue Nationale des Technologies de l'Information E5*, 2005.

[147] K. Mase : Recognition of facial expression from optical flow. *IEICE Transactions*.

[148] S. McKenna, S. Gong et Y. Raja : Modelling Facial Colour and Identity with Gaussian Mixtures. *Pattern Recognition*, 31(12):1883–1892, 1998.

[149] A. Mehrabian : Communication without words. *Psychology Today*, 2.4:53–56, 1968.

[150] P. Murphy : *Dynamic Bayesian Networks : Representation,Inference and Learning*. Thèse de doctorat, University of California Berkeley, 2002.

[151] P. Naim, P. Wuillemin, P. Leray, O. Pourret et A. Becker : *Réseaux Bayésiens*. Eyrolles, 2008.

[152] J. Naveteur : Douleur chronique et activité électrodermale. *Springer Dossier : Psychophysique et électrophysiologie*, 21, 2008.

[153] T. L. Nwe, F. S. Wei et L.C. De Silva : Speech based emotion classification. *Proceedings of IEEE Region 10 International Conference on Electrical and Electronic Technology*, 1: 297–301, 2001.

[154] A. Ortony et T.Turner : *what's basic about basic emotions*, volume 97. Psycological review, 1990.

[155] P. Pal, A.N. Iyer et R.E. Yantorno : Emotion detection from infant facial expressions and cries. *Proc. IEEE Int'l Conf. Acoustics, Speech and Signal Processing (ICASSP'06)*, 2, 2006.

[156] M. Paleari, R. Benmokhtar et B. Huet : Evidence theory-based multimodal emotion recognition. *Springer-Verlag Berlin Heidelberg 2009 ,(Eds.) : MMM 2009, LNCS 5371*, 2009.

[157] M. Pantic et I. Patras : Detecting facial actions and their temporal segmentation in nearly frontal-view face image sequences. *Proc IEEE International Conference on Systemsn Man and Cybernetics Hawaii*, 2005.

[158] M. Pantic et I. Patras : Dynamics of facial expression : Recognition of facial actions and their temporal segments from face profile image sequences. *IEEE Transaction on System Systemsn, Man, and Cybernetics*, 36, 2006.

[159] M. Pantic et L. J. M. Rothkrantz : Expert system for automatic analysis of facial expressions. *Image and Vision Computing Journal*, 18:881–905, 2000.

[160] M. Pantic et L.J.M. Rothkrantz : Toward an affect-sensitive multimodal human-computer interaction. *Proceedings of the IEEE*, 91:1370–1390, 2003.

[161] M. Pantic et L.J.M. Rothkrantz : Facial action recognition for facial expression analysis from static face images. *IEEE Trans. Systems, Man, and Cybernetics Part B*, 34(3):1449–1461, 2004.

[162] M. Pardas et A. Bonafonte : Facial animation parameters extraction and expression detection using hmm. *Signal Processing : Image Communication*, 17:675–688, 2002.

[163] M. Pardas et E. Sayrol : Motion estimation based tracking of active contours. *Pattern recognition letters*, 22:1447–1456, 2001.

[164] K. Pearson : On lines and planes of closest fit to systems of points in space. *In Philosophical Magazine*, 2:559–572, 1901.

[165] M. Pei, E. D. Goodman et W. F. Punch : Feature extraction using genetic algorithms. *In Proceeding of International Symposium on Intelligent Data Engineering and Learning98 (IDEAL98),Hong Kong*, 1998.

[166] H. Peng, F. Long et C. Ding : Feature selection based on mutual information : Criteria of max-dependency, max-relevance, and min-redundancy. *on Pattern Analysis and Machine Intelligence*, 27(8):1226–1238, 2005.

[167] S. Petridis et M. Pantic : Audiovisual discrimination between laughter and speech. *IEEE Int'l Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, 2008.

[168] R. W. Picard : Affective computing. *MIT Press, Cambridge, MA*, 1997.

[169] R. W. Picard, E. Vyzas et J. Healey : Toward machine emotional intelligence : analysis of affective physiological state. *IEEE Transaction Pattern Analysis*, 23:1175–1191, 2001.

[170] R. Plutchik : A general psychoevolutionary theury of emotion, 1980.

[171] J. Posner, J. A. Russell et B. S. Peterson : The circumplexe model of affect : An interactive approach to affective neuroscience, cognitive developement, and psychopathologie. *Developement and psychopatologie*, 17(3):715–734, 2005.

[172] Procomp : www.thoughttechnology.com/proinf.htm.

[173] F. Prêteux : On a distance function approach for grey-level mathematical morphology. *E. R. Dougherty Eds Mathematical Morphology in Image Processing*, 1992.

[174] P. Pudil, J. Novovicova et J. Kittler : Floating search methods in feature selection. *Pattern Recognition Letter*, 15(11):1119–1125, 1994.

[175] P. Radeva et E. Marti : Facial features segmentation by model-based snakes. *Proc. International Conference on Computer Analysis and Image Processing*, pages 515–520, 1995.

[176] P. Rani, N. Sarkar, C. A. Smith et J. A. Adams : Affective communication for implicit human-machine interaction. *IEEE International Conference on Systems, Man, and Cybernetics*, 2003.

[177] P. Rani, J. Sims, R. Brackin et N. Sarkar : Online stress detection using psychophysiological signals for implicit human-robot cooperation. *Robotica*, 20, 2002.

[178] C.R. Rao : The utilization of multiple measurements in problems of biological classification. *In Journal of the Royal Statistical Society*, 10:159–203, 1948.

[179] A. Rivière et B. Godet : L'affective computing : rôle adaptatif des émotions dans l'interaction homme - machine. *Travail d'Etude et de Recherche (TER) Maîtrise de sciences cognitives*, 2003.

[180] M. Robnik-Sikonja et I. Kononenko : Theoretical and empirical analysis of relieff and rrelieff. *Machine Learning*, 53(1), 2003.

[181] H. Rowley, S. Baluja et T. Kanade : Neural Network-based Face Detection. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 20(1):23–38, 1998.

[182] J. Russell : A circumplex model of affect. *Journal of personality and social Psychology*, 39, 1980.

[183] J.A. Russell et G. Pratt : A description of the effective quality attributed to environments. *Journal of Personality and Social Psychology*, 38, 1980.

[184] B. Sahiner, H.P. Chan, N. Petrick, R. F. Wagner et L. M. Hadjiiski : Stepwise linear discriminant analysis in computer-aided diagnosis : the effect of finite sample size. *Proceedings of SPIEMedical Imaging : Image Processing*, 3661, 1999.

[185] R. E. Schapire et Y. Singer : Improved boosting algorithms using confidence-rated predictions. *Machine Learning.*, 37(3):297–336, 1999.

[186] J. Scheirer, R. Fernandez, J. Klein et R.W. Picard : Frustrating the user on purpose : a step toward building an affective computer. *In Interacting with Computers*, 14:93–118, 2002.

[187] K. R. Scherer : On the nature and function of emotion : A component process approach. *Lawrence Erlbaum Associates, Publishers, Londres*, 1984.

[188] B. Scholkopf, A. J. Smola et K. Muller : Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*, 10(5):1299–1319, 1998.

[189] N. Sebe, E. Bakker, I. Cohen, T. Gevers et T. S. Huang : Bimodal emotion recognition. *In Proc. 5th International Conference on Methods and Techniques in Behavioral Research*, 2005.

[190] N. Sebe, I. Cohen, T. Gevers et T.S. Huang : Emotion recognition based on joint visual and audio cues. *Proc. 18th Int'l Conf. Pattern Recognition (ICPR '06)*, 2006.

[191] N. Sebe, M.S. Lew, I. Cohen, Y. Sun, T. Gevers et T.S. Huang : Authentic facial expression analysis. *Proc. IEEE Int'l Conf.Automatic Face and Gesture Recognition (AFGR)*, 2004.

[192] C. E. Shannon : A mathematical theory of communication. *Bell System Technical Journal*, 27(3):379–423, 1948.

[193] J. Shi et C. Tomasi : Good features to track. *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 593–600, 1994.

[194] L. C. De Silva et P. C. Ng : Bimodal emotion recognition. *IEEE International Conf. on Automatic Face and Gesture Recognition*, pages 332–335, 2000.

[195] L.C. De Silva, T. Miyasato et R. Nakatsu : Emotion recognition using multimodal information. *IEEE Int. Conf. on Information, Communications and Signal Processing (ICICS'97)*, 1997.

[196] L.C. De Silva et P. C. Ng : Bimodal emotion recognition. *Fourth IEEE International Conference on Automatic Face and Gesture Recognition*, 2000.

[197] R. Sinha et O.A. Parsons : Multivariate response patterning of fear and anger. *Cognition and Emotion*, 10:173–198, 1996.

[198] M. Song, J. Bu, C. Chen et N. Li : Audio-visual-based emotion recognition : A new approach. *Proc. Int'l Conf. Computer Vision and Pattern Recognition (CVPR '04)*, 2004.

[199] G. Stemmler, M. Heldmann, C.A. Pauls et T. Scherer : Constraints for emotion specificity in fear and anger : the context counts. psychophysiology. 38, 2001.

[200] K. Takahashi : Remarks on emotion recognition from multi-modal bio-potential signals. *IEEE International Conference on Industrial Technology*, 2004.

[201] M. Tekalp : Face and 2d mesh animation in mpeg-4. *Tutorial Issue on the MPEG-4 Standard Image Communication Journal Elsevier*, 1999.

[202] Y. Tian, T. Kanade et J. Cohn : Dual state parametric eye tracking. *Proc. 4th IEEE International Conference on Automatic Face and Gesture Recognition*, pages 110–115, 2000.

[203] Y. Tian, T. Kanade et J. Cohn : Recognizing lower face action units for facial expression analysis. *Proceedings of the 4th IEEE International Conference on Automatic Face and Gesture Recognition (FG'00)*, 2000.

[204] Y. Tian, T. Kanade et J.F. Cohn : Recognizing action units for facial expression analysis. *Trans IEEE Pattern Analysis and Machine Intelligence*, 23:97–115, 2001.

[205] S. S. Tomkins : Affect theory. *In K. R. Scherer, P. Ekman (Eds.) Approaches to emotion, Hillsdale, NJ : Erlbaum.*, 1984.

[206] Y. Tong, W. Liao et Q. Ji : Facial action unit recognition by exploiting their dynamics and semantic relationships. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 29(10):1683–1699, 2007.

[207] N. Tsapatsoulis, K. Karpouzis, G. Stamou, F. Piat et S. Kollias : A fuzzy system for emotion classification based on the mpeg-4 facial definition parameter set. *Proc.10th European Signal Processing Conference Finland*, 2000.

[208] J. Valat : Le comportement émotionnel. *Licence de Psychologie, Université Montpellier II*, 2008.

[209] M.F. Valstar, H. Gunes et M. Pantic : How to distinguish posed from spontaneous smiles using geometric features. *Proc. Ninth ACM Int'l Conf. Multimodal Interfaces (ICMI'07)*, 2007.

[210] L. Vandendorpe : Elec 2900 traitement des signaux. *UniversitŽe catholique de Louvain.*

[211] V.N. Vapnik : an overview of statistical learning theory. *IEEE Trans. Neutral Netw.*, 10:988–999, 1999.

[212] V. Vezhnevets et A. Degtiareva : Robust and accurate eye contour extraction. *Proc. GraphicsCon Russia*, pages 81–84, 2003.

[213] O. Villon et C.L. Lisetti : Toward building adaptive users psycho-physiological maps of emotions using bio-sensors. *Proceedings of KI*, 2006.

[214] P. Viola et M. Jones : Robust real-time object detection. *2nd international workshop on statistical and computational theories of vision - modeling, learning, computing, and sampling vancouver, canada*, 2001.

[215] S. R. Vrana, B. N. Cuthbert et P. J. Lang : Fear imagery and text processing. *Psychophysiology*, 23:247–253, 1986.

[216] F. Wallhoff : Feedtum : Facial expressions and emotion database, 2005.

[217] J. Wang, L. Yin, X. Wei et Y. Sun : 3d facial expression recognition based on primitive surface feature distribution. *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition (CVPR'06)*, 2, 2006.

[218] Y. Wang et L. Guan : Recognizing human emotion from audiovisual information. *Proc. Int'l Conf. Acoustics, Speech, and Signal Processing (ICASSP'05)*, 2005.

[219] Y. Y. Wang et J. Li : Feature-selection ability of the decision-tree algorithm and the impact of feature-selection/extraction on decision-tree results based on hyperspectral data. *International Journal of Remote Sensing*, 29, 2008.

[220] P. Weber et L. Jouffe : Reliability modelling with dynamic bayesian networks. *the 5th IFAC symposium on fault detection, supervision and safety of technical processes, SafeProcess*, 2003.

[221] J. Whitehill et C.W. Omlin : Haar features for facs au recognition. *Proc. IEEE Int'l Conf. Automatic Face and Gesture Recognition (AFGR'06)*, 2006.

[222] F. Xiaoyi, L. Baohua, L. Zhen et Z. Jiling : Automatic facial expression recognition with aam-based feature extraction and svm classifier. *MICAI 2006 : Advances in Artificial Intelligence*, 2006.

[223] B.V. Xu, A. Krzyzak et C.Y. Suen : Methods of combining multiple classifiers and their application to handwriting recognition. *IEEE Transactions on Systems, Man Cybernetics*, 22:418–435, 1992.

[224] Q. Xu, P. Zhang, L. Yang, W. Pei et Z. He : A facial expression recognition approach based on novel support vector machine tree. *In ISNN '07 : Proceedings of the 4th international symposium on Neural Networks*, pages 374–381, Berlin, Heidelberg, 2007. Springer-Verlag.

[225] Y. Yacoob et L. Davis : Computing spatio-temporal representations of human faces. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition CVPR '94*, 1994.

[226] G. Yang et T. S. Huang : Human Face Detection in Complex Background. *Pattern Recognition*, 27(1):53–63, 1994.

[227] J. Yang et V. Honavar. : Feature subset selection using a genetic algorithm. *In IEEE Intelligent Systems*, 1998.

[228] M. Yeasin, B. Bullot et R. Sharma : Recognition of facial expressions and measurement of levels of interest from video. *IEEE Trans. Multimedia*, 8(3):500–507, 2006.

[229] Y. Yoshitomi, Sung-Ill Kim, T. Kawano et T. Kilazoe : Effect of sensor fusion for recognition of emotional states using voice, face image and thermal image of face. *9th IEEE International Workshop on Human Interactive Communication RO-MAN 2000*, 2000.

[230] L. Yu et H. Liu : Feature selection for high-dimensional data : A fast correlation based filter solution. *In 20th International Conference on Machine Learning*, 2003.

[231] A.L. Yuille, P.W. Hallinan et D.S. Cohen : Feature extraction from faces using deformable templates. *International Journal of Computer Vision*, 8:99–111, 1992.

[232] Z. Zeng, Y. Fu, G.I. Roisman, Z. Wen, Y. Hu et T.S. Huang : Spontaneous emotional facial expression detection. *J. Multimedia*, 1(5):1–8, 2006.

[233] Z. Zeng, Y. Hu, M. Liu, Y. Fu et T.S. Huang : Training combination strategy of multi-stream fused hidden markov model for audio-visual affect recognition. *ACM MM*, 2006.

[234] Z. Zeng, J.Tu, M. Lu, T. S. Huang, B. Pianfetti, D. Roth et S. Levinson : Audio-visuel affect recognition. *IEEE Transactions on Multimedia*, 9(2), 2007.

[235] Z. Zeng, M. Pantic, G. I. Roisman et T. S. Huang : A survey of affect recognition methods : audio, visual and spontaneous expressions. *In ICMI '07 : Proceedings of the 9th international conference on Multimodal interfaces*, pages 126–133, New York, NY, USA, 2007. ACM.

[236] Z. Zeng, J. Tu, M. Liu, T.S. Huang, B. Pianfetti, D. Roth et S.Levinson : Audio-visual affect recognition. *IEEE Trans. Multimedia*, 9(2):424–428, 2007.

[237] Z. Zeng, J. Tu, M. Liu, T. Zhang, N. Rizzolo, Z. Zhang, T. S. Huang, D. Roth et S. Levinson : Bimodal hci-related affect recognition. *ICMI*, 2004.

[238] Z. Zeng, J. Tu, P. Pianfetti, M. Liu, T. Zhang, Z. Zhang, T.S. Huang et S. Levinson : Audio-visual affect recognition through multi-stream fused hmm for hci. *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition (CVPR'05)*, 2005.

[239] Z. Zeng, Z. Zhang, B. Pianfetti, J. Tu et T.S. Huang : Audio-visual affect recognition in activation-evaluation space. *Proc.13th ACM Int'l Conf. Multimedia (Multimedia'05)*, 2005.

[240] L. Zhang : Estimation of eye and mouth corner point positions in a knowledge based coding system. *Proc. SPIE Digital Compression Technologies and Systems for Video Communications*, pages 21–28, 1996.

[241] Y. Zhang et Q. Ji : Active and dynamic information fusion for facial expression understanding from image sequences. *IEEE Trans. Pattern Anal. Mach. Intell.*, 27(5):699–714, 2005.

[242] Y. Zhu et S. C. Schwartz : Discriminant analysis and adaptive wavelet feature selection for statistical object detection. *ICPR*, 4, 2002.

Titre :

Reconnaissance automatique des émotions par données multimodales : expressions faciales et signaux physiologiques

Résumé :

Cette thèse présente une méthode générique de reconnaissance automatique des émotions à partir d'un système bimodal basé sur les expressions faciales et les signaux physiologiques. Cette approche de traitement des données conduit à une extraction d'information de meilleure qualité et plus fiable que celle obtenue à partir d'une seule modalité.
L'algorithme de reconnaissance des expressions faciales qui est proposé, s'appuie sur la variation de distances des muscles faciaux par rapport à l'état neutre et sur une classification par les séparateurs à vastes marges (SVM). La reconnaissance des émotions à partir des signaux physiologiques est, quant à elle, basée sur la classification des paramètres statistiques par le même classifieur.
Afin d'avoir un système de reconnaissance plus fiable, nous avons combiné les expressions faciales et les signaux physiologiques. La combinaison directe de telles informations n'est pas triviale étant donné les différences de caractéristiques (fréquence, amplitude de variation, dimensionnalité). Pour y remédier, nous avons fusionné les informations selon différents niveaux d'application. Au niveau de la fusion des caractéristiques, nous avons testé l'approche par l'information mutuelle pour la sélection des plus pertinentes et l'analyse en composantes principales pour la réduction de leur dimensionnalité. Au niveau de la fusion de décisions, nous avons implémenté une méthode basée sur le processus de vote et une autre basée sur les réseaux Bayésien dynamiques. Les meilleurs résultats ont été obtenus avec la fusion des caractéristiques en se basant sur l'Analyse en Composantes Principales.
Ces méthodes ont été testées sur une base de données conçue dans notre laboratoire à partir de sujets sains et de l'inducteur par images IAPS. Une étape d'auto évaluation a été demandée à tous les sujets dans le but d'améliorer l'annotation des images d'induction utilisées. Les résultats ainsi obtenus mettent en lumière leurs bonnes performances et notamment la variabilité entre les individus et la variabilité de l'état émotionnel durant plusieurs jours.

Mots clés :

Reconnaissance des émotions, expression faciale, signaux physiologiques, fusion de caractéristiques, fusion de décisions, SVM, RBD, ACP, Information mutuelle, vote.

Title :

Automatic emotion recognition from multimodal data : facial expressions and physiological signals

Abstract :

This thesis presents a generic method for automatic recognition of emotions from a bimodal system based on facial expressions and physiological signals. This data processing approach leads to better extraction of information and is more reliable than single modality.
The proposed algorithm for facial expression recognition is based on the distance variation of facial muscles from the neutral state and on the classification by means of Support Vector Machines (SVM). And the emotion recognition from physiological signals is based on the classification of statistical parameters by the same classifier. In order to have a more reliable recognition system, we have combined the facial expressions and physiological signals. The direct combination of such information is not trivial giving the differences of characteristics (such as frequency, amplitude, variation, and dimensionality). To remedy this, we have merged the information at different levels of implementation. At feature-level fusion, we have tested the mutual information approach for selecting the most relevant and principal component analysis to reduce their dimensionality. For decision-level fusion we have implemented two methods; the first based on voting process and another based on dynamic Bayesian networks. The optimal results were obtained with the fusion of features based on Principal Component Analysis. These methods have been tested on a database developed in our laboratory from healthy subjects and inducing with IAPS pictures. A self-assessment step has been applied to all subjects in order to improve the annotation of images used for induction. The obtained results have shown good performance even in presence of variability among individuals and the emotional state variability for several days.

Keywords :

Emotion recognition, facial expression, physiological signals, feature fusion, decision fusion, SVM, DBN, PCA, Mutual information, vote.