



HAL
open science

Reconnaissance automatique de l'émotion à partir de signaux EEG

Anne-Claire Conneau

► **To cite this version:**

Anne-Claire Conneau. Reconnaissance automatique de l'émotion à partir de signaux EEG. Imagerie médicale. Télécom ParisTech, 2016. Français. NNT : 2016ENST0033 . tel-02957939

HAL Id: tel-02957939

<https://pastel.hal.science/tel-02957939>

Submitted on 5 Oct 2020

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

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



EDITE - ED 130

Doctorat ParisTech

THÈSE

pour obtenir le grade de docteur délivré par

Télécom ParisTech

Spécialité "SIGNAL et IMAGES"

présentée et soutenue publiquement par

Anne-Claire CONNEAU

le 09 juin 2016

Reconnaissance automatique de l'émotion à partir de signaux EEG

Directeur de thèse : **Gaël RICHARD**

Co-directeur de thèse : **Slim ESSID**

Jury

M. Mohamed Chetouani,
M. Jean-Claude Martin,
Mme Catherine Pelachaud,
Mme Magalie Ochs,

Professeur, UPMC
Professeur, Université Paris-Sud
Directrice de recherches, CNRS-LTCI
Maître de Conférences, PolyTech Marseille
Aix-Marseille Université

Rapporteur
Rapporteur
Présidente du jury
Examinatrice

TELECOM ParisTech

Ecole de l'Institut Mines-Télécom - Membre de ParisTech

46 rue Barrault 75013 Paris - (+33) 1 45 81 77 77 - www.telecom-paristech.fr

A Françoise Conneau,

Remerciements

Une thèse ne s'élabore pas en solitaire mais elle repose sur un encadrement de qualité, ce dont j'ai eu la chance de bénéficier. C'est pourquoi mes premiers remerciements vont tout d'abord à mon co-directeur de thèse, Slim Essid, qui m'a apporté un soutien constant tout au long de la thèse. Ses conseils précieux et sa confiance ont fortement contribué à son bon déroulement. Je lui suis reconnaissante pour son écoute et sa disponibilité. Je tiens également à remercier mon directeur de thèse, Gaël Richard, qui a su se montrer présent dans ces moments de difficulté qui font parti du parcours d'une thèse.

Je tiens à remercier tout particulièrement les membres du jury pour l'intérêt qu'ils ont porté à mes travaux et pour le temps et le travail qu'ils ont consacré pour cette soutenance : Mohamed Chetouani et Jean-Claude Martin, rapporteurs, Catherine Pelachaud, présidente du jury et Magalie Ochs, examinatrice. Merci aux rapporteurs pour leur lecture attentive et l'ensemble de leurs remarques sur le manuscrit, et à Catherine pour son regard bienveillant sur mes travaux et de manière générale depuis le début de la thèse.

Je remercie Chloé Clavel pour son aide sur les stimuli du corpus et plus particulièrement sur les vidéos du corpus SAFE. Merci à Mohamed Chetouani pour l'intérêt qu'il a porté au corpus et Ayoub Hajlaoui pour ses travaux sur le corpus EMOGEE. Je remercie également toutes les personnes qui, de près ou de loin, ont contribué à la mise en œuvre du corpus.

Je garderais de ces années passées au sein de l'équipe AAO de Télécom ParisTech les souvenirs de belles rencontres dans une ambiance joyeuse. Merci entre autres à Aymeric, François, Mounira, Rémi, Manuel, Benoît, Cécilia, Nico, Clément, Antoine, Angélique, Nicolas, Sébastien, Fabrice.

Le bon déroulement d'une thèse ne se joue pas uniquement sur les heures passées au labo, mais également grâce aux moments agréables passés avec les proches et au soutien de l'entourage. C'est pourquoi je tiens à remercier chaleureusement mes soutiens parisiens Vivien et Maeva pour leur investissement et la curiosité dont ils ont fait preuve tout au long de ces années. Je remercie aussi Audric et Sandra, Raphaël, Chouk et Thomas, Lucie, Maxime et Coline, Corentin et Marine, Renaud, Anne-Sophie, Cooc, Florent pour tous ces moments heureux passés et à venir.

Merci également à la grande famille de Louis, dont le soutien m'a été précieux, et je les remercie d'être venu assister à ma soutenance. J'adresse un merci tout particulier à Martine et Jacques pour m'avoir toujours accueillie chaleureusement.

Je remercie Isa et César pour le soutien et la bienveillance dont ils ont fait preuve et qui m'ont été essentiels dans la dernière ligne droite de la thèse. Avec un merci particulier à Isa pour s'être montrée vivement intéressée par mes travaux lors de la soutenance.

Un grand merci à ma maman pour sa confiance, son soutien et pour ses très nombreuses cartes d'encouragement ainsi qu'à Benjamin, Elisa et au bonheur des sourires d'Aurélie.

Enfin, je ne remerciais jamais assez Louis de m'avoir aidée et soutenue tout au long de cette thèse et d'avoir été en plus un auditeur patient et attentif pour les répétitions de soutenance.

Résumé

Alors que l'utilisation de l'électro-encéphalographie (EEG) a longtemps été confinée au domaine médical, l'intérêt dans les interfaces cerveau-machine basées sur ces signaux s'est développé au cours de ces dernières années pour les applications grand public. Les enregistrements EEG ont tout particulièrement retenu l'attention des chercheurs dans le domaine de l'informatique affective, affective computing, dans le but de réaliser des travaux sur l'analyse du comportement humain, et ce plus particulièrement sur la reconnaissance automatique des émotions. Comparé à d'autres modalités utilisées dans de précédents travaux sur la reconnaissance des émotions, telles que la parole, les expressions du visage, les mouvements et d'autres signaux physiologiques, l'EEG a l'avantage de pouvoir capturer des informations liées à l'état émotionnel interne qui n'est pas forcément traduit par des manifestations extérieures observables.

La reconnaissance des émotions est habituellement envisagée sous l'angle d'un problème de classification où le choix de caractéristiques appropriées est essentiel dans le but de s'assurer une précision de reconnaissance satisfaisante. Une de nos problématiques repose sur le fait que, dans le cadre des caractéristiques EEG, un consensus n'a toujours pas été réalisé sur un ensemble standard de caractéristiques qui permettrait de garantir une distinction performante des émotions d'un sujet humain. Nous explorons une grande variété de caractéristiques temporelles, spectrales et spatiales pouvant être potentiellement utiles dans le cadre de la reconnaissance des émotions et nous les comparons à d'autres, exposées dans de précédents travaux, en utilisant un protocole expérimental rigoureux. Nous évaluons plus particulièrement l'efficacité de plusieurs caractéristiques spectrales, n'ayant pas été précédemment proposées pour le problème de la classification. Nos résultats montrent que les nouvelles caractéristiques spectrales que nous proposons sont compétitives comparées à celles précédemment utilisées. Elles nous amènent de plus vers une configuration mono-canal performante de la reconnaissance des émotions, ce qui implique un fort potentiel pour les applications destinées au grand public.

Au sein des corpus existants et accessibles destinés à la reconnaissance des émotions en informatique affective, l'aspect de la dynamique de l'émotion n'est pas pris en considération. Ces corpus présentent également un manque de variabilité dans les données et ne possèdent les enregistrements que d'un nombre limité de participants. Ces raisons nous ont menées à proposer un corpus multi-modal destiné à l'analyse de l'état émotionnel qui s'attache à répondre au mieux à certaines faiblesses des corpus existants. Nous employons différentes stratégies d'élicitation de l'émotion, par le biais de l'utilisation de stimuli visuels et audio-visuels, et nous proposons également une approche novatrice dans la stratégie

d'annotation de l'émotion ressentie, en intégrant en plus de la retranscription de l'émotion ressentie de manière globale, la retranscription de ses variations.

Ce nouveau corpus permettra dans un premier temps de renforcer la validation de l'approche proposée dans nos travaux. D'autre part, en accord avec les résultats encourageants obtenus, il sera possible d'envisager à plus long terme la caractérisation de la dynamique de l'état émotionnel de l'utilisateur qui ouvre la voie à de nouveaux modèles qui pourraient prédire, par exemple, l'augmentation de l'anxiété de l'utilisateur en fonction de la situation dans laquelle il est placé.

Abstract

While the usage of electroencephalographic (EEG) recording has been for long confined in the medical field, the recent years have seen a growing interest in EEG-based brain-computer interfaces (BCI) for general public applications. In particular EEG-recording has attracted the attention of researchers in the field of affective computing as part of the effort to perform human-behaviour analysis tasks, especially automatic emotion recognition. Compared to other modalities which have been considered in previous work on emotion recognition, such as speech, facial expressions, gestures or other physiological signals, EEG has the advantage of capturing information related to internal emotional states not necessarily resulting in any observable external manifestations.

Emotion recognition is usually approached as a classification problem where the choice of appropriate features is critical to ensure satisfactory recognition accuracy. As far as EEG-features are concerned, a consensus has not yet been reached as to a standard set of attributes that could guarantee a successful characterization of a human-subject's emotions. We explore a wide range of temporal, spectral and spatial features potentially useful for emotion recognition, comparing them to previous proposals through a rigorous experimental evaluation. In particular we assess the effectiveness of various spectral features that were not previously envisaged for the problem of classifying emotions. Our results show that the new spectral shape features that we propose are very competitive compared to previously used ones. They are additionally amenable to successful emotion recognition in single-channel setups, which holds a great potential for general public applications.

Within the existing and accessible datasets for emotion recognition in affective computing, the aspect of the dynamics of emotion is not being considered. Those datasets also show a lack of variability in the data and contain the records of a limited number of participants. These reasons have led us to propose a new multimodal dataset for the analysis of the emotional state that attaches to respond better to certain weaknesses of the existing datasets. We use different strategies for the emotion elicitation, through the use of both visual and audio-visual stimuli. We also provide an innovative approach to the strategy on the annotation of the emotion experienced by the user by integrating more than the transcript of the emotion globally felt, but also the transcript of its variations.

This new dataset will, in a first stage, enhance the validation of the approach we propose in our work. Secondly, in agreement with the encouraging results, it will be possible to consider a longer-term characterization of the dynamics of the emotional state of the user, which opens the way for new models that could predict, for example, the increase in the user's anxiety according to the situation in which he is placed.

Table des matières

1	Introduction	5
1.1	Introduction	6
1.2	Vers un système de reconnaissance automatique des émotions basé sur l'EEG	6
1.3	Objectifs et Contributions	7
1.4	Organisation du document	8
2	La reconnaissance des émotions : Etat de l'art et Préalables	11
2.1	A la recherche de l'émotion	13
2.1.1	Une définition possible? Les théories de l'émotion	13
2.1.2	Le cerveau et les émotions	18
2.1.3	Bilan et discussion	21
2.2	Le point de vue en informatique affective et les différentes modalités pour la reconnaissance des émotions	21
2.2.1	Les signaux physiologiques périphériques	22
2.2.2	Les signaux audio-visuels	25
2.2.3	Les signaux EEG	27
2.2.4	Les systèmes multi-modaux	27
2.3	Une vue d'ensemble des systèmes précédents pour la reconnaissance des émotions basée sur l'EEG	32
2.3.1	Les pré-traitements des signaux EEG	33
2.3.2	Le système de classification	33
2.4	L'expérience des émotions	36
2.4.1	Retranscrire l'émotion	36
2.4.2	Eliciter une émotion	47
2.5	Les corpus existants	50
2.5.1	INTERFACE'06 [Savran et al., 2006]	50
2.5.2	DEAP Dataset [Koelstra et al., 2012]	53
2.5.3	MAHNOB-HCI [Soleymani et al., 2012a]	53
2.5.4	Discussion	54
2.6	Conclusion	55
3	EMOGEE : un nouveau corpus pour la reconnaissance des émotions basée sur l'EEG	57
3.1	Introduction	59
3.2	Stratégie d'élicitation : les stimuli	59

3.2.1	Dimensionnement des stimuli	60
3.2.2	Stimuli visuels statiques	60
3.2.3	Stimuli audio-visuels	61
3.3	L'installation expérimentale	65
3.3.1	Le système matériel	65
3.3.2	Le protocole expérimental	70
3.4	Stratégie d'annotation	74
3.4.1	Annotation statique	74
3.4.2	Annotation dynamique	74
3.5	Les données enregistrées	76
3.6	Analyse statistique des annotations	77
3.6.1	Analyses des annotations sur les stimuli visuels	78
3.6.2	Analyse des annotations sur les stimuli audio-visuels	82
3.6.3	Bilan	84
3.7	Conclusion	84
4	Etude comparative des caractéristiques	87
4.1	Introduction	88
4.2	Les descripteurs existants pour la reconnaissance des émotions	89
4.2.1	Les descripteurs temporels	89
4.2.2	Les descripteurs spectraux	90
4.3	Les descripteurs proposés	91
4.4	Les paramètres d'extraction de caractéristiques	93
4.5	Données et Protocole d'évaluation	94
4.6	Expériences et Résultats	97
4.6.1	Etude comparative des caractéristiques	97
4.6.2	Etudes des descripteurs proposés	99
4.7	Conclusion	102
5	Conclusions et Perspectives	105
5.1	Apports	107
5.2	Perspectives	108
5.2.1	Perspectives à court terme	108
5.2.2	Perspectives à long terme	108
A	Annexes	I
A.1	Les principes de l'électro-encéphalographie	I
A.1.1	Les bases neurophysiologiques, d'après Blinowska and Durka [2006]	I
A.1.2	Les rythmes EEG	II
A.2	Tableaux annexes	III
A.2.1	Extraits issus du corpus SAFE	III
A.2.2	Blocs d'images IAPS issus de ENTERFACE'06	V
A.3	Liste des acronymes	VIII

B Bibliographie	XIII
B.1 Publications de l'auteur	XIII

Chapitre 1

Introduction

Sommaire

1.1 Introduction	6
1.2 Vers un système de reconnaissance automatique des émotions basé sur l'EEG	6
1.3 Objectifs et Contributions	7
1.4 Organisation du document	8

1.1 Introduction

Nous ressentons chaque jour de manière involontaire une grande variété d'émotions. Elles peuvent différer de par leur intensité ou leur caractère positif ou négatif. Ressentir une émotion, c'est un moteur dans la réflexion, l'appréciation ou encore la créativité, mais peut également, notamment dans le cas d'émotions négatives comme la peur, être un frein à certaines de nos actions.

Avoir la possibilité de connaître l'émotion ressentie par une personne, c'est aussi avoir la potentialité de lui venir en aide en suivant l'évolution de ses souffrances liées à des phobies dans le cadre d'un système cherchant à les en soulager. Cela peut par exemple concerner des personnes âgées en perte de mobilité et qui s'exposent à la crainte d'une perte de stabilité et donc d'une chute dans leurs déplacements. On peut également citer le cas de personnes souffrant d'apathie, c'est-à-dire qui ne sont plus en mesure de ressentir des émotions. Ce système, c'est celui que vise à créer le projet européen VERVE¹ (Vanquishing fear and apathy through E-inclusion) dans lequel s'est inscrit une majeure partie de cette thèse.

1.2 Vers un système de reconnaissance automatique des émotions basé sur l'EEG

Ce n'est que depuis peu que l'informatique affective s'est emparée de l'électroencéphalographie (EEG) afin de l'intégrer dans les systèmes de reconnaissance automatique de l'émotion. Bien que son utilisation soit plus accessible que les moyens d'investigation de l'activité cérébrale utilisés en neurosciences (PET², MEG³, IRM⁴), elle demande à faire face à quelques obstacles si l'on souhaite en faire une utilisation convenable. Tout d'abord, d'un point de vue pratique, le système d'acquisition est lourd à mettre en place, certains casques pouvant compter jusqu'à 256 électrodes (BioSemi ActiveTwo⁵). Les signaux EEG sont également facilement sujets à différents artefacts. De simples actions, tel que cligner des yeux ou serrer la mâchoire, sont sources de bruits dans les signaux enregistrés.

Hormis ces problèmes pratiques, un autre problème d'un point de vue théorique est également présent. Alors que les systèmes de reconnaissance de tâches motrices sont effectués indépendamment du sujet enregistré, les systèmes de reconnaissance d'activités cognitives telles que l'émotion présentent une difficulté majeure : les résultats obtenus sont dépendants au sujet. On peut remarquer que la plupart des travaux dans la reconnaissance automatique de l'émotion est réalisée pour chaque sujet de manière dépendante [Horlings et al., 2008; Koelstra et al., 2012; Schaaff and Schultz, 2009b; Wang et al., 2014b]. Ceci est dû au fait que ressentir une émotion est une tâche subjective, les réactions provoquées par un même stimulus varient d'une personne à l'autre. Quelques travaux s'aventurent toutefois

1. Projet Européen VERVE n.288914 (FP7/2007-2013)

2. Tomographie par émission de positons

3. Magnétoencéphalographie

4. Imagerie par résonance magnétique

5. <http://www.biosemi.com/headcap.htm>

dans des systèmes de classification qui ne présentent pas de dépendance au sujet, mais les résultats obtenus ne sont pas significativement plus convaincants dans cette condition comme le montrent les travaux de Wang et al. [2014b] et Jatupaiboon et al. [2013b].

Le problème de l'utilisation d'une telle modalité est amplifié par le fait que la mise en place d'un enregistrement de signaux EEG est lourd en temps et en énergie, ce qui conduit souvent à des corpus disposant soit d'un faible nombre de participants ou bien d'une quantité de sessions ou d'essais assez limités. On peut également citer comme autre grande contrainte le peu de données disponibles pour chaque participant. Il faut donc pouvoir s'y adapter, notamment grâce à un modèle de validation croisée que l'on peut intégrer dans l'architecture de classification.

Il est néanmoins important de noter que l'EEG présente l'avantage de pouvoir contenir des informations sur des activités internes liées au phénomène émotionnel, tout en étant acquis par un système non-invasif. Les autres signaux physiologiques, qui s'attachent à être un recueil de manifestations extérieures, ne sont quant à eux que le reflet de modifications du comportement des organes internes.

Tous ces éléments amènent à adopter différentes stratégies afin de mettre en place un système de reconnaissance optimal.

La reconnaissance des émotions est habituellement approchée comme un problème de classification où le choix de caractéristiques appropriées est critique pour s'assurer d'une précision de reconnaissance satisfaisante. À l'heure actuelle, concernant les caractéristiques extraites à partir des signaux d'EEG, un consensus n'a pas été établi quant à un ensemble de caractéristiques appropriées qui pourrait garantir une caractérisation réussie des émotions chez un sujet humain. Il est toutefois reconnu dans le domaine des neurosciences qu'une grande quantité d'information pertinente est véhiculée par les propriétés spectrales des signaux EEG, où les différents types d'activités humaines résultent en différents motifs spectraux dans des bandes de fréquences spécifiques.

1.3 Objectifs et Contributions

L'utilisation de dispositifs d'enregistrements de signaux d'EEG se démocratise et ils possèdent un nombre de plus en plus réduit d'électrodes en devenant de même plus mobiles avec l'apparition de casques d'acquisition de signaux EEG sans fil. La problématique dans la reconnaissance des émotions à partir de signaux EEG peut alors être envisagée sous l'angle d'une application grand public, à savoir dans le cadre d'un dispositif allégé pouvant aller éventuellement jusqu'à une configuration mono-canal.

Nos travaux s'orientent vers la mise en place d'un système de reconnaissance automatique des émotions à partir de signaux EEG, en proposant de nouveaux descripteurs qui n'ont pas été utilisés dans ce domaine auparavant. On s'oriente pour cela vers un système de classification binaire supervisée qui dans un premier temps visera à discriminer deux classes d'émotions : de *valence* positive versus négative.

Une autre contribution importante des travaux menés au cours de cette thèse a été la création d'un nouveau corpus de données multimodales pour la reconnaissance des émo-

tions. Ce corpus répond à certaines faiblesses rencontrées dans les quelques corpus existants et accessibles, et il s'attache à offrir la possibilité d'étudier un nouvel aspect de l'émotion à savoir l'analyse de sa dynamique. Cela est permis par l'utilisation de stimuli vidéo et l'annotation des variations dans l'émotion ressentie.

1.4 Organisation du document

Ce document présente dans une première partie un état de l'art et des préalables nécessaires à une bonne lecture de ce document (**Chapitre 2**). Il est par la suite organisé en deux parties distinctes correspondant aux deux principales contributions des travaux de cette thèse :

- la création d'un corpus multimodal pour la reconnaissance des émotions (**Chapitre 3**),
- et l'étude et le développement d'un système de reconnaissance des émotions basé sur les signaux EEG (**Chapitre 4**).

Une étude de l'existant

Concernant l'étude des travaux existants, les premières questions que l'on vient à se poser sont comment définir une émotion et par quels moyens se manifeste-t-elle ? Le phénomène émotionnel est complexe et touche à de nombreux domaines. L'informatique affective s'est appliquée à étudier une grande variété de manifestations de l'émotion chez l'homme par le biais de modalités variées. Ce n'est que plus récemment que des systèmes de reconnaissance des émotions basés sur l'EEG ont émergés et avec eux l'apparition de corpus dédiés à cette application.

En amont de la création d'un corpus de données liées à l'émotion il est nécessaire de se pencher sur les facteurs qui peuvent influencer l'expérience. La littérature est riche en ce domaine et les préalables présentés dans un premier temps font état de nombreuses réflexions qui prennent appui sur les théories psychologiques de l'émotion. Deux étapes sensibles interviennent dans la mise en place d'un protocole expérimental lié à l'émotion, à savoir la méthode employée pour éliciter l'émotion et la retranscription de l'émotion ressentie par le sujet. Ces travaux nous permettent, par le biais de discussions, d'apporter des solutions aux problématiques auxquelles on doit faire face dans le cadre de la création d'un corpus.

La création d'un corpus

La création d'un corpus émotionnel fait l'objet de nombreux défis à relever. L'étude des travaux existants sur les méthodes employées pour le déclenchement et la retranscription de l'expérience émotionnelle permet d'apporter des réponses au problème suivant : dans quelle mesure peut-on s'affranchir au mieux de la part de subjectivité dans le cadre du dispositif expérimental ?

Les stratégies employées sont explicitées et discutées dans la section suivante et sont issues d'un travail de réflexion mené sur l'existant. Elles cherchent à répondre aux défis d'élicitation, d'annotation et d'enregistrement de données multimodales.

Enfin une analyse statistique des annotations fournies par les participants lors de l'expérience nous permet de détecter les divergences entre les différents annotateurs et d'apporter des informations sur la fiabilité de la stratégie d'annotation et sur la pertinence des stimuli sélectionnés.

Le système de reconnaissance

Les manifestations émotionnelles sont aussi bien physiques que psychiques. C'est cette dernière manifestation qui est reflétée par les informations contenues dans les signaux EEG. Le dernier chapitre est consacré à une étude comparative des descripteurs existants et de descripteurs proposés dans le cadre d'un système de reconnaissance des émotions basé sur l'EEG. Cette étude cherche à mettre en évidence la caractérisation de ces manifestations émotionnelles par le biais de descripteurs pertinents pouvant contenir des informations temporelles, spatiales et/ou fréquentielles. La spécificité de ce système tient notamment à la recherche d'une configuration allégée des données EEG pouvant aller jusqu'au mode mono-canal.

Chapitre 2

La reconnaissance des émotions : Etat de l'art et Préalables

Sommaire

2.1 A la recherche de l'émotion	13
2.1.1 Une définition possible? Les théories de l'émotion	13
2.1.2 Le cerveau et les émotions	18
2.1.3 Bilan et discussion	21
2.2 Le point de vue en informatique affective et les différentes modalités pour la reconnaissance des émotions	21
2.2.1 Les signaux physiologiques périphériques	22
2.2.2 Les signaux audio-visuels	25
2.2.3 Les signaux EEG	27
2.2.4 Les systèmes multi-modaux	27
2.3 Une vue d'ensemble des systèmes précédents pour la reconnaissance des émotions basée sur l'EEG	32
2.3.1 Les pré-traitements des signaux EEG	33
Normalisation	33
Sous-échantillonnage	33
2.3.2 Le système de classification	33
Les caractéristiques	34
Intégration Précoce	36
Intégration Tardive	36
2.4 L'expérience des émotions	36
2.4.1 Retranscrire l'émotion	36
2.4.1.1 Les représentations de l'émotion	37
Les catégories d'émotions : une distinction sémantique	37
Le cône de Plutchik	38
Les dimensions émotionnelles	40
Discussion	40

2.4.1.2 Les stratégies d'annotation	42
L'annotation statique	42
L'annotation dynamique	45
Discussion	46
2.4.2 Eliciter une émotion	47
Le principe de l'élicitation des émotions	47
Les modalités des stimuli provoquant une émotion	49
Discussion	50
2.5 Les corpus existants	50
2.5.1 ENTERFACE'06 [Savran et al., 2006]	50
2.5.2 DEAP Dataset [Koelstra et al., 2012]	53
2.5.3 MAHNOB-HCI [Soleymani et al., 2012a]	53
2.5.4 Discussion	54
2.6 Conclusion	55

Résumé

La reconnaissance des émotions a fait l'objet de nombreux travaux, et ce dans des domaines variés, tels que la psychologie, la philosophie ou les neurosciences. Une première approche de l'étude de ces travaux consiste à analyser les tentatives de définition du phénomène émotionnel. Ces tentatives de définition se situent principalement dans la sphère de la psychologie et posent les bases de l'informatique affective. Les éléments impliqués dans ce phénomène sont au moins la présence d'un facteur déclenchant (le stimulus), les réactions physiologiques et les réactions psychiques.

Les travaux menés en informatique affective sur la reconnaissance des émotions ont abordé cette dernière à travers la prospection de différentes modalités, telle que l'électro-encéphalographie (EEG). Ces modalités ont été étudiées de manière isolée, mais également multimodale afin de conserver un maximum d'informations complémentaires liées au phénomène émotionnel.

Cette tâche de reconnaissance repose sur la mise en place d'un système de classification dont les paramètres sont établis de manière appropriés. Ce système est composé de plusieurs étapes dont un pré-traitement des données qui veille à conserver l'information utile, le choix d'une stratégie d'intégration des caractéristiques, précoce ou tardive, et d'un classifieur adéquats. Les signaux EEG acquis dans le cadre de tâches cognitives ont pour particularité d'avoir un contenu lié au sujet. Les classifieurs les plus adaptés à la variabilité des données EEG sont les Machines à Vecteurs Support (SVM).

Un tour d'horizon des trois principaux corpus accessibles liés à la reconnaissance des émotions basée sur l'EEG permet de mettre en évidence les prémisses des difficultés auxquelles nous aurons à faire face pour la création de notre corpus.

2.1 A la recherche de l'émotion

2.1.1 Une définition possible ? Les théories de l'émotion

Les émotions prennent une part importante dans notre vie quotidienne. Le premier souvenir est souvent lié à une émotion forte qui a pu être ressentie à cet instant. Elles jouent un rôle essentiel aussi bien sur le plan social que cognitif, et elles ont une influence sur nos capacités de mémoire ou d'apprentissage.

Cependant, dès lors que l'on souhaite donner une définition précise de l'émotion, on se retrouve à coup sûr perdu dans des notions vagues mais surtout diverses.

Les émotions ne sont pas une humeur, ni même un tempérament. Elles correspondent à des états de courtes durées, allant de quelques secondes à quelques minutes, contrairement à ces derniers qui durent plusieurs jours pour l'humeur et sur une période encore plus longue pour le tempérament, pouvant être défini sur le temps d'une vie [Gunes et al., 2008].

Définir une émotion, c'est être en mesure d'analyser la perception du phénomène émotionnel. Plusieurs tentatives de définition du processus émotionnel ont été réalisées. On retrouve parmi ces définitions plusieurs courants de pensées qui évoluent au cours du contexte scientifique dans lequel il se situe. Cette partie ne fournira que les grandes lignes des théories les plus pertinentes, c'est-à-dire celles qui auront un impact dans le cadre de la reconnaissance des émotions.

Le sens commun voudrait que le processus émotionnel se déroule de la manière suivante : un stimuli provoque chez le sujet une émotion qui déclenche à son tour un ensemble de réactions physiologiques qui lui serait associées (Figure 2.1). Comme nous allons le voir, ce mode de fonctionnement, qui dans un premier temps paraît pourtant aller de soi pour la plupart d'entre nous, est cependant mis à mal par plusieurs théories.

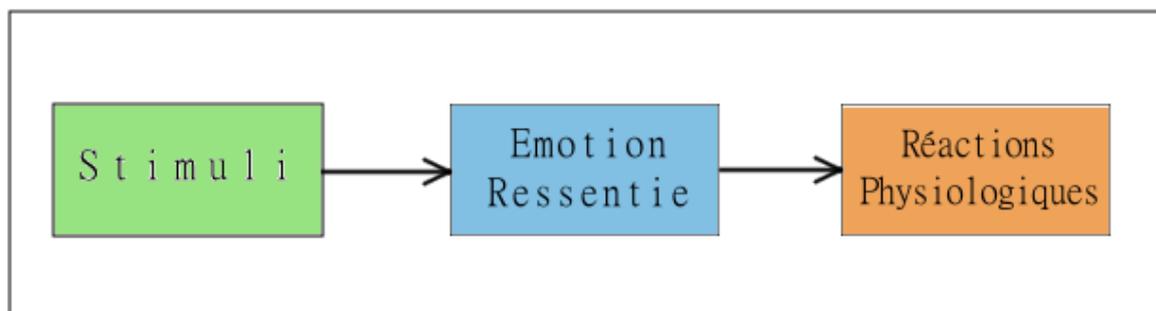


FIGURE 2.1 – Le processus émotionnel selon le *Sens Commun*.

Charles Darwin [Darwin, 1872] ne s'est pas spécifiquement intéressé aux émotions chez l'homme, mais à la manière dont elles s'expriment chez les êtres vivants en général. Il s'attèle à retranscrire les principes et les moyens mis en œuvre dans l'expression des émotions, avec une attention toute particulière aux mécanismes mis en œuvre chez les animaux. Il prend l'axe d'un caractère instinctif à la présence des émotions, c'est la raison pour laquelle

il accorde une telle importance aux moyens d'expression des émotions chez les animaux. Il est néanmoins important de noter qu'à aucun moment il ne cherchera à donner une définition du terme *émotion*. Les travaux qui s'appuient sur cette théorie considèrent que les émotions seraient héritées de nos ancêtres et défendent l'importance de la fonction communicative des émotions Ekman [1999]; Rimé and Scherer [1989]. On peut proposer les travaux de Hess and Thibault [2009] comme étude détaillée de la théorie évolutionniste proposée par Darwin.

Les deux grands théoriciens William James et Walter Cannon, issus du domaine de la physiologie, ont opposé leurs versions de la description du phénomène émotionnel à l'aube du XXème siècle. William James soutient la thèse selon laquelle la *perception* d'un stimulus engendre des réactions physiques spécifiques à chaque émotion [James, 1884]. Cette spécificité permet de prendre conscience et d'identifier l'émotion ressentie (Figure 2.2a). Il s'agit d'une expérience subjective de l'émotion. Un autre postulat important de sa théorie repose sur la possibilité d'une différenciation des manifestations du corps pour huit émotions qu'il définit comme *standards* (la surprise, la curiosité, le ravissement, la peur, la colère, la convoitise, la cupidité, l'affection).

William James [James, 1884] -

The bodily changes follow directly the PERCEPTION of the exciting fact, and that our feeling of the same changes as they occur IS emotion¹.

Cannon développe quant à lui une nouvelle théorie des émotions qui remet en cause la théorie de James-Lange. Il la met à mal en s'appuyant sur de nombreux arguments réfutant les facteurs viscéraux et les processus cérébraux sur lesquels cette théorie repose [Cannon, 1927]. Cependant sa plus grande critique porte sur l'absence de différenciation dans les réactions viscérales pour plusieurs émotions et dans des états non-émotionnels. En accord avec Bard [1928], il soutient la thèse selon laquelle l'expression des émotions résulte du traitement du stimulus effectué par le thalamus (Figure 2.2b).

Walter B. Cannon [Cannon, 1927] -

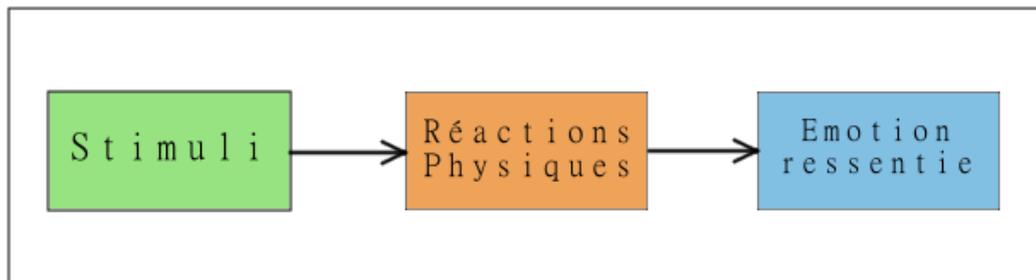
The peculiar quality of the emotion is added to simple sensation when the thalamic processus are roused².

Les critiques concernant ce modèle sont peu nombreuses car cette théorie n'a été au final que très peu exploitée. Elle est plutôt utilisée comme support pour une critique de la théorie viscérale des émotions de James-Lange.

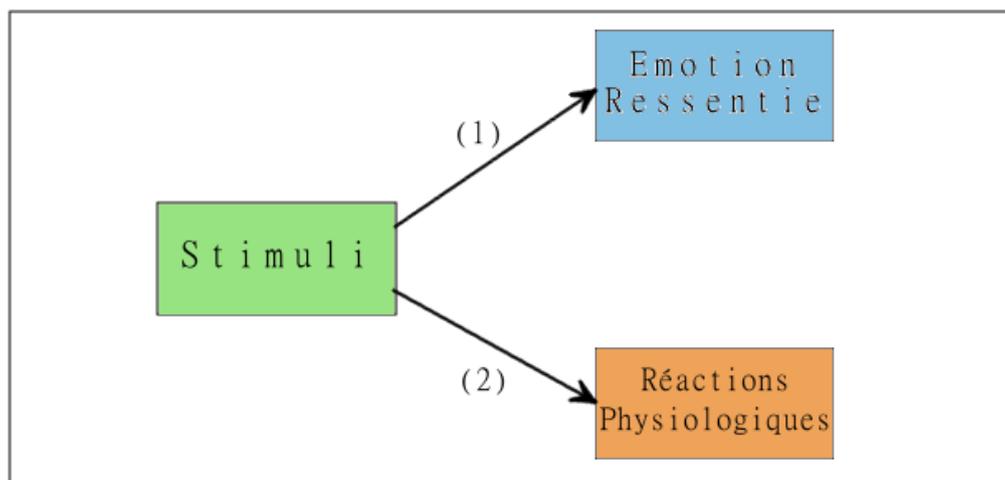
Les deux psychologues américains, Stanley Schachter et Jerome Singer, sans vouloir contredire la théorie de James-Lange, qui ne peut être totalement réfutée aux vues des travaux de l'époque menés sur la différenciation de la peur et de la colère [Ax, 1953; Schachter,

1. Les modifications corporelles suivent directement la PERCEPTION d'un fait excitant, et c'est notre impression de modifications identiques alors qu'elles se produisent qui EST l'émotion.

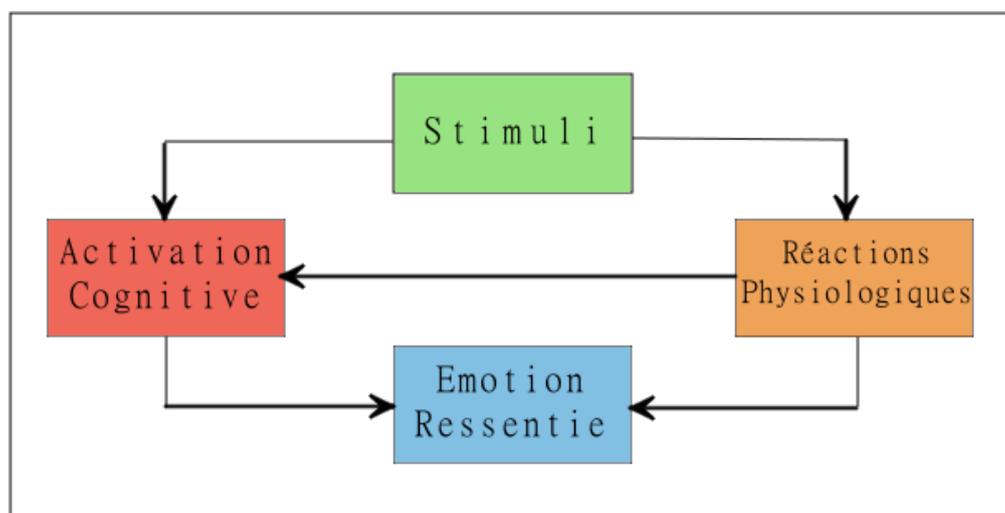
2. La qualité spécifique de l'émotion est ajoutée à une simple sensation lorsque les processus thalamiques sont réveillés.



(a) La théorie de James-Lange (1884).



(b) La théorie de Cannon-Bard (1927). Le thalamus réalise deux opérations distinctes : il envoie une impulsion au cortex (1) et une autre à l'hypothalamus (2).



(c) La théorie de Schachter-Singer (1962).

FIGURE 2.2 – Les principales théories des émotions

1957], en proposent toutefois une modification [Schachter and Singer, 1962]. On doit effectivement faire face à l'époque à un immobilisme dans la découverte de patterns associés à un état émotionnel particulier. En accord avec d'autres travaux [Hunt et al., 1958] ils sont amenés à proposer une théorie selon laquelle les facteurs cognitifs seraient déterminants dans le processus de l'émotion. Ils soutiennent la théorie selon laquelle un état émotionnel est fonction à la fois d'une activation physiologique et cognitive (Figure 2.2c). Leur théorie repose sur le rôle du système nerveux sympathique, le siège des réactions physiologiques, qu'ils savent nécessaire mais dont ils questionnent son caractère suffisant pour se sentir dans un état émotionnel. Ils ne remettent pas en cause l'implication de ces réactions physiologiques mais ils doutent cependant de l'existence de différenciations bien marquées entre chaque état émotionnel. Peuvent-elles à elles seules expliquer l'état émotionnel ? Ils soumettent alors les trois postulats suivants :

1. l'état émotionnel est fonction de réactions physiologiques, identiques à plusieurs étiquettes émotionnelles, et de l'activation cognitive, qui quant à elle va pouvoir jouer le rôle de facteur discriminant,
2. si l'individu a la connaissance de l'élément déclencheur des réactions physiologiques, par exemple par l'administration d'adrélanine ou d'épinéphrine, il ne fait pas l'expérience d'un ressenti émotionnel,
3. dans un cadre cognitif impliquant par exemple un grand danger (comme par exemple être en présence d'un ours, d'un homme armé,...), l'individu n'aura un ressenti émotionnel que dans le cas où il ressent une activation physiologique.

S. Schachter et J. Singer [Schachter and Singer, 1962] -
 [...] emotional states are a function of the interaction of [...] cognitive factors with a state of physiological arousal³.

Les principales critiques de cette théorie ont été proposées par Cotton [1981]; Reizenzein [1983], dont la critique dominante repose sur la dénonciation d'une méprise et d'une surévaluation du rôle de l'activation cognitive dans l'émotion .

Nous avons ici exposé les quatre grandes théories du phénomène émotionnel. Elles sont essentielles en ce sens qu'elles constituent l'origine de la plupart des recherches menées dans le cadre de la compréhension du lien qui unit les réactions et les sensations qui interviennent dans le cadre du processus émotionnel (Section 2.2). Nous pouvons toutefois ajouter à ces dernières, d'autres courants de pensées, issus de différentes disciplines, qui apportent un éclairage particulier.

Sartre propose dans son *Esquisse d'une théorie des émotions* [Sartre, 1965] une approche phénoménologique de l'émotion. En réalisant dans un premier temps une critique des théories traditionnelles, aussi bien dans le domaine de la psychologie que de la psychanalytique émergente, il s'oriente vers une vision phénoménologique en définissant l'émotion

3. [...] les états émotionnels sont fonction de l'interaction de [...] facteurs cognitifs avec un état d'activation physiologique.

comme le "mode d'existence de la conscience". Cependant son raisonnement le guide vers une aporie dont il sera lui-même conscient, ce qui le mènera à écarter son essai de son œuvre majeure *L'être et le néant*.

Jean-Paul Sartre [Sartre, 1965] -
Une émotion est une transformation du monde.

Mais tout n'est pas à écarter dans cette étude. L'intérêt de son ouvrage réside dans le fait qu'il lance des pistes, même si elles se trouvent être sans démonstrations, ce qui est une des faiblesses de son essai, et il permet de mettre en exergue la complexité de la définition d'une émotion.

Damasio [Damasio, 2008, 1996] ne propose pas à proprement parler une théorie des émotions, mais il s'interroge plutôt sur le rôle qu'ont les émotions dans nos prises de décision et propose ce qu'il appelle l'*hypothèse des marqueurs somatiques* (Figure 2.3). Ces marqueurs seraient des mécanismes émotionnels qui influencent, de manière consciente ou inconsciente, la prise de décision en fonction du souvenir d'expériences passées. Ils sont qualifiés de *somatiques* en ce sens qu'ils sont liés à la structure et à la régulation de l'état corporel, en intervenant aussi bien sur le corps à proprement parler que sur sa représentation mentale.

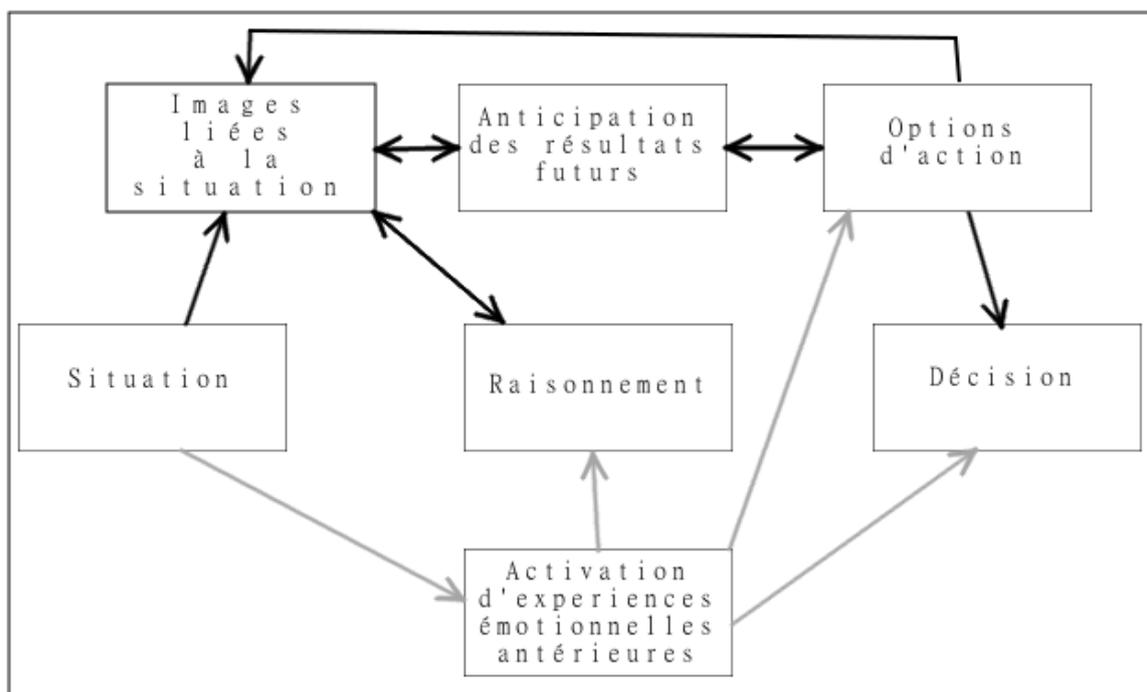


FIGURE 2.3 – Le processus émotionnel dans la prise de décision : l'hypothèse des marqueurs somatiques formulée par Damasio [2008].

4. Les états émotionnels sont fonction de l'interaction de facteurs cognitifs avec un état d'activation physiologique.

Antonio Damasio [Damasio, 1996] -

[...] 'marker' signals influence the processes of response to stimuli, at multiple levels of operation, some of which occur overtly (consciously, 'in-mind') and some of which occur covertly (non-consciously, in a non-minded manner). The marker signals arise in bioregulatory processes, including those which express themselves in emotions and feelings, but are not necessarily confined to those alone⁴.

Cette théorie faite par Damasio, tend à se rapprocher d'un autre courant majeur qui est celui de la théorie de l'évaluation cognitive, ou *théorie de l'appraisal*⁵ [Scherer et al., 2001]. Cette théorie, qui trouve ces origines dans les travaux de Arnold [1960], est portée par Scherer [Scherer et al., 2001] et Frijda [Frijda, 1988]. Elle trouve un écho retentissant dans le domaine de l'informatique affective, mais surtout dans la communauté des neurosciences affectives. Klaus Scherer va à contre courant des pratiques habituelles, en refusant de diviser le phénomène émotionnel en deux classes : conscient et inconscient [Scherer, 2005a]. Selon lui, la plupart des processus émotionnels implique une activité inconsciente et seuls quelques uns aboutissent parfois à une fonction consciente. Comme l'indique son sous-titre, *The Bulk of the Iceberg*⁶, faisant référence à une métaphore de Freud (Figure 2.4), sa proposition s'oriente vers la théorie psychanalytique à savoir *la science de l'inconscient*. Une augmentation de la densité du traitement mental peut permettre d'aboutir à une plus grande émergence du traitement dans la conscience. Avec cette théorie il ouvre le débat sur l'implication de ces deux processus, conscient et inconscient, dans l'élicitation et la différenciation des émotions.

Klaus R. Scherer [Scherer, 2004] -

Emotion is defined as an episode of massive, synchronous recruitment of mental and somatic resources to adapt to, or cope with, a stimulus event that is subjectively appraised as being highly pertinent to the needs, goals, and values of the individual⁷.

Les théories sont nombreuses et font l'objet d'un grand nombre de critiques. Nous renvoyons le lecteur vers les principaux papiers de Frijda [1986]; Kleinginna and Kleinginna [1981] et Scherer [2000] et les livres de Scherer et al. [2010] pour une lecture plus approfondie à ce sujet et vers Coppin and Sander [2012] en ce qui concerne les théories plus contemporaines.

2.1.2 Le cerveau et les émotions

Nous faisons ici une description très brève des zones cérébrales prenant part dans le mécanisme émotionnel, afin de se concentrer davantage sur leur implication dans la com-

5. Evaluation.

6. "La partie immergée de l'Iceberg"

7. L'émotion est définie comme un épisode d'engagement imposant et synchrone de ressources mentales et somatiques pour s'adapter, ou faire face, à un événement qui est évalué subjectivement comme étant extrêmement pertinent aux besoins, aux objectifs et aux valeurs de l'individu.

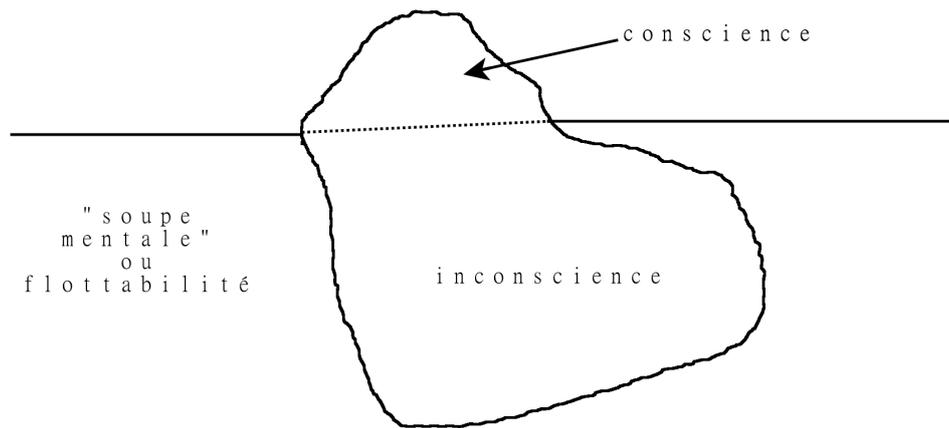


FIGURE 2.4 – La métaphore de l'Iceberg de Freud appliquée au processus émotionnel. La force de flottabilité est liée à la masse de liquide déplacée par l'objet.

préhension des émotions et d'étudier les liens possibles existant avec les principales théories de l'émotion.

Les principales parties du cerveau qui sont évoquées dans le processus du phénomène émotionnel (Figure 2.5) sont :

- le *cortex* (a) : substance périphérique des hémisphères cérébraux, il fournit une information détaillée des stimuli. L'information qu'il fournit met plus de temps à arriver dû aux différents traitements qui ont lieu au niveau cortical [Davis et al., 1966; White, 1989],
- le *thalamus* (b) : il réalise des échanges avec le cortex [Schlögl, 2000], probablement concernant des informations sensorielles spécifiques, et il intervient dans la déconnexion du cortex de stimulations sensorielles en phase de sommeil.
- l'*hippocampe* (c) : il tient un rôle au niveau de la mémoire [Squire, 1992]. LeDoux [2000] suppose qu'il permettrait aux réponses de type peur d'être activées par la mémoire (consciente ou explicite) des expériences passées,
- l'*hypothalamus* (d) : il réagit à plusieurs facteurs tels que les informations nerveuses provenant du coeur, la faim, les stimuli olfactifs. Il forme avec le thalamus (b), et l'épithalamus le diencephale.
- l'*amygdale* (e) : situé à l'extrémité de l'hippocampe, il serait essentiel pour décoder les stimuli représentant une menace [LeDoux, 2007], et dans le traitement des informations visuelles [Pessoa and Adolphs, 2010].

En se basant sur les travaux de Darwin [Darwin, 1872], des chercheurs ont également porté leur attention sur l'évolution du cerveau humain par rapport aux autres espèces animales [Berridge and Kringelbach, 2013; LeDoux, 2000]. Le cortex préfrontal s'étant plus développé chez l'homme, cela a eu pour effet l'apparition de nouvelles sous-régions préfrontales qui n'ont pas leur équivalent chez les espèces non-primates, tel que le cortex insulaire (ou insula) où intervient le dégoût [Murphy et al., 2003].

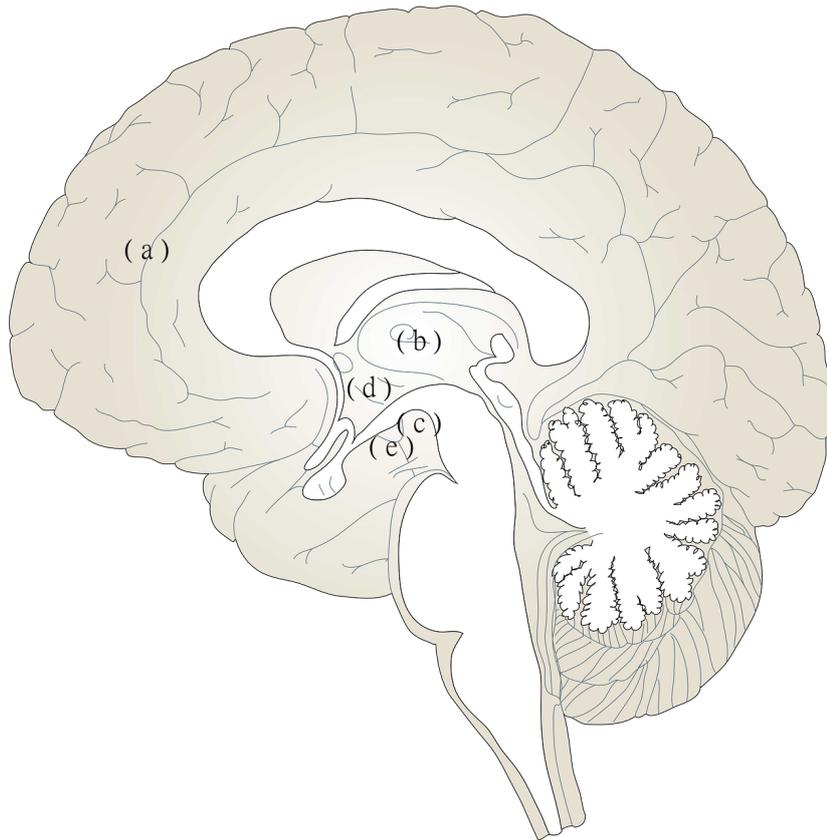


FIGURE 2.5 – Schéma du cerveau.

D'un point de vue théorique, les neurosciences [Brosch and Sander, 2013; LeDoux, 2000] sont plus adeptes de la théorie de l'*appraisal* [Scherer et al., 2001]. Le domaine des neurosciences met en jeu le système limbique (amygdale, hypothalamus, hippocampe) dans le processus de l'émotion [MacLean, 1949], cependant LeDoux et al. [LeDoux, 2000] et Murphy et al. [Murphy et al., 2003] s'accordent sur le fait qu'il n'est pas le seul à entrer en jeu dans la structure neuronale de l'émotion. Plusieurs courants se sont développés pour la caractérisation des émotions. Dans un premier temps un système simple où prédomine l'activité de l'hémisphère droit, et un système dual entre les hémisphères droit et gauche [Davidson, 1984]. Ces hypothèses ont été évaluées par Murphy et al. [Murphy et al., 2003] qui ont regroupé un ensemble de 106 études sur la détection de l'émotion en neuro-imagerie (PET, IRMf⁸). Ils ont de même évalué un modèle multi-système basé sur les émotions dites basiques du programme d'Ekman [Ekman and Friesen, 1976] (peur, dégoût, colère, joie, tristesse, surprise) supposées être invariantes suivant les populations [Ekman, 1992]. Il en ressort que la différence d'activation entre l'hémisphère droit et l'hémisphère gauche n'est pas significative, l'hémisphère gauche étant par ailleurs légèrement plus activé. Concernant les émotions du programme d'Ekman, il a pu être constaté une différence entre d'un côté la peur, le dégoût et la colère, où des régions spécifiques s'activent (respectivement l'amygdale, l'insula/le globus oallidus, et le cortex orbito-frontal latéral), et d'un autre côté la joie et la tristesse.

8. Imagerie par résonance magnétique fonctionnelle

L'utilisation de l'électro-encéphalographie est plus restreinte dans le domaine des neurosciences dans le cadre de la reconnaissance des émotions, mais elle n'en est pas pour autant totalement écartée [Daly et al., 2014]. Cet outil d'enregistrement de l'activité cérébrale, non-invasif est plus accessible à des domaines non spécialistes, et est le plus couramment utilisé dans l'informatique affective. L'EEG, en plus de permettre d'obtenir des connaissances sur l'activité qui se produit dans le cortex cérébral [Constant and Sabourdin, 2012], qui est le siège de nombreux traitements réalisés par le cerveau : visuels [Kastner and Ungerleider, 2000; Wandell et al., 2007], auditifs [Davis et al., 1966], motrices [Fink et al., 1997], ou encore spatiales [Fried et al., 1982], présente l'avantage d'être un dispositif plus léger que les systèmes d'investigation de l'activité cérébrale utilisées par les neuroscientifiques.

Les études menées sur le cerveau ne sont plus réellement à leur état de balbutiement, mais elles connaissent, à l'heure de la rédaction de ce manuscrit, un réel essor qui est appuyé par des efforts étatiques de par le monde pour leur développement. Parmi les plus grands projets nous pouvons citer le Human Brain Project⁹ (HBP) pour l'Union Européenne, le Brain Initiative¹⁰ pour les Etats-Unis, et dernièrement le China Brain Science Project.

2.1.3 Bilan et discussion

Les travaux menés dans l'étude du phénomène émotionnel au sein du domaine de la psychologie visent à apporter des éléments de réponse quant à sa composition et à sa structuration. S'ils s'accordent sur une structure composée d'au moins trois éléments, à savoir la présence d'un élément déclencheur (le stimulus), la présence consciente ou non de manifestations corporelles (réactions physiologiques) et du ressenti d'une émotion, il n'en reste pas moins que chacun en propose une organisation différente. Cela met en lumière la difficulté de trouver un consensus sur une définition stricte du phénomène émotionnel et de situer clairement le rôle joué par l'activité cérébrale. La tendance tend cependant à confirmer que cette activité ne peut être représentée à elle seule par un même élément dans la structure mais qu'elle intervient sous la forme de sous-éléments répartis en amont et en aval du phénomène émotionnel. C'est donc dans cette incertitude des pistes à explorer que s'inscrivent les travaux effectués au cours de cette thèse.

2.2 Le point de vue en informatique affective et les différentes modalités pour la reconnaissance des émotions

Les théories énoncées précédemment trouvent un certain écho dans le domaine de l'informatique comme nous allons le voir tout au long de cette section. L'informatique affective

9. <https://www.humanbrainproject.eu/fr>

10. <http://www.braininitiative.nih.gov/>

ne va cependant pas chercher à mettre en évidence les mécanismes sous-jacents du phénomène émotionnel, plutôt laissé aux soins du domaine des neurosciences, mais elle va s'atteler à explorer les manifestations de ce phénomène. Il s'agit ici de s'orienter vers des travaux de reconnaissance des émotions à défaut d'une compréhension. Les explorations menées dans ce domaine aspirent à déboucher sur une description détaillée du phénomène émotionnel, qu'il soit visible par le fait de manifestations extérieures (expressions du visage, modification de la voix), ou moins perceptible pour l'homme (activités physiologiques et notamment cérébrales).

La reconnaissance des émotions a été abordée dans plusieurs disciplines. Tout d'abord dans le domaine des neurosciences mais également dans le domaine de l'informatique affective notamment à l'aide de différents capteurs permettant la réalisation d'enregistrements de signaux physiologiques périphériques (conductance de la peau, ECG, respiration, température, etc.) et de signaux physiologiques issus de l'activité cérébrale avec l'électro-encéphalographie (EEG).

Les principales modalités utilisées en informatique affective sont :

- les signaux physiologiques périphériques (non-EEG),
- la parole ou les expressions du visage,
- l'électro-encéphalographie (EEG).

Dans un souci de clarté nous désignerons les signaux physiologiques périphériques sous le terme raccourci de *signaux périphériques* comme il est parfois d'usage [Chanel, 2009; Khalili and Moradi, 2008]. Si toutefois les signaux EEG sont également des signaux physiologiques, ils sont cependant à distinguer des périphériques, comme nous allons l'expliquer dans la section qui suit.

Nous orientons le lecteur qui souhaiterait un contenu plus extensif sur les systèmes d'informatique affective vers le livre de Pelachaud [2013] qui fournit des bases solides et nécessaires pour des travaux dans ce domaine.

2.2.1 Les signaux physiologiques périphériques

Plusieurs signaux physiologiques périphériques ont été utilisés comme témoins d'une manifestation émotionnelle. Lorsque nous ressentons une émotion, plusieurs manifestations corporelles surviennent. L'enregistrement de ces signaux lors de manifestations émotionnelles permet d'analyser les différentes manifestations possibles. Le champ de ces manifestations recouvre plusieurs niveaux : celles extérieures et visibles avec par exemple les expressions du visage, des modifications dans la voix, mais également des manifestations plus subtiles et non détectables à l'œil humain comme les micro-expressions du visage, les variations de la température corporelle, de même que des manifestations non détectables par une personne extérieure avec des changements dans la respiration, du rythme cardiaque.

Ensemble	Mesures	Activité physiologique
Système cardio-vasculaire	électrocardiographie (ECG), pression sanguine (BP)	rythme cardiaque, circulation sanguine
Activité électrodermale	conductance de la peau (SC)	transpiration
Système Respiratoire	respiration	fréquence et volume
Système Musculaire	électromyographie (EMG)	musculaire (dont visage)

TABLEAU 2.1 – Les principaux signaux périphériques utilisés dans le cadre de la reconnaissance de l'émotion en informatique affective.

L'intérêt de l'enregistrement de signaux périphériques réside dans le fait qu'ils résultent d'une activité du système nerveux périphérique (SNP). Le SNP correspond à l'ensemble des nerfs et ganglions situé à l'extérieur du cerveau et de la moelle épinière. C'est le pont entre les organes et le système nerveux central (SNC) dans lequel est inclus le cerveau. Au sein du SNP, on distingue le *système nerveux somatique* (SNS) et le *système nerveux autonome* (SNA). Ces signaux périphériques correspondent plus précisément à l'enregistrement des activités provenant du SNA, qui a la particularité de ne pas pouvoir être déclenché par un contrôle intentionnel ou conscient.

Le développement d'outils plus légers et moins coûteux pour l'enregistrement des signaux périphériques a permis l'essor de ces modalités en informatique affective. De nombreuses études se sont penchées sur l'implication du SNA dans le processus émotionnel [Ekman et al., 1983; Kreibig, 2010; Levenson, 2014]. Pour Stephens et al. [2010], cette implication ne peut être remise en cause, d'autant plus qu'il estime que les réactions obtenues en laboratoire sont moins intenses que dans la vraie vie, face à de vraies situations, et cela n'empêche pas pour autant d'obtenir des résultats optimistes. La solidité de cet argument a toutefois été débattu par Stemmler et al. [2001] qui observent que l'activité du SNA est statistiquement différente entre une peur réelle et une peur induite par le souvenir d'un événement personnel.

Après avoir établi le lien entre une activité du système autonome et les émotions [Ax, 1953], on a majoritairement cherché à attribuer des différenciations dans l'activité du SNA suivant les émotions ressenties, ce qui appuie la théorie de James-Lange selon laquelle un certain nombre d'émotions ont des *manifestations* corporelles distinctes [James, 1884] (son premier postulat : surprise, curiosité, ravissement, peur, colère, convoitise, cupidité, affection). On cherche à mettre au jour ce *langage naturel*, sans pour autant être en désaccord avec la théorie cognitivo-physiologique [Schachter and Singer, 1962]. Selon Schachter and Singer [1962], l'existence de distinctions physiologiques doit rester une question ouverte.

Kreibig [2010], en regroupant les données de 134 travaux, a essayé de faire ressortir les principaux patterns des activations physiologiques pour 22 catégories d'émotions.

Cet intérêt pour les signaux périphériques trouve ses origines dans la quête de différenciateurs physiologiques au sens de la théorie de James-Lange (voir la Section 2.1.1 et la Figure 2.2a). Cependant, bien que les études aient été nombreuses, aucun discriminant n'a pu être clairement établi à ce jour.

Des résumés des recherches menés sur l'activité du SNA sont nombreux et l'on peut

citer parmi eux les papiers de Kreibig [2010]; Levenson [1988, 2003, 2014]; Stephens et al. [2010].

Les principales informations physiologiques (Tableau 2.1) utilisées dans le domaine de l'informatique affective sont les suivantes :

- l'*électromyographie* (EMG) [Cheng and Liu, 2008] : les électrodes permettent d'enregistrer l'activité musculaire. Dans le cadre de travaux sur la reconnaissance des émotions, elles sont utilisées pour capturer l'activité des muscles du visage dont les principaux sont les muscles sourciliers et les zygomatiques majeurs [BrBradley, 2000; Rani et al., 2006].
- La *température* [Khan et al., 2006; Mcfarland, 1985; Rimm-Kaufman and Kagan, 1996] : l'étude de la température de la peau est quant à elle moins couramment utilisée car les variations sont faibles et mettent du temps à survenir. Il est difficile en post-traitement de synchroniser les données de manière fiable ces informations aux instants d'élicitation de l'émotion.
- La *respiration* [Bloch et al., 1991; Boiten et al., 1994; Feleky, 1916] : on distingue deux événements liés à la respiration qui peuvent être mesurés : le volume d'air inspiré ainsi que la fréquence de respiration.
- l'*électrocardiographie* (ECG) [Agrafioti et al., 2012; Anttonen and Surakka, 2005; Cai et al., 2009; Ma and Liu, 2009] : le rythme cardiaque est un indicateur précieux dans l'étude des signaux périphériques liés aux émotions. C'est un des signaux les plus fréquemment utilisés dans les travaux de reconnaissance des émotions basée sur les signaux périphériques [Kreibig, 2010].
- La *conductance de la peau* ou Galvanic Skin Response (GSR) [Khalifa et al., 2002] : on mesure ici la conductance électrique de la peau. Le signal subit des variations en lien avec la quantité de transpiration produite. Cette conductance est mesurée le plus souvent sur le poignet ou dans la paume de la main.
- La *pression sanguine (volumique)* ou Blood Volume Pulse (BVP/BP) [Gupta et al., 2012; Stemmler et al., 2001] : ces mesures fournissent des informations quant au flux sanguin à travers le corps. Une pression élevée résulte d'un débit sanguin plus important.
- L'*électro-oculographie* (EOG) ou le suivi du regard [Haak et al., 2009; Soleymani et al., 2012b] : l'EOG permet d'enregistrer des informations liés aux mouvements oculaires, à savoir en premier lieu les changements de direction effectués mais également les clignements des paupières.

Une illustration des signaux communément utilisés en informatique affective, à savoir l'ECG, l'EOG et l'EMG, est fournie dans la Figure 2.6.

Les signaux physiologiques périphériques sont pour la plupart enregistrés conjointement : pression sanguine et conductance (transpiration) [Levenson, 2003], EMG et conductance [Nakasone et al., 2005], EDA et ECG et EMG [Kallinen, 2004], cardiorespiratoire [Rainville et al., 2006] (voir le Tableau 2.2).

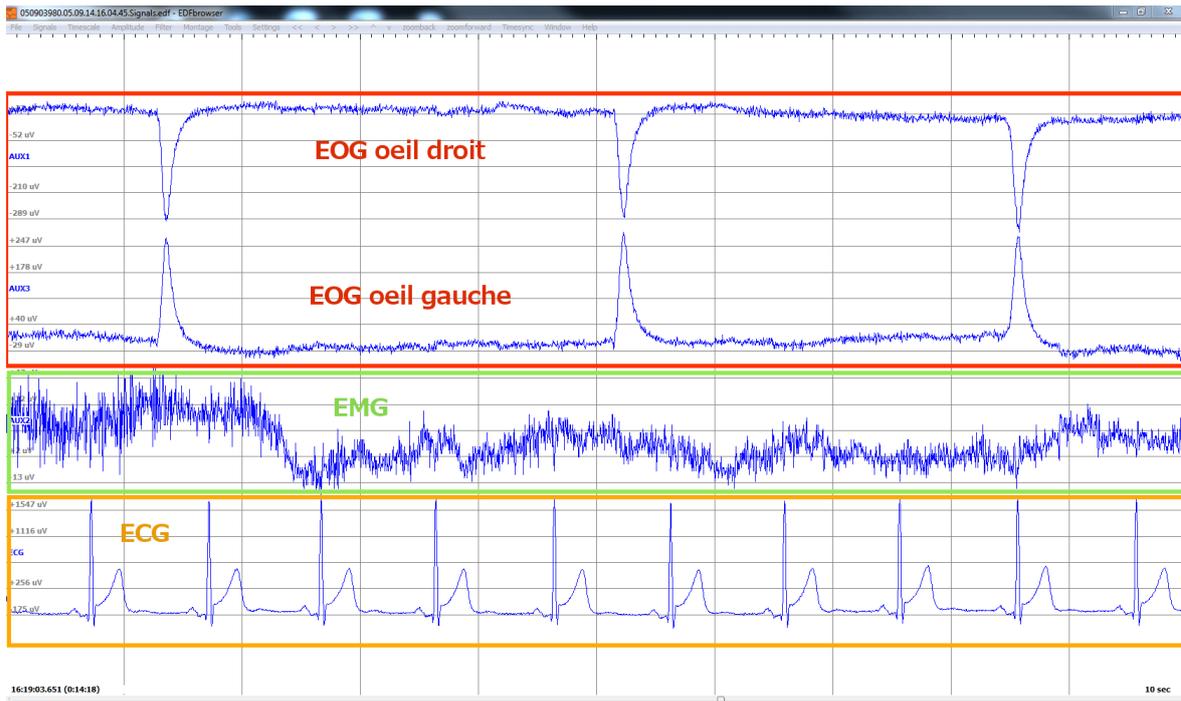


FIGURE 2.6 – Exemple d’enregistrements de signaux périphériques EOG, EMG et ECG sur une durée de 10 secondes.

L’étude des signaux périphériques a toujours une place prépondérante dans l’informatique affective dans le cadre de la reconnaissance des émotions [Jang et al., 2013].

2.2.2 Les signaux audio-visuels

Dans une autre mesure on emploie des méthodes d’acquisition audio-visuelles dont la qualité des données s’améliore au fur et à mesure que les technologies progressent. Les émotions sont souvent traduites par les expressions du visage et des modifications vocales. Ces modalités sont pertinentes en ce sens que l’expression des émotions joue un rôle essentiel dans les communications sociales [Ekman, 1999; Massey, 2002], comme le soutiennent les théoriciens auprès desquels les travaux de Darwin trouvent un écho particulier [Ekman, 2006] (voir Section 2.1.1).

Les modalités visuelles se concentrent principalement sur l’analyse des expressions du visage, dont l’universalité est reconnue pour certaines émotions [Adolphs, 2002; Darwin, 1872; Ekman, 1992], et plus récemment sur les micro-expressions, qui peuvent être détectées grâce à des technologies plus pointues qui ont permis d’aboutir à une meilleure résolution de la vidéo enregistrée par les caméras actuelles [Pfister et al., 2011; Polikovskiy et al., 2009; Shreve et al., 2011; Vanpé, 2011; Yan et al., 2013].

D’autres travaux portent sur la reconnaissance des émotions en se basant sur l’analyse des mouvements corporels [Castellano et al., 2007; Glowinski et al., 2008; Schindler et al., 2008], et nous pouvons conseiller la lecture du papier de Kleinsmith and Bianchi-Berthouze [2013] pour plus d’informations sur les travaux menés dans ce domaine.

Référence	Signaux Physiologiques	Nbr. de participants	Catégories d'émotion	Méthodes d'élicitation	Méthodes de classification	Meilleurs résultats
Kim and Andreé [2008]	EMG, ECG, conductance de la peau, respiration	3 hommes âgés de 25 à 38 ans	positive/high arousal, negative/high arousal, positive/low arousal, negative/low arousal	4 morceaux de musique choisis par le participant	Sequential Backward Selection (SBS) + Extended Linear Discriminant Analysis (GLDA)	91%
Chen et al. [2007]	GSR, pression sanguine, ECG, respiration, EEG	1 60 femmes âgées de 18 à 23 ans	excited -negative, excited positive, calm -neutral peur, joie, neutre	images 3 vidéos (5-8 minutes/vidéo)	Fisher Linear SVM CCA (canonical correlation analysis)	53% 47% 85%
Li and Chen [2006]	température, respiration, conductance de la peau, ECG	12 hommes	joie, colère, tristesse, peur, relaxation	vidéos	Linear SVM one vs. all	42%
Takahashi [2004]	BVP, GSR, EEG	1	arousal, valence	IAPS (2s)	neural network for regression, accuracy is computed as the number of samples that fall in a 20% bandwidth of the correct value	97%, 90%
Hang et al. [2004a]	EMG, SC, température, BVP, ECG, respiration	1	positive and negative state of mind	20 audio (60s)	AdaBoost, MI	74%
Kundhwa et al. [2011]	EEG, EDA, Pulse	1	neutre, colère, haine, peine, amour platonique, amour romantique, joie, reverence	images personnalisées	Hybrid Linear Discriminant Analysis	81%
Picard et al. [2001]	EMG, blood volume, pulse, conductance de la peau, respiration	1	joie, colère, tristesse, plaisir	musique	SVM	76%
Zong and Cheouani [2009]	ECG, EMG, SC, RSP (respiration changes)	10	amusement, contentement, dégoût, peur pas d'émotion (neutre), tristesse	IAPS	Fisher, SVM	92%, 90%
Maoui and Pruski [2010]	SKT (température), respiration	15	3 niveaux d'intensité pour l'engagement, l'anxiété, l'ennui, la frustration, la colère	résoudre des anagrammes, jouer à pong	KNN, Regression Tree, Bayes Network, SVM	79%, 83%, 78%, 86%
Rani et al. [2006]	ECG, EMG, GSR, température, BVP, heart sound	1	(colère, tristesse, joie, plaisir), (positive/negatif), (high/low arousal)	musique choisie par le participant	LDA, KNN, MLP, SFS, Fisher ANOVA	(92%), (85%), (96%)
Wagner et al. [2005]	EMG, ECG, GSR, respiration	50 enfants âgés de 7 à 8 ans	tristesse, colère, stress, surprise	multimodal (audio, visuel, cognitif)	SVM	62%

TABLEAU 2-2 – Liste et description de publications sur le reconnaissance de l'émotion utilisant les signaux physiologiques périphériques.

Les différentes intonations de la voix peuvent être révélatrices d'une émotion mais les travaux sur la reconnaissance des émotions dans la voix peuvent également être mis en œuvre dans le but de distinguer les émotions réelles des émotions actées [Nwe et al., 2003; Oudeyer, 2003; Schuller et al., 2011; Ververidis and Kotropoulos, 2006].

L'utilisation de ces modalités est souvent mise en pratique dans la reconnaissance de l'émotion dans un contenu multimédia [Clavel, 2006; Clavel et al., 2007; Dhall et al., 2013; Kim et al., 2010; Wang et al., 2014a]. La combinaison de ces deux modalités, audio et visuelles, sera discutée dans le paragraphe sur les systèmes multi-modaux.

Nous pouvons orienter le lecteur vers les papiers de El Ayadi et al. [2011]; Ramakrishnan [2012] concernant la reconnaissance des émotions basée sur la voix, de Fasel and Luetten [2003] concernant l'utilisation des expressions du visage, et plus généralement vers Zeng and Pantic [2009].

2.2.3 Les signaux EEG

Alors que l'utilisation de l'enregistrement électro-encéphalographique (EEG) a été longtemps confinée au domaine médical, les dernières années ont vu un intérêt croissant dans les interfaces cerveau-machine à base de EEG (BCI : Brain-Computer Interface) pour les applications destinées au grand public. En particulier l'enregistrement EEG a attiré l'attention des chercheurs dans le domaine de l'informatique affective dans le cadre d'un l'effort pour effectuer des tâches d'analyse de comportement humain, en particulier pour la reconnaissance automatique des émotions.

Comparé à d'autres modalités qui ont été prises en compte dans les travaux antérieurs sur la reconnaissance des émotions, comme la parole, les expressions du visage, les gestes ou d'autres signaux physiologiques [Chanel et al., 2006; Keltner et al., 2003; Schuller et al., 2011], l'EEG a l'avantage de capturer des informations relatives à des états émotionnels internes ne résultant pas nécessairement à des manifestations externes observables (en particulier à travers les modalités audio, visuelles ou de mouvement). Contrairement aux signaux périphériques, les signaux EEG reflètent une activité du système nerveux central (SNC). Plus de précisions sur le fonctionnement de l'électro-encéphalographie sont données en annexe de ce document (référence Annexe A).

Les Tableaux 2.3 et 2.6 résume quelques travaux qui ont été menés sur la reconnaissance automatique des émotions basée sur l'EEG.

2.2.4 Les systèmes multi-modaux

Le processus émotionnel peut également être étudié d'un point de vue pluridisciplinaire. L'émotion étant multimodale il est pertinent, dans le cadre de la reconnaissance des émotions en informatique affective, de s'orienter vers ce type de système afin de pouvoir bé-

Auteur	Titre	Participants/Dataset	Stimuli	Signaux	Matériel	Classes d'émotion	Classificateur
Chanel et al. [Chanel et al., 2006]	Emotion Assessment: Arousal Evaluation using EEG's and Peripheral Physiological signals	4 participants	Pictures (IAPS)	EEG, respiration belt, temperature, blood pressure, GSR	Biosemi Active Two with 64 electrodes	Calm, exciting	Naive Bayes Classifier, Fisher Discriminant Analyser
Herz et al. [Herz et al., 2007]	Using machine learning to predict learner emotional state from brainwaves	17 participants	Pictures (IAPS)	EEG	Pendant EEG, a portable wireless EEG, 3 electrodes	anger, boredom, confusion, contempt, curious, disgust, eureka and frustration	IBK (implementation of k-nearest neighbors)
Hoflings et al. [Hoflings et al., 2008]	Emotion Recognition using brain activity	ENTERFACE06 + 10 participants	Pictures (IAPS)	EEG	Biosemi Active Two with 64 electrodes	Positif, neutre, négatif	SVM, Neural Network, Naive Bayes Classifier
Khafili et al. [Khafili, 2009]	Emotion Recognition Using Brain and Peripheral Signals : Using Correlation Dimension to Improve the results of EEG	ENTERFACE06	Pictures (IAPS)	EEG, Galvanic Skin Resistance (GSR), Respiration, Blood pressure, Temperature	Biosemi Active Two with 64 electrodes	positively excited, negatively excited	Quadratic Discriminant Classifier
Li et al. [Li and Liu, 2009]	Emotion classification based on gamma-band EEG	10 participants	Pictures (smile and cry)	EEG, EMG	62 channel electrode cap	happiness, sadness	Linear Support Vector Machine (SVM)
Lin et al. [Lin et al., 2008]	Support Vector Machine for EEG Signal Classification during Listening to Emotional Music	26 participants	Music	EEG, EOG	32 channel Neuroscan module	joy, angry, sadness, pleasure	SVM
Nie et al. [Nie et al., 2011]	EEG-based Emotion Recognition during Watching Movies	6 participants	Movie clips	EEG	62 channel electrode cap	positive, negative	SVM
Takahashi et al. [Takahashi and Tsukaguchi, 2004]	Remarks on Emotion Recognition from Multi-Modal Bio-Potential Signals	12 participants	Film clips	EEG, pulse, GSR	Three electrodes (headband)	joy, anger, sadness, fear, relax	SVM

TABLEAU 2.3 – Liste et description de publications sur le reconnaissance de l'émotion utilisant les signaux périphériques et l'EEG.

néficer d'une plus grande richesse d'information. Cette richesse possède deux avantages essentiels : tout d'abord de s'assurer la présence d'information dans au moins une des modalités [Wagner et al., 2011], mais également, dans le cas où de l'information utile se situe au même moment dans différentes modalités, de pouvoir améliorer les performances du système de reconnaissance [Gunes and Piccardi, 2007; Kim and Andre, 2006; Paleari et al., 2010; Schuller et al., 2002]. En effet, dans ce cas de figure la richesse de l'information réside dans sa diversité et sa complémentarité.

Plusieurs travaux ont pris ce parti, en faisant usage de différentes combinaisons de modalités. Nous conseillons la lecture du papier de D'mello and Kory [2015] qui porte sur l'analyse de 90 systèmes multimodaux pour la reconnaissance de l'émotion. On peut se permettre de classer les systèmes multi-modaux en trois catégories, essentiellement suivant leur fréquence d'utilisation :

- les systèmes audio-visuels,
- l'utilisation conjointe de signaux physiologiques et EEG (Tableau 2.3),
- des combinaisons plus atypiques des différentes modalités exposées précédemment comme par exemple une utilisation conjointe d'informations de l'expression du visage et de signaux EEG [Basu and Halder, 2014] (Tableau 2.4).

Dans le cas où plus de deux modalités sont présentes, le système de reconnaissance est réalisé soit par paires de modalités, en système bimodal [Kessous et al., 2010], soit en fusionnant l'ensemble des modalités [Castellano et al., 2008; Wagner et al., 2011]. L'emploi de la multimodalité, du fait qu'elle contienne une plus grande richesse d'information, permet d'obtenir au mieux de meilleurs résultats de classification, sinon de s'assurer d'un résultat proche du meilleur obtenu avec une seule modalité (Tableau 2.5).

L'utilisation de données audio-visuelles pour la reconnaissance des émotions trouve un écho tout particulier dans le domaine de l'informatique affective. Ceci s'explique par deux raisons : d'une part le système d'acquisition est léger (une caméra et un microphone), et d'autre part les applications que l'on peut en faire sont variées et peuvent s'appliquer à des domaines aussi larges que la surveillance [Clavel, 2006] ou les interactions hommes-machines (HMI : Human-Computer Interface) [Truong et al., 2007; Vanpé, 2011]. Nous invitons le lecteur à s'orienter vers les papiers de Jaimes and Sebe [2007] et de Haq and Jackson [2010], qui résument les principaux travaux menés en reconnaissance de l'émotion basée sur des informations audio-visuelles.

La combinaison de signaux périphériques et EEG peut tout de même s'apparenter à un système multimodal du fait qu'elle reflète l'activité de deux systèmes nerveux distincts : respectivement le SNA et le SNC [Hagemann et al., 2003]. Cependant si des signaux périphériques sont enregistrés conjointement aux signaux EEG, c'est principalement dans le but de réaliser et d'améliorer la réjection d'artefacts [Sarilo et al., 2005; Wang et al., 2011, 2014b; Yuen et al., 2011]. Ces signaux annexes ne sont que rarement pris en compte dans les tâches de classification, soit dans un cadre de comparaison indépendantes des modalités

	Audio	Visuel (Visage)	Visuel (Corps)	Physiologique
Haak et al. [2009]		X		X
Bailenson et al. [2008]		X		X
Castellano et al. [2008]		X	X	X
Gunes and Piccardi [2006]		X	X	
Kim and Andre [2006]	X			X
Kessous et al. [2010]	X	X	X	
Chang et al. [2009]		X		X
Wagner et al. [2011]	X	X	X	
Schuller et al. [2002]	X		X	

TABLEAU 2.4 – Différentes combinaisons de modalités.

[Chanel et al., 2007; Koelstra et al., 2010b], soit dans un cadre de classification multimodale [Chanel et al., 2006; Kuncheva et al., 2011; Soleymani et al., 2012b].

Référence	Modalité 1	Modalité 2	Modalité 3	Nbr. de participants	Catégories d'émotion	Méthodes d'élicitation	Méthodes de classification	Meilleurs Résultats			
								Modalité 1	Modalité 2	Modalité 3	Multi-Modal
Bailenson et al. [2008]	Visuel (visage)	Signaux physiologiques : BP, ECG, SK, température	-	41 (not user specific)	amusement, sadness	vidéos	SVM	62%	41%	-	66%
Chang et al. [2009]	Visuel (visage)	Signaux Physiologiques : GSR, Temp., HR	-	240?	4 émotions (peur, amour, joie, surprise)	Vidéo (1h)	LVQNN	90%	88,33%	-	95%
Kim and Andre [2006]	Audio	Signaux Physiologiques : BVP, EMG, RSP, SC, Temp	-	3 (not user specific)	Valence/ Arousal (quadrants)	Jeu (Quiz)	LDA	54%	51%	-	55%
Chen et al. [1998]	Audio	Visual	-	2 [Silva et al., 1997]	6 émotions (joie, tristesse, colère, dislike, surprise, peur)	Acté		75%	69%	-	97%
Sebe et al. [2006]	Audio	Visual	-	38	7 émotions (joie, tristesse, peur, surprise, colère, dégoût, neutre) +4 HCI-related affects (intérêt, ennui, confusion, frustration)	Acté	BN	45%	56%	-	89%
Chanel et al. [2006]	EEG	Signaux Physiologiques	-	4 (eINTERFACE'06) [Savran et al., 2006]	calme, excitant	Images (IAPS)	FDA	55%	53%	-	54%
Dobrišek et al. [2013]	Audio	Visual	-	43 (eINTERFACE'05) [Martin et al., 2006]	6 émotions (colère, dégoût, peur, joie, tristesse, surprise)	Acté	SVM	72,5%	52,8%	-	77,5%
Kuncheva et al. [2011]	EEG	Signaux Physiologiques : GSR, BP, Resp.	-	1	Positif/Négatif	Sons	AdaBoost.M1 Naive Bayes	-	-	-	73,80%
Soleymani et al. [2012b]	EEG	Eye Gaze	-	24	Valence/Arousal	20 vidéos	SVM	62,1%	71,1%	-	76,4%
Wagner et al. [2011]	Visuel (Visage)	Audio	Mouvements (Manette Wii)	21 (not user specific)	Positif/Négatif	Acté	Naive Bayes	51%	48%	42%	55%
Schuller et al. [2002]	Audio	Visuel (corps)	-	15 (not user specific)	7 émotions (colère, dégoût, peur, joie, tristesse, surprise, neutre)	Acté	DTW	84-85%	85%	-	89%
Palenri et al. [2010]	Visuel	Audio	-	44 (eINTERFACE'05)	6 émotions (colère, dégoût, peur, joie, tristesse, surprise)	Acté	NN	60-65,5%	56,5%	-	50,3%
Caridakis and Castellano [2007]	Visuel (Visage)	Audio	Visuel (corps)	10	8 émotions	Acté	NN	48,3%	57,10%	67,10%	78,3%

TABLEAU 2.5 – Liste et description de publications sur la reconnaissance de l'émotion utilisant des systèmes multi-modaux.

2.3 Une vue d'ensemble des systèmes précédents pour la reconnaissance des émotions basée sur l'EEG

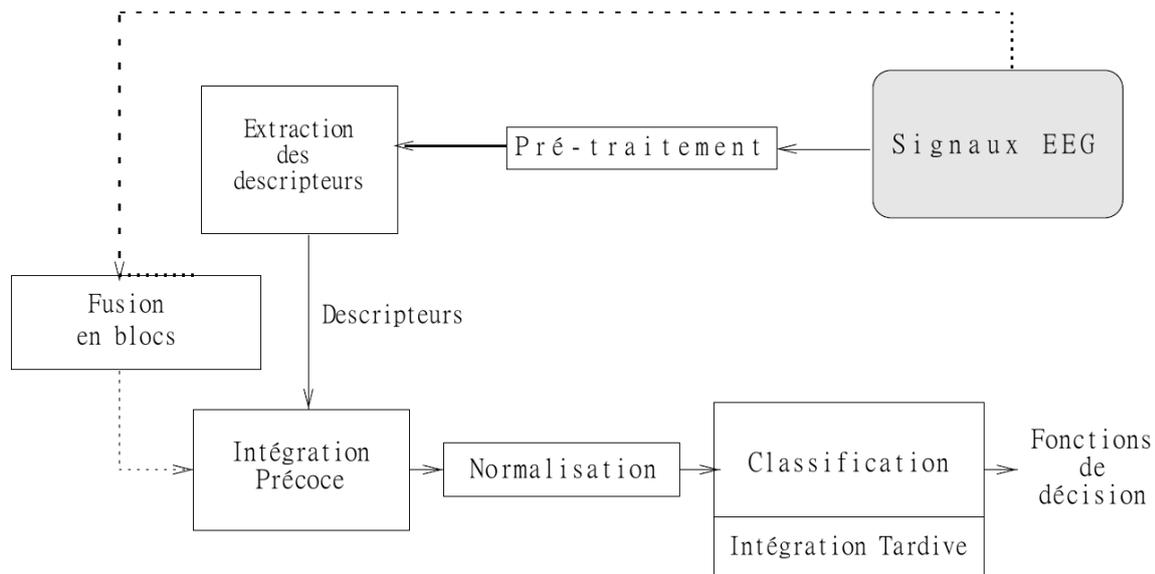


FIGURE 2.7 – Diagramme du système de classification.

La reconnaissance de l'émotion basée sur l'EEG est un sujet de recherche relativement récent qui possède de nombreuses difficultés et certains challenges. Un des défis majeurs, qui est inhérent à l'analyse des signaux EEG en général, est lié au fait que les signaux enregistrés sont facilement contaminés par de lourds artefacts provenant de sources de bruit variées, à la fois physiologiques (notamment dus à l'activité oculaire et aux muscles de la tête) et environnementales (dus aux interférences électromagnétiques). C'est la raison pour laquelle l'enregistrement de signaux EEG est habituellement réalisé dans des conditions comportant de nombreuses contraintes (souvent en insistant auprès du participant de rester le plus stable possible), ce qui ajoute à la complexité du processus d'enregistrement des données.

En réalité, la disponibilité de données de qualité pouvant être utilisées dans le cadre du développement d'algorithmes d'apprentissage automatique servant à traiter notre problème est critique. Des efforts importants ont été faits en ce sens [Koelstra et al., 2012; Sarvan et al., 2006], bien que les types de stimuli et les stratégies d'annotation de l'émotion adoptés varient suivant les travaux, ce qui rend difficile l'exploitation de plus d'un dataset à la fois. Les principaux corpus accessibles seront présentés dans la Section 2.5.

Afin de s'assurer d'une bonne utilisation du système de classification mis en place pour le système de reconnaissance automatique des émotions, des pré-traitements doivent être appliqués aux différentes données (Section 2.3.1). Ce pré-traitement se rapporte à deux élé-

ments essentiels survenant en amont de la classification par le biais d'un nettoyage et d'une normalisation des données, mais également le choix de la taille des fenêtres temporelles et le choix d'une stratégie d'intégration temporelle dont les détails sont fournis dans la section 2.3.2. Dans le système de classification, ces étapes sont réalisées avant l'extraction des descripteurs.

2.3.1 Les pré-traitements des signaux EEG

Les signaux EEG sont reconnus pour être sensibles à de nombreuses perturbations entraînant l'apparition d'artefacts dans les signaux enregistrés. La première étape consiste donc à retrouver les informations utiles au sein de ces signaux bruités. Les artefacts sont créés essentiellement par des activités musculaires telles que le clignement des yeux ou le fait de serrer les mâchoires, mais également par des éléments extérieurs comme l'éclairage de la pièce. Ce dernier artefact ne représente pas une réelle contrainte puisqu'il en résulte une composante fréquentielle continue et constante à 50 Hz dans les signaux EEG.

Normalisation

La normalisation du signal est effectuée afin de palier les problèmes de variabilité des signaux créés par des conditions d'enregistrement qui ne peuvent que très difficilement scrupuleusement identiques. Pour chaque canal EEG, le signal est normalisé à partir du signal $x_c(n)$:

1. $\tilde{x}_c(n) = x_c(n) - \bar{x}_c(n)$ avec $\bar{x}_c(n) = \frac{1}{L} \sum_{l=0}^{L-1} x_c(l)$ qui représente la moyenne du signal,
2. $\hat{x}_c(n) = \frac{\tilde{x}_c(n)}{\max_n |\tilde{x}_c(n)|}$.

Cette normalisation a pour but de limiter les effets de cette variabilité sur le système de classification.

Sous-échantillonnage

Les enregistrements étant souvent longs avec des fréquences d'échantillonnage élevées, on réalise un sous-échantillonnage des données. Cela permet d'alléger les données en prenant garde de conserver une fréquence suffisamment élevée pour l'extraction de descripteurs fréquentiels situés dans la bande γ d'ondes EEG qui peut contenir de l'information utile.

2.3.2 Le système de classification

Le système de classification est composé de deux étapes :

- une phase d'apprentissage du classifieur qui détermine un modèle de chaque classe à partir de données d'apprentissage,

- et une phase de test qui permet d'évaluer les performances du système de classification à l'aide de données de test.

Plusieurs classifieurs peuvent être utilisés dans le cadre de systèmes pour la reconnaissance des émotions basés sur des signaux EEG. Il est toutefois difficile de déterminer avec certitude quel est le meilleur classifieur. Les travaux de [Sohaib et al. \[2013\]](#) tendent à montrer que, parmi un choix composé d'un arbre de régression, une méthode des k plus proches voisins (K-Nearest Neighbor ou KNN), un réseau bayésien (Bayesian Network ou BN), une machine à vecteurs support (Support Vector Machine ou SVM) et des réseaux de neurones artificiels (Artificial Neural Networks ou ANN), les classifieurs de type KNN et SVM sont les plus appropriés. Les résultats de [Valenzi et al. \[2014\]](#) parviennent à la même conclusion concernant le choix des SVM pour les systèmes de classification de l'émotion basés sur l'EEG.

Cela se conforme aux observations de [Chanel et al. \[2007\]](#) mais également de [Wang et al. \[2011\]](#), où la performance des classifieurs SVM est meilleure que celles obtenues avec des réseaux de neurones ou un algorithme KNN pour des descripteurs temporels et fréquentiels. De même, l'emploi de SVM permet d'obtenir des résultats plus constants sur les deux axes de valence et d'activation [[Horlings et al., 2008](#)].

Les caractéristiques

Un certain nombre d'études récentes ont examiné la question de la sélection de caractéristiques pour la reconnaissance des émotions basée sur l'EEG. Les caractéristiques spatio-temporelles et spatio-spectrales ont été explorées. En ce qui concerne les plus anciennes, les caractéristiques dites standard [[Picard et al., 2001](#)], relatives à des informations statistiques extraites de la forme d'onde temporelle enregistrée par différentes électrodes, comme la moyenne, l'écart-type, etc., se sont avérées utiles [[Chanel et al., 2006](#); [Khalili, 2009](#)]. Cependant les caractéristiques contenant de l'information spectrale ont été plus largement utilisées.

Les travaux de Li et al. [[Li and Lu, 2009](#)] sont axés sur les caractéristiques spatio-spectrales où des *Common Spatial Patterns* (CSP) sont extraits de la bande γ ($>30\text{Hz}$)¹¹. Avec une méthode de sélection de bande fréquentielle associée aux CSP qui sont calculés à partir des données filtrées, les auteurs ont réalisé une classification binaire le long de la dimension valence, correspondant à une discrimination entre les émotions positives et négatives. Xu et al. [[Xu and \(Kostas\) Plataniotis, 2012](#)] ont exploré les caractéristiques spatio-temporelles, avec des caractéristiques de *Higher-Order Crossings* (HOC) [[Kedem and Frey, 1994](#)] et des caractéristiques du domaine spectral, parmi lesquelles les ondelettes et l'énergie des composantes spectrales de différentes sous-bandes fréquentielles au sein de la gamme 8-30 Hz. Les auteurs ont obtenu de meilleurs résultats avec des caractéristiques de sous-bandes spectrales qu'avec des caractéristiques standards. Plus récemment, Jenke [[Jenke et al., 2013](#)] s'est axé sur la sélection d'électrode et de caractéristique compte tenu des statistiques des signaux filtrés dans le répertoire des bandes α , β , θ , ainsi que des paramètres de Hjorth [[Hjorth, 1970](#)]. Un résumé de quelques travaux portant sur la reconnaissance des émotions

11. voir "Les rythmes EEG" en Annexe A.1.2

Référence	Nombre d'électrodes EEG	Nombre de participants	Stimuli	Emotions	Descripteurs	Classifieur	Résultats EEG
Khalili [2009]	54	1 (eINTERFACE'06)	Images (IAPS)	positivement excité, négativement excité, calme	statistiques	QDA	67%
Schaaff and Schultz [2009a]	4	5	Images (IAPS)	plaisant, neutre, déplaisant	fréquentiels	SVM (noyaux RBF)	44%
Yuen et al. [2011]	62	6	Images (visages)	colère, tristesse, surprise, joie, neutre	statistiques	Réseaux de neurones	95%
Sohaib et al. [2013]	8	15	Images (IAPS)	neutre, calme excité positif, calme excité négatif	min, max, moyenne, écart-type Alzoubi et al. [2009]	SVM	56%
Li and Lu [2009]	62	10	Images	joie, tristesse	CSP	SVM	94%
Koelstra et al. [2010b]	32	6	Clips vidéos	2 classes de valence	PSD, CSP (2 composantes)	SVM	59%
Liu et al. [2015]	30	12	Vidéos	neutre, heureux, triste, tendu, dégoût	PSD, AI	SVM	85%
Jie et al. [2014]	40	32 (DEAP Dataset)	Clips vidéos	2 classes de valence	Sample entropy	SVM	80%
Valenzi et al. [2014]	32	9	Vidéos	amusé, dégoûté, triste, neutre	fréquentielles	SVM	97%
Jenke et al. [2013]	64	16	Images (IAPS)	heureux, curieux, colère, triste, calme	statistiques fréquentielles	QDA	58%
Bastos-Filho et al. [2012]	4	32 (DEAP Dataset)	Clips vidéos	calme, stress	PSD	KNN	70%
Schaaff and Schultz [2009b]	4	5	Images (IAPS)	plaisant, neutre, déplaisant	statistiques fréquentielles	SVM	47%
Wang et al. [2011]	64	5	Vidéos	joie, relaxé, triste, peur	PSD	SVM	67%

TABLEAU 2.6 – Principaux résultats dans la littérature sur la reconnaissance des émotions basée sur l'EEG (PSD : Power Spectral Density, CSP : Common Spatial Pattern, AI : Asymmetry Index; SVM : Support Vector Machine, QDA : Quadratic Discriminant Analysis, KNN : K-Nearest Neighbor).

basée sur l'EEG est fourni dans le Tableau 2.6.

Intégration Précoce

Les descripteurs peuvent être extraits sur des tailles de fenêtre temporelle différentes, et avec des taux de recouvrement qui varient. Cela nous amène donc à se poser la question de la méthode à utiliser pour effectuer un regroupement de ces données.

En optant pour une taille de fenêtre temporelle de référence, on va chercher à recalculer les données sur la base de cette fenêtre de temps. Cette méthode représente ce que l'on appelle l'intégration précoce du fait qu'elle intervient en amont du classifieur [Joder et al., 2009] (voir Figure 2.7). Deux options sont envisageables pour mettre en place ce type d'intégration :

- La taille de fenêtre temporelle la plus petite sert de format de référence. Si le segment temporel à intégrer est plus grand que le format de référence, il est alors divisé en plusieurs fenêtres dont la référence temporelle est identique au format de référence, et les données de ce segment sont copiées dans chacune de ces fenêtres.
- L'autre option repose sur le choix de la taille de fenêtre temporelle la plus grande comme format de référence. Les sous-segments qui se trouvent inclus temporellement au segment de référence sont moyennés pour former la nouvelle fenêtre temporelle correspondante.

Intégration Tardive

A l'inverse de l'intégration précoce présentée précédemment, l'intégration tardive intervient en aval du classifieur (voir Figure 2.7). Le bloc de descripteur est envoyé au classifieur, qui pour chaque segment fournit une probabilité de "bonne classification". La probabilité finale du bloc correspond au produit de chaque probabilité [Joder et al., 2009].

2.4 L'expérience des émotions

Dans cette partie, deux composantes distinctes sont abordées dans le but de mieux aborder la partie concernant la création de notre corpus (**Chapitre 3**). Ces deux composantes sont l'élicitation et la retranscription de l'émotion. Ces sujets sont depuis longtemps abordés dans la littérature et les mécanismes sous-jacents qui y sont impliqués méritent d'être discutés afin de mettre en lumière les choix dans les stratégies qui sont les nôtres.

2.4.1 Retranscrire l'émotion

De nombreux facteurs viennent influencer la manière dont nous allons traduire une émotion que nous avons pu ressentir à un moment donné, comme l'illustre bien Scherer [2005a]

(voir Figure 2.8). Dans cette partie nous nous intéressons aux différentes stratégies adoptées pour retranscrire l'émotion. Parmi ces stratégies, on peut notamment distinguer l'utilisation de catégories et la représentation à l'aide de différentes dimensions.

Un des défis majeurs dans la retranscription d'une émotion et de pouvoir adopter une stratégie intuitive, à savoir peu coûteuse en temps et en réflexion. C'est à ces critères que vont chercher à répondre les différentes stratégies d'annotation. Après avoir exposé les théories sur lesquelles reposent les méthodes de retranscriptions de l'état émotionnel, les principales stratégies d'annotations seront explicitées.

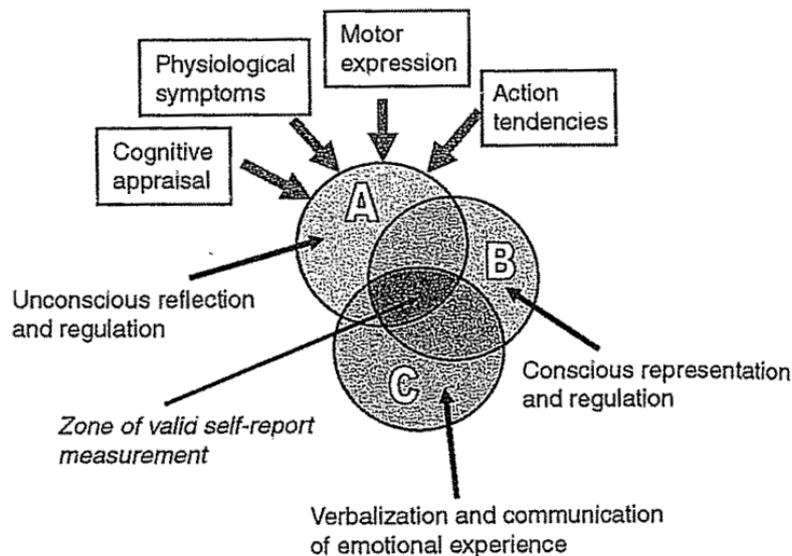


FIGURE 2.8 – Les trois modes de représentation des changements dans les composantes de l'émotion : l'inconscience, la conscience et la verbalisation, tiré de Scherer [2005a].

2.4.1.1 Les représentations de l'émotion

Le participant doit être en mesure de définir l'émotion qu'il vient de ressentir de manière immédiate/spontanée. On doit chercher à limiter l'introspection qui devient sur le long terme trop coûteuse au niveau de la charge cognitive.

Il faut donc trouver une représentation simple, mais non simpliste, du phénomène émotionnel. Parmi ces représentations, nous pouvons distinguer deux approches principales :

- une approche catégorielle,
- et une approche dimensionnelle.

Les catégories d'émotions : une distinction sémantique

Les catégories d'émotions ont pour objectif de traduire le phénomène émotionnel par des concepts définis et limités.

- **Le Big Six**, est un cas de définition catégorielle de l'émotion proposé par Ekman [1992] où il propose différentes catégoriques d'émotions dites basiques : la joie, la tristesse, la peur, la colère, le dégoût et la surprise. Il s'appuie pour cela sur l'observation d'expressions du visage à travers différentes cultures. Pour ces catégories d'émotions, les expressions se trouvent être identiques, on peut alors parler d'expressions *universelles*. Ekman [1999] justifie l'utilisation du terme "basique" à travers deux significations de ce dernier. Si dans un premier temps il s'était orienté vers la théorie darwinienne pour exposer le caractère universel de ces catégories, il relativise toutefois cette notion en y ajoutant un point de vue ontologique, qui appuie le rôle social des émotions par le biais des conventions. Il ne renie pas cependant l'aspect de la théorie de Darwin dans l'influence qu'a le passé ancestral dans l'évaluation que nous faisons des événements présents, mais également dans la fonction primaire qu'a l'émotion de mobiliser l'organisme afin de répondre rapidement lors d'importantes rencontres interpersonnelles.

- Les **termes émotionnels**. Plusieurs listes de termes émotionnels ont été proposées, aussi bien en anglais (GALC [Scherer, 2005b], ANEW [Bradley and Lang, 1999]) qu'en français (EMOTAIX [Piolat and Bannour, 2009]), cherchant à couvrir autant que faire ce peut le champ des états émotionnels. Bien qu'ils soient issues du langage courant, une des difficultés de l'utilisation de ces termes émotionnels réside dans un aspect linguistique, plus particulièrement au niveau du signifié des termes utilisés. Le signifié renvoie à l'objet défini par le terme, or dans le cadre de mots cherchant à identifier des catégories d'émotions, il peut y avoir une confusion et une incertitude non pas par la dénotation de ce mot mais bien par la connotation, à savoir l'ensemble de ces éléments de sens (subjectifs, culturels, inconscients, ...) qui pourrait venir perturber le signifié. Parmi les autres contraintes liées à l'utilisation de termes émotionnels, on peut évoquer la dépendance à la langue maternelle de l'individu, mais également des listes qui se veulent très exhaustives et aboutissent à des bases d'une taille importante. Ces listes sont difficilement exploitables dans le cadre de travaux de reconnaissance des émotions en informatique affective.

Le cône de Plutchik

Nous avons choisi de distinguer le cône de Plutchik [Plutchik, 1984] de la représentation des deux approches catégorielles et dimensionnelles des émotions, car ce choix de représentation se situe entre les deux, c'est un pont entre chaque approche.

Suivant une approche clairement darwinienne, puisqu'il pose comme postulat que le concept des émotions doit s'appliquer chez l'ensemble des êtres vivants, aussi bien chez les hommes que chez les animaux inférieurs, il recherche dans le comportement social [Van Hooff, 1973] et d'adaptation [Scott, 1958] une liste finie d'émotions fondamentales.

Cette recherche d'émotions fondamentales repose sur une analogie avec les couleurs primaires et secondaires, à laquelle il souhaite associer une représentation. Les éléments lui servant à poser son modèle structurel sont les suivants :

1. Les émotions varient par leur *intensité* (peur/panique), leur *similarité* (la honte et la

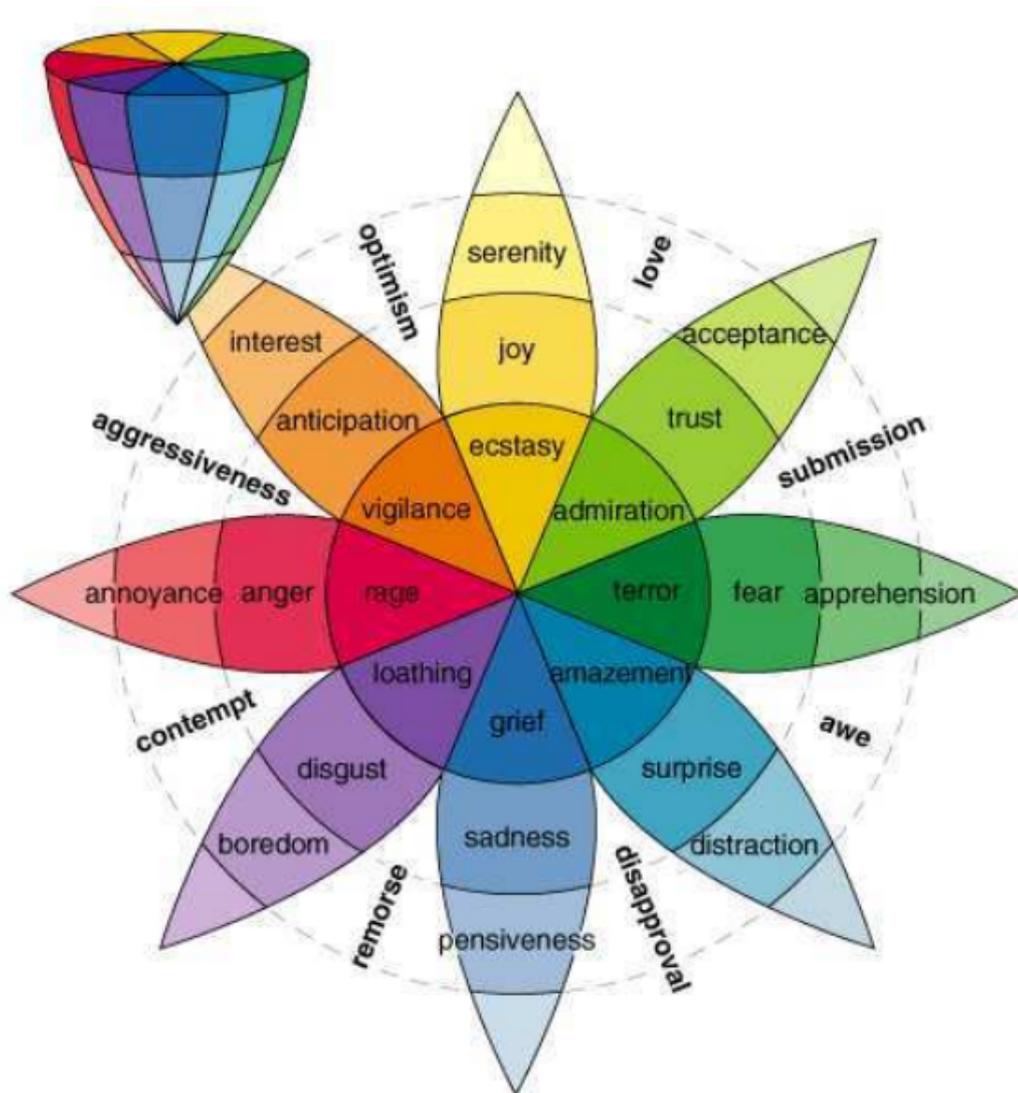


FIGURE 2.9 – Le cône des émotions de Plutchik [Plutchik, 1984]. Les 8 émotions élémentaires sont placées par opposition au centre de la roue.

culpabilité sont plus similaires que la joie et le dégoût), et par leur *polarité* (la joie est l'opposé de la tristesse, la haine de l'amour) ,

2. Des émotions sont primaires ou fondamentales, et d'autres sont secondaires ou dérivées, de la même manière que certaines couleurs sont primaires et les autres secondaires.

Afin d'aboutir à une représentation de cette analogie, il est amené à se poser les questions suivantes : combien y a-t-il d'émotions fondamentales et comment devraient-elles être étiquetées ?

Le cône qu'il propose (Figure 2.9) est constitué de 8 émotions primaires qu'il oppose par paires. Plus on descend dans le cône, plus les émotions sont difficiles à distinguer les unes des autres. L'axe vertical illustre les variations en intensité ou en activation des émotions. Il place les émotions dans différentes dimensions qui se décomposent de la manière suivante : au centre de la roue se trouvent les *émotions élémentaires*, et l'on trouve ensuite les *émotions secondaires*, qui correspondent à un mélange d'émotions élémentaires dont le degré d'intensité diminue au fur et à mesure que l'on s'éloigne du centre.

Les dimensions émotionnelles

Parmi les dimensions les plus fréquemment utilisées dans le cadre de l'annotation d'une émotion on distingue trois dimensions [Mehrabian and Russell, 1974; Russell and Mehrabian, 1977] :

- la *valence*, qui représente l'aspect positif ou négatif d'une émotion (plaisir vs. déplaisir),
- l'*activation*, le fait d'être réactif à un stimuli (calme vs. excité), aussi bien de manière physique que de manière psychologique.
- et la *dominance*, qui fournit une indication sur la maîtrise que la personne a de son émotion (dominé vs. dominant).

Tous ces axes représentent des dimensions bipolaires et indépendantes. Il est important de noter ici que l'axe d'activation ne correspond pas forcément à l'intensité de l'émotion. En effet, plusieurs théories quant à la place de l'intensité au sein de ces dimensions ont été exposées. Le cas de l'intensité de l'émotion ressentie est quant à lui plutôt particulier, et a été la source de nombreux questionnements résumés dans Coppin and Sander [2012]. L'intensité peut soit être considérée comme étant prise en compte de manière implicite dans les axes de valence, d'activation et de dominance, ou bien il est considéré comme étant à part. Il faut alors associer un axe d'intensité pour chacune des dimensions utilisées.

Discussion

Nous n'avons présenté ici que certaines représentations de l'émotion, à savoir les modèles les plus couramment évoqués et utilisés dans la littérature de l'informatique affective.

Le choix d'une approche catégorielle de l'émotion implique la nécessité d'avoir des termes émotionnels distincts du point de vue sémantique lorsque l'on cherche à se tendre

vers une ensemble d'émotions dites *basiques* (Tableau 2.7). Combien d'émotions pouvons nous qualifier de *basiques*? Bann and Bryson [2012] proposent une analyse de cette distinction sur les principales propositions d'émotions de base, et parviennent à la conclusion que la série de termes proposée par Ekman [Ekman, 1999] fournit des catégories plus distinctes au sein de ces dernières. Ils aboutissent cependant à la meilleure distinction avec une nouvelle combinaison de huit termes émotionnels : *accepting, anxious, ashamed, contempt, joyful, miserable, pleased, stressed*¹².

Théories	Emotions de base identifiées
Izard [Izard, 1992]	colère, mépris, dégoût, détresse, peur, culpabilité, intérêt, joie, honte, surprise
Russell [Russell, 1980]	en colère, déprimé, peiné, excité, misérable, heureux, détendu, somnolent
Plutchik [Plutchik, 1984]	anticipation, joie, confiance, peur, colère, surprise, tristesse, dégoût
Ekman [Ekman, 1999]	colère, dégoût, peur, joie, tristesse, surprise
Tomkins [Tomkins and Mc Carter, 1964]	colère, intérêt, mépris, dégoût, détresse, peur, joie, honte, surprise
Johnson-Laird [Johnson-Laird and Oatley, 1989]	colère, dégoût, anxiété, bonheur, tristesse

TABLEAU 2.7 – Les principales théories des émotions *basiques*.

D'un point de vue expérimental, cela conduit à une dépendance des propositions à la langue dans laquelle elles ont été faites, mais également à l'interprétation qu'en fait le participant, de manière consciente ou non, et de la perception qu'il en a. Prenons comme exemple les termes de *joyful* et *pleased* proposés par Bann and Bryson [2012]. Pour un non-anglophone, ou plus précisément un *non-native English speaker*, ces deux mots peuvent facilement être placés sous le même terme de *content*¹³. On pourra également remarquer que les définitions de *joyful*¹⁴ et de *pleased*¹⁵ proposées par l'*Oxford English Dictionary* peuvent prêter à confusion.

C'est pour ces différentes raisons que certains travaux menés dans le domaine de l'informatique affective se tournent, comme nous allons le voir dans les sections suivantes, vers une représentation dimensionnelle de l'émotion qui réduit le nombre de termes employés au nombre de dimensions considérées. Ce choix de représentation est également consolidé par le fait que ces catégories d'émotions peuvent naturellement être transposées dans les représentations dimensionnelles du phénomène émotionnel. Les termes émotionnels sont susceptibles d'être classés suivant les dimensions émotionnelles [Bradley and

12. Acceptant, anxieux, honteux, mépris, joyeux, misérable, satisfait, stressé.

13. Le dictionnaire Larousse propose deux traductions anglaises pour *content* : au sens de *satisfait* (*pleased*) et de *heureux* (*happy*).

14. Joyful : Feeling, expressing, or causing great **pleasure and happiness** (*Oxford English Dictionary*).

15. Pleased : Feeling or showing **pleasure and satisfaction**, especially at an event or a situation (*Oxford English Dictionary*).

Lang, 1999; Warriner et al., 2013]. Pour une lecture approfondie dans ce domaine, nous orientons le lecteur vers les papiers de Galati and Sini [2000]; Goossens [2005]; Leveau et al. [2011] concernant les directions prises dans la terminologie française, et vers Bann and Bryson [2012]; Johnson-Laird and Oatley [1989] pour un aperçu des travaux en terminologie anglaise.

Les bases des théories qui s'appliquent à l'annotation de l'émotion ressentie ont été posées dans cette section. On va pouvoir à présent décrire les applications qui en ont été faites d'un point de vue expérimental en informatique affective. Même si l'on peut déjà commencer à s'interroger sur la pertinence du choix d'une annotation discrète plutôt que continue (ou inversement) [Barrett, 1998], la première interrogation réside dans l'opposition d'une annotation statique ou dynamique dans les méthodes d'annotations qui découlent de ces théories.

2.4.1.2 Les stratégies d'annotation

Les stratégies d'annotations fréquemment rencontrées dans le cadre de l'informatique affective reposent sur les représentations des descripteurs dimensionnels de l'émotion. Elles proviennent essentiellement de la représentation tridimensionnelle (PAD : Plaisir, Activation, Dominance) de l'émotion par Mehrabian and Russell [1974]. D'autres stratégies peuvent toutefois être adoptées, comme la retranscription verbale du ressenti ou encore le choix parmi une liste de termes émotionnels.

Toutefois, les choix symboliques sont privilégiés afin de diminuer les écarts dans les annotations effectuées par différents sujets. Ce sont le choix de ces représentations dimensionnelles qui vont être discutées dans cette section.

L'annotation statique

Nous entendons par *annotation statique*, la retranscription d'une émotion unique ou globalement ressentie par le participant soumis à un ou plusieurs stimuli. Cette annotation a pour problématique majeure : quel est le nombre de dimensions suffisantes pour décrire de façon raisonnablement précise l'émotion ?

Le modèle bi-dimensionnel

Le *circumplex* proposé par James A. Russell [Russell, 1980] est un modèle bidimensionnel et circulaire formé avec les axes de valence et d'activation, respectivement pour l'axe horizontal et vertical (Figure 2.10).

Pour réduire le nombre de dimensions du PAD à un espace bidimensionnel, il souligne qu'une des propriétés du langage de l'affect est la possibilité de définir n'importe quel terme lié à l'affect comme une combinaison des deux composantes de valence et d'activation [Russell and Pratt, 1980].

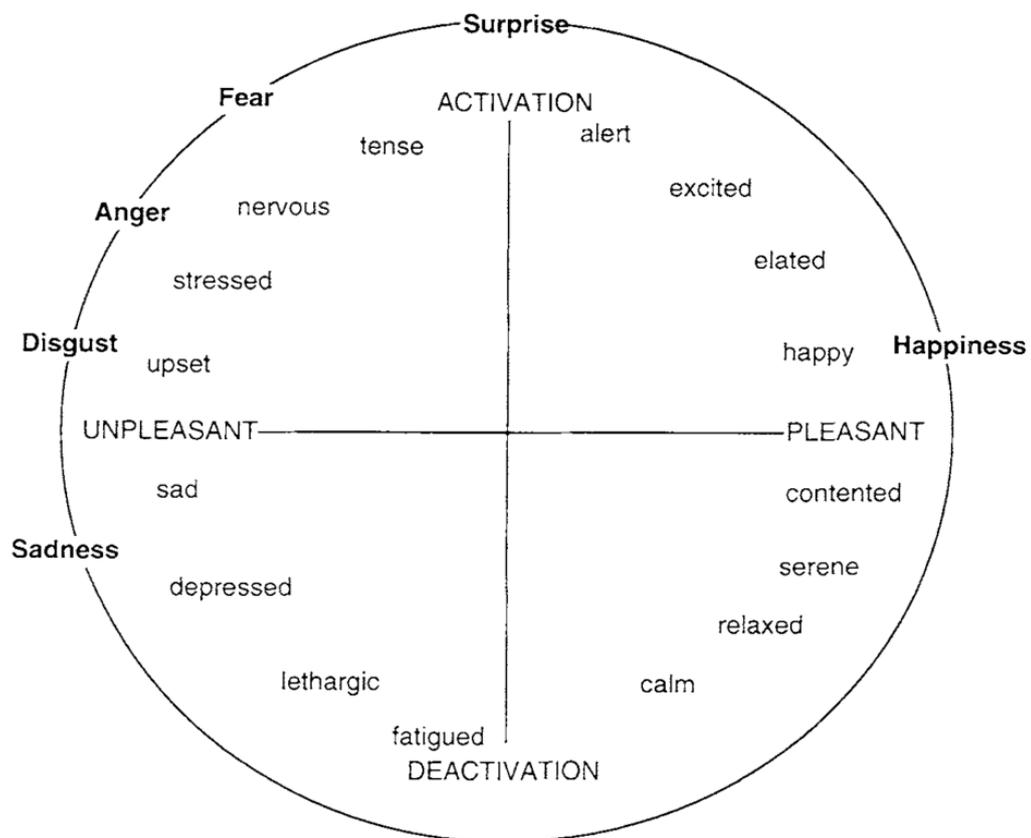
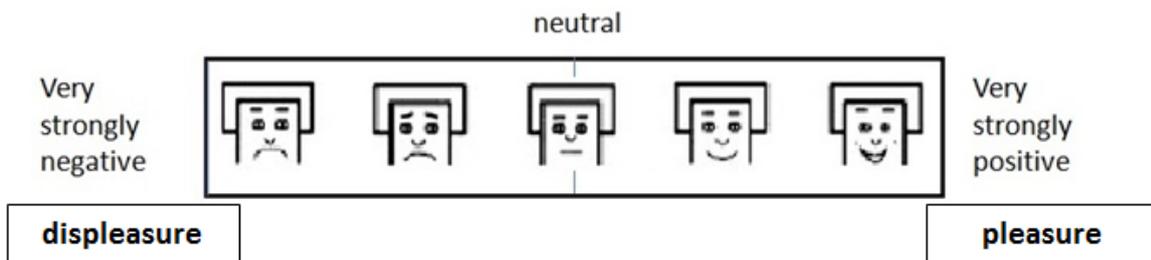


FIGURE 2.10 – *Circumplex* de **Russell and Barrett [1999]**, modifié à partir de la version de **Feldman Barrett and Russell [1998]**. Le disque interne représente un schéma de l'affect central (*core affect*), et le disque extérieur représente plusieurs émotions prototypiques liées à leurs caractéristiques affectives.

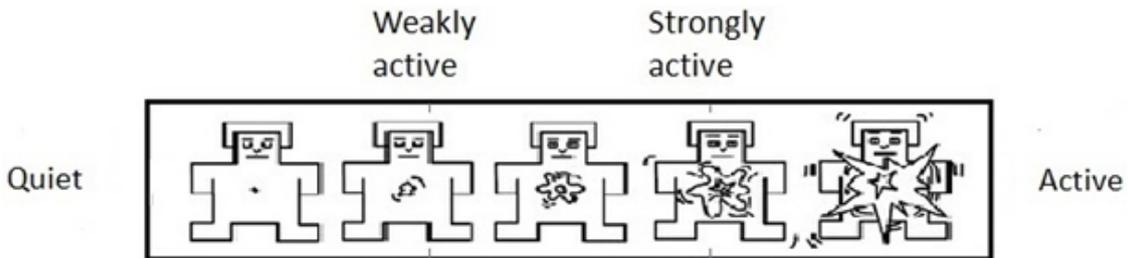
Ce modèle a fait l'objet de nombreuses discussions et modifications [Mehrabian, 1997; Yik et al., 2011], même de la part de Russell lui-même, en collaboration avec Barrett [Russell and Barrett, 1999].

Self-Assessment Manikin [Bradley and Lang, 1994]

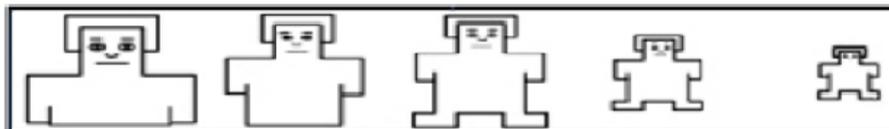
Le modèle tridimensionnel (PAD) [Mehrabian and Russell, 1974] a été largement étudié et appliqué dans de nombreux domaines, aussi bien dans les neurosciences [Posner et al., 2005] que dans le domaine de la publicité [Havlena and Holbrook, 1986]. Il est à l'origine d'un modèle d'annotation très répandu : le Self-Assessment Manikin (SAM) [Bradley and Lang, 1994]. Cette représentation est basée sur l'utilisation de petits personnages, également appelés *mannequins*, qui possèdent des variations différentes suivant chaque axe (Figure 2.11). C'est une représentation illustrée du PAD. Différents niveaux de graduations peuvent être utilisés, mais le plus souvent on a recours à des versions possédant une graduation de 5 ou 9 mannequins.



(a) Représentation de la dimension *Valence*.



(b) Représentation de la dimension *Activation*.



(c) Représentation de la dimension *Dominance*.

FIGURE 2.11 – Exemple d'une représentation des axes de valence et d'activation à l'aide du Self-Assessment Manikin.

Les variantes dans l'utilisation de cette méthode se situent sur son échelle. Elle est souvent discrète [Savran et al., 2006; Soleymani et al., 2012a], quelquefois continue [Koelstra

et al., 2012], et utilisée avec différentes graduations (allant de 1 à 5, ou de 1 à 9). Il s'agit d'un outil simple à implémenter qui en fait un outil de prédilection pour les expériences menées en informatique affective. L'avantage principal du SAM, à savoir sa facilité de compréhension pour l'annotateur, fait qu'il est utilisé dans de divers domaines cherchant à obtenir des informations sur les réponses émotionnelles [Barrett, 1998; Morris, 1995].

Il ne fait pas appel à une réflexion sur la signification par exemple de termes émotionnels dans l'utilisation de formes verbales, mais seulement à une projection de son ressenti sur des figures, ce qui allège la charge cognitive demandée à l'annotateur et il en découle naturellement un gain de temps. Ces petits mannequins se sont également révélés être plus attractifs pour les annotateurs qu'une retranscription verbale de l'émotion ressentie, ce qui permet de retenir leur attention plus longtemps [Lang, 1985]. Des personnes de tous les âges [Lang, 1985] et de toutes les cultures [Morris, 1995] peuvent s'y identifier facilement.

L'annotation dynamique

Différentes stratégies ont été adoptées dans le passé pour annoter l'émotion de manière dynamique. Ces stratégies reposent sur des retranscriptions sur des dimensions continues de l'émotion ressentie. On peut notamment citer deux d'entre elles : FeelTrace [Cowie et al., 2000] et GTrace¹⁶. Bien qu'elles cherchent toutes les deux à retranscrire une dynamique de l'émotion, les méthodes d'annotation diffèrent. En effet, FeelTrace (Figure 2.12) adopte une annotation conjointe des axes de Valence et d'Activation en se déplaçant dans ce plan, tandis que la méthode adoptée dans GTrace (Figure 2.13) correspond à une retranscription des variations dans l'émotion ressentie sur l'axe Valence (ou Activation) dans un premier temps, puis sur l'axe Activation (respectivement Valence).

Les contraintes liées à ces systèmes d'annotation dynamique d'un point de vue expérimental seront débattues plus en détails dans le **Chapitre 3**.

16. <http://emotion-research.net/toolbox/toolboxlabellingtool.2011-07-08.0084929013>

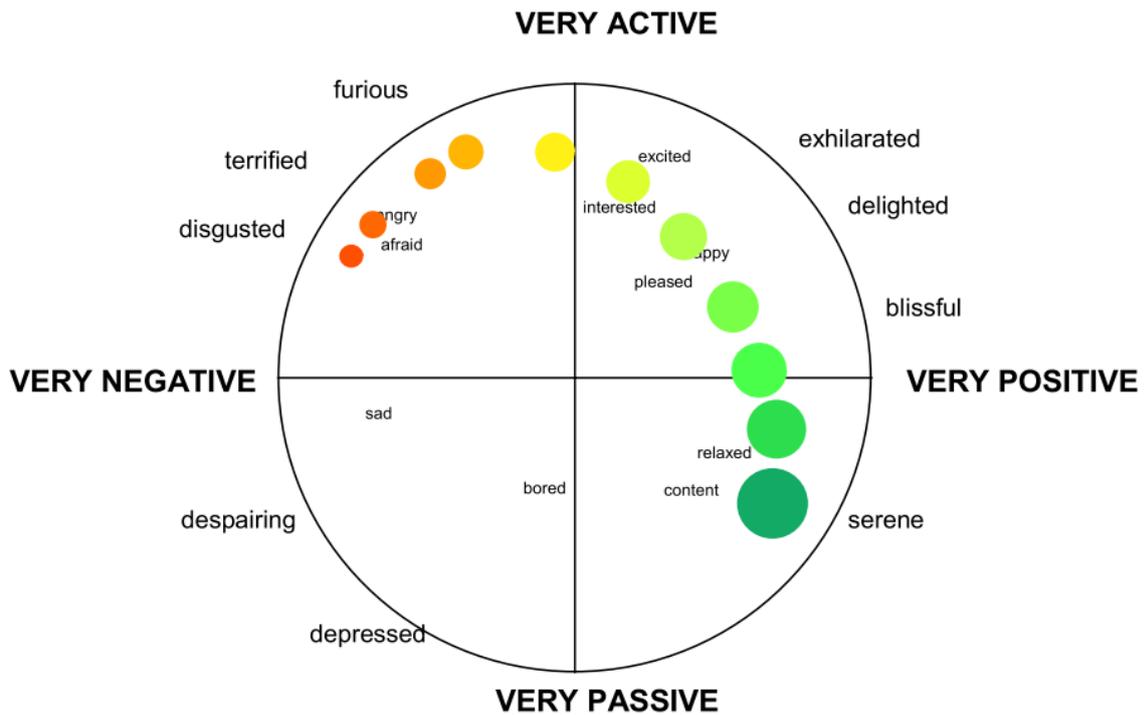


FIGURE 2.12 – Le système d’annotation bi-dimensionnel de Feeltrace.

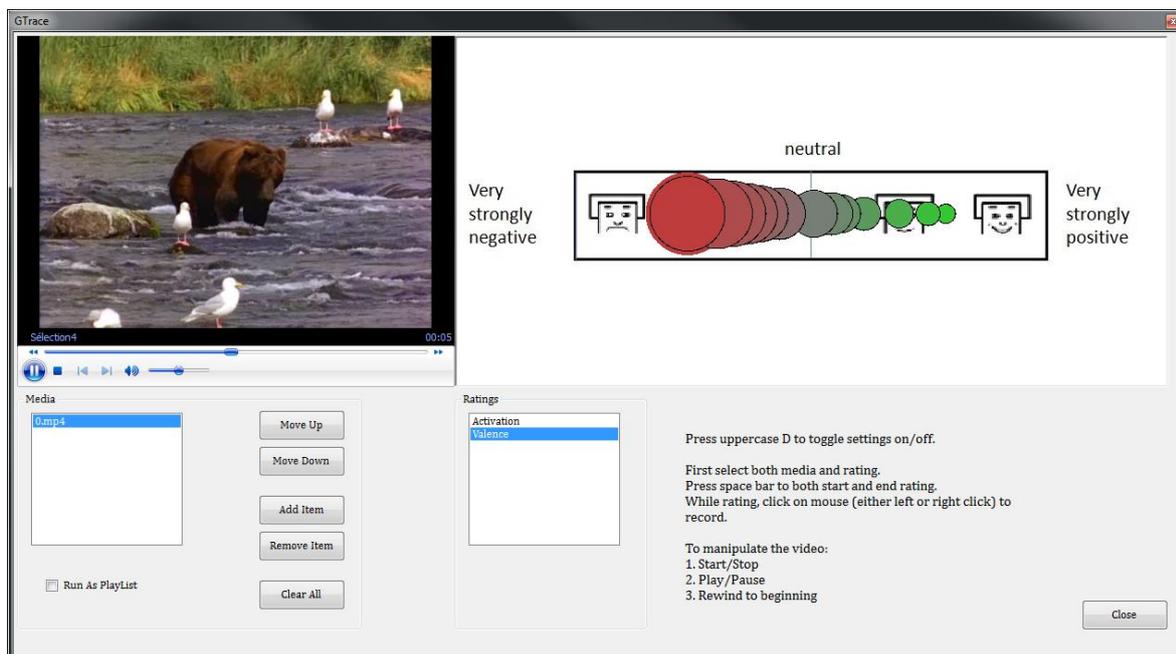


FIGURE 2.13 – Le système d’annotation de GTrace.

Discussion

Comme nous avons pu le constater, l’annotation statique pose en premier lieu la question du nombre de dimensions nécessaires à la retranscription d’une émotion, et il s’en suit avec l’annotation dynamique une accentuation de l’interrogation sur l’intérêt d’une annotation discrète ou continue.

Le nombre de dimensions à prendre en compte a tout d'abord son importance sur la fidélité de l'émotion retranscrite face à l'émotion vécue, mais également au niveau du système de reconnaissance automatique car elles vont permettre de spécifier les classes d'émotions à discriminer. Fontaine et al. [2007] déclarent que quatre dimensions¹⁷ sont nécessaires lorsque l'on souhaite représenter de manière satisfaisante, à savoir distinctes, les différences et les similitudes des termes émotionnels. Ils remettent en cause, en se positionnant comme partisans de la théorie de l'*appraisal* [Ellsworth and Scherer, 2003; Scherer, 2005b], la suprématie que connaît le modèle bidimensionnel *Valence-Activation*. A ce jour, le nombre optimal de dimensions nécessaires à une retranscription solide de l'émotion reste toutefois une question ouverte.

L'annotation statique, qui cherche à renseigner sur une émotion dans sa globalité ou à un instant t , élude quelque peu l'intérêt d'une annotation continue par rapport à une annotation discrète, et ce point est plus particulièrement mis en avant par l'annotation dynamique puisque l'on cherche à illustrer les variations qui ont lieu au cours d'une émotion. On constate cela dans l'émergence de nouvelles approches qui privilégient un modèle continu, en abandonnant la catégorisation en différentes classes pour les systèmes de reconnaissance des émotions [Gunes and Pantic, 2010; Nicolle et al., 2012; Soleymani et al., 2011; Wöllmer et al., 2008]. Ces approches demandent toutefois de relever certains défis à l'heure actuelle [Hudlicka and Gunes, 2012; Metallinou and Narayanan, 2013], comme la gestion du décalage temporel de l'annotation [Mariooryad and Busso, 2013].

2.4.2 Eliciter une émotion

Comment provoquer une émotion et de surcroît faire en sorte qu'elle soit semblable d'un sujet à un autre? C'est à ces questions que va tenter de répondre cette section. Pour ce faire, la discussion du principe de l'élicitation des émotions, notamment aux travers des mécanismes sous-jacents impliqués, va permettre d'éclaircir certains aspects théoriques. En ce qui concerne les aspects pratiques, ils seront mis en lumière par un état de l'art des méthodes employées dans le domaine de l'informatique affective.

Le principe de l'élicitation des émotions

Réussir à déclencher une émotion est une problématique majeure dans la réalisation du protocole expérimental. Bien que différentes tentatives de définition ait été données de la part de plusieurs disciplines, on peut s'accorder sur le fait que rien n'est plus subjectif que l'émotion ressentie, véhiculée par un stimulus. On peut d'ailleurs emprunter les termes de "signifiant" et de "signifié" au domaine de la linguistique pour aborder la question du stimulus dans le cadre de l'élicitation de l'émotion en s'autorisant l'interpolation suivante pour les définir :

17. evaluation- pleasantness, potency-control, activation-arousal, and unpredictability (plaisir, contrôle, activation, et imprévisibilité).

- le *signifié*, auquel on peut attribuer une plus grande part de subjectivité, à l'objet/ la représentation de ce stimulus,
- le *signifiant* qui a trait à l'esthétique du stimulus.

En linguistique ces termes représentent respectivement le concept (représentation mentale d'une chose), et l'image acoustique d'un mot. Le signifiant est affaire de goût, d'appréciation et peut influencer sur le positionnement de l'état émotionnel en terme de valence. Le signifié est la dimension du stimulus sur laquelle la part de subjectivité est la plus importante et comporte donc une grande part d'imprévisibilité pour l'expérimentateur.

Lorsque l'on cherche à déclencher une émotion dans un cadre expérimental, on est amené à accepter qu'une partie de l'expérience ne puisse pas être entièrement contrôlée, quelque soit les réflexions menées en amont sur les éléments déclencheurs. On ne peut qu'essayer de réduire cette part d'imprévisibilité sans toutefois pouvoir l'éliminer.

La grande thématique liée à ce sujet repose donc sur la place de la conscience [Lambie and Marcel, 2002; Lewis, 2011; Scherer, 2005a; Tsuchiya and Adolphs, 2007] et de l'inconscience [Berridge and Winkielman, 2003; Scherer, 2005a] dans le processus émotionnel. S'interroger sur leurs portées revient en effet à discuter du poids de la subjectivité. Scherer [2004] a par ailleurs dressé un inventaire des variations intervenant sur différentes composantes caractérisant le phénomène émotionnel en fonction des informations conscientes que nous pouvons percevoir d'un stimulus (voir la Figure 2.14).

Il a été avancé que le processus émotionnel n'est pas simplement lié à l'apparition d'un stimulus, comme cela est proposé par la théorie de James-Lange ou encore de Cannon-Bard, mais que plusieurs éléments interfèrent avec ce stimulus. Ces éléments se situent essentiellement dans des processus cognitifs, comme cela est affirmé notamment avec l'hypothèse des marqueurs somatiques de Damasio [1996] et les théories cognitivo-physiologique [Schachter and Singer, 1962] et de l'*appraisal* [Scherer et al., 2001]. Plusieurs études soulignent que le contexte dans lequel intervient le stimulus a également son importance [Pastor et al., 2008; Siemer et al., 2007; Stemmler et al., 2001].

Design feature										
	Intensity	Duration	Synchronization	Event/situation focus	Intrinsic appraisal	Transactional appraisal	Rapidity of change	Behavioral impact	Induction via art/music	
Type of affective state: brief definition (<i>examples</i>)										
Preferences: evaluative judgments of stimuli in the sense of liking or disliking, or preferring or not, over another stimulus (<i>like, dislike, positive, negative</i>)	L	M	VL	VH	VH	M	VL	M	H	
Utilitarian emotions: relatively brief episodes of synchronized response of all or most organismic subsystems to the evaluation of an external or internal event as being of major significance for personal goals and needs (<i>angry, sad, joyful, fearful, ashamed, proud, elated, desperate</i>)	H	L	VH	VH	M	VH	VH	VH	H	
Aesthetic emotions: evaluations of auditory or visual stimuli in terms of intrinsic qualities of form or relationship of elements (<i>moved, awed, surprised, full of wonder, admiration, bliss, ecstasy, fascination, harmony, rapture, solemnity</i>)	L-M	L	M	H	VH	L	VH	L	VH	

FIGURE 2.14 – Caractéristiques de trois types d'affects proposés par Scherer [2005a]. (VL : très faible, L : faible, M : Moyen, H : élevé, VH : très élevé) [Scherer, 2004].

Pour une lecture plus approfondie concernant la théorie de l'élicitation, et plus particulièrement ses mécanismes sous-jacents, nous conseillons les papiers de Koelsch [2010]; Philippot [1993].

Nom	Référence	Modalité	Nbr. de stimuli	Annotation
IAPS	Lang et al. [1997]	images	700	Val., Act., Dom.
IADS	Bradley and Lang [2007]	sons	167	Val., Act., Dom.
EMDB	Carvalho et al. [2012]	vidéos	52	Val., Act., Dom.
GAPED	Dan-Glauser and Scherer [2011]	images	730	Val., Act.
ANEW	Bradley and Lang [1999]	texte	700	Val., Act., Dom.
NAPS	Marchewka et al. [2014]	images	1356	Val., Act., App.-Av.
DEAP	Koelstra et al. [2012]	vidéos	120	Val., Act., Dom.

TABLEAU 2.8 – Les bases de données validées de stimuli à contenu émotionnel (Val. : Valence, Act. : Activation, Dom. : Dominance, App.-Av. : Approach-Avoidance).

Les modalités des stimuli provoquant une émotion

Différentes stratégies d'élicitation de l'émotion ont été mises en œuvre, et on peut les classer en deux catégories : les émotions induites par un stimuli extérieur et celles provoquées par l'individu lui-même. Cette dernière catégorie est réalisée soit en demandant au participant de se remémorer des souvenirs forts du point de vue émotionnel [Alzoubi et al., 2009; Chanel et al., 2007], mais il peut également lui être demandé de réaliser des expressions du visage qui semblent par nature provoquer chez lui une émotion [Laird, 1984; Laird and Strout, 2007].

Au sein des stimuli extérieurs on retrouve des déclencheurs pouvant solliciter différents sens, tels que l'odorat [Alaoui-Ismaïli et al., 1997; Chen and Dalton, 2005], le toucher [Singh et al., 2014], la vue et l'ouïe avec l'utilisation d'images [Savran et al., 2006], de sons [Daly et al., 2014; Du and Lee, 2015] ou encore de vidéos [Koelstra et al., 2012; Liu et al., 2015].

Des études se sont penchées sur le cas de stimuli multi-modaux [Mühl et al., 2011; Paulmann and Pell, 2011], par l'utilisation conjointe d'images et de sons [Gerdes et al., 2014], de musique [Baumgartner et al., 2006b; Bos, 2006], et plus particulièrement de musique classique [Baumgartner et al., 2006a]

Toutefois, dans le domaine de l'informatique affective, les stimuli audio-visuels sont privilégiés : utilisation de musique [Kim and André, 2008; Lin et al., 2008, 2010] ou de sons [Du and Lee, 2015; Kuncheva et al., 2011], d'images (essentiellement de la base IAPS [Lang et al., 1993]) [Horlings et al., 2008; Jatupaiboon et al., 2013a; Yuen et al., 2011], de vidéos [Gross and Levenson, 1995; Liu et al., 2015; Philippot, 1993; Rottenberg et al., 2007; Soleymani et al., 2012a; Wang et al., 2014b].

Les bases de stimuli les plus couramment utilisées en informatique affective sont données dans le Tableau 2.8.

Nous orientons le lecteur vers les papiers de Coan and Allen, Allen [2007]; Gerrards-Hesse et al. [1994] pour de plus amples informations concernant les méthodes d'élicitation de l'émotion dans le cadre d'expérience appliquées à l'informatique affective.

Discussion

Si de nombreuses variétés de stimuli ont été utilisées pour déclencher une émotion dans le cadre d'expériences et d'études dans le domaine des sciences cognitives, les stimuli audio et/ou visuels semblent être ceux qui se prêtent le mieux des applications dans le domaine de l'informatique affective.

Toutefois il reste une problématique moins étudiée, qui est celle de la durée de diffusion d'un stimulus nécessaire pour avoir une quasi-certitude d'avoir déclenché une émotion chez le participant. Il est cependant couramment admis dans le cadre de l'emploi de stimuli visuels et audio-visuels non-musicaux que cette durée est de l'ordre de quelques secondes [Ekman, 1984] et que l'émotion peut persister jusqu'à une dizaine de secondes après l'arrêt de l'exposition à ce stimuli [Levenson, 1988].

2.5 Les corpus existants

Différents corpus ont été créés dans le cadre de la reconnaissance des émotions basée sur les signaux EEG. Cependant, peu d'entre eux sont rendus accessibles. On peut citer, à l'état actuel, trois corpus majeurs : ENTERFACE'06 [Savran et al., 2006], DEAP [Koelstra et al., 2012] et MAHNOB-BCI [Soleymani et al., 2012a], dont les principales caractéristiques sont données dans le Tableau 2.9. L'ensemble de ces corpus proposent un ensemble multi-modal d'enregistrements avec une combinaison de signaux EEG et de signaux périphériques variés.

2.5.1 ENTERFACE'06 [Savran et al., 2006]

Le protocole expérimental de ce corpus se compose de trois sessions de 15 minutes. Chaque session (Figure 2.15) contient 30 blocs, 10 pour chaque catégorie d'état émotionnel. Un bloc correspond à une succession de 5 images appartenant toutes à un seul et même état émotionnel. L'ensemble du protocole se compose d'un total de 90 blocs, donc 450 images. Après chaque bloc de 5 images le sujet est invité à faire une auto-évaluation de son état émotionnel en utilisant la technique du *Self-Assessment Manikin* (SAM) en donnant une note comprise entre 1 et 9 pour les composantes de *Valence* et d'*Activation*. Les images composant chaque bloc sont prises à partir de données de référence : l'IAPS (*International Affective Picture System* [Lang et al., 1997]), qui est une base de données comprenant 1196 photos avec un contenu émotionnel varié. Pour le protocole ENTERFACE'06, trois sous-ensembles d'images sont sélectionnés dans IAPS, un pour chaque état émotionnel, à savoir calme, excité positivement et excité négativement (Figure 2.16), en fonction de seuils spécifiques le long des axes valence et l'excitation. Cela abouti à une sélection de respectivement 106, 71 et 150 photos pour les trois classes émotionnels considérées. La répartition des stimuli dans le plan Valence/Activation est fournie dans la Figure 2.17.

	eNTERFACE'06	DEAP	MANHNOBHCI
Equipement EEG	Biosemi Active Two	Biosemi Active Two	Biosemi Active Two
Nombre d'électrodes EEG	54	32	32
Fréquence d'échantillonnage EEG	256 Hz pour le Sujet 1, 1024 Hz sinon	512 Hz	512 Hz
ECG	-	-	X
Fréquence d'échantillonnage ECG	-	-	256 Hz
EMG	-	X	-
Fréquence d'échantillonnage EMG	-	512 Hz	-
Respiration	X	X	X
Fréquence d'échantillonnage Respiration	-	512 Hz	256 Hz
Température	-	X	X
Fréquence d'échantillonnage Température	-	512 Hz	256 Hz
EOG	-	X	-
Fréquence d'échantillonnage EOG	-	512 Hz	-
Conductance de la peau (GSR)	X	X	X
Fréquence d'échantillonnage GSR	-	512 Hz	256 Hz
Blood Volume Pressure (BVP)	X	X	-
Fréquence d'échantillonnage BVP	-	512 Hz	-
Autres modalités	fNIRS, Vidéo	Vidéo du visage (50 fps)	Audio (44.1 kHz), Vidéo (60 fps), suivi du regard (60 Hz)
Granularité de l'annotation	25 s		60 s
Stimuli	Images (IAPS)	Vidéos	Vidéos
Nombre de stimuli	450	40	20
Nombre de participants	5	32	27
Nombre de sessions/participant	3	-	-
Nombre de blocs/participants	30	-	-
Valence	X	X	X
Activation	X	X	X
Autres échelles d'annotation	-	Dominance, Familiarité, Affection	Dominance, Prévisibilité, clavier émotionnel
Valeurs limites sur Valence/Activation	1-5	1-9	1-9
Annotation Discrète ou Continue	Discrète	Continue	Discrète

TABLEAU 2.9 – Comparatif des bases de données existantes et accessibles

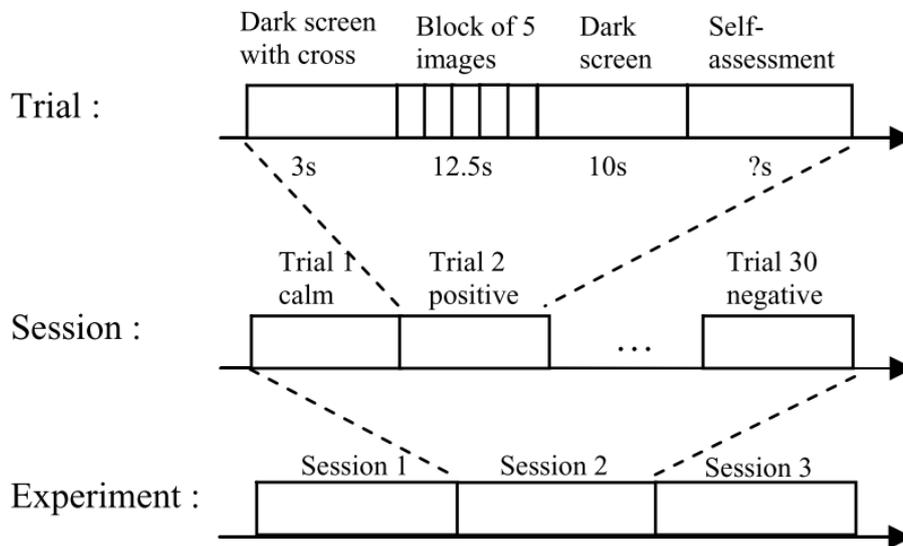


FIGURE 2.15 – Description du protocole ENTERFACE'06 [Savran et al., 2006].



FIGURE 2.16 – Images issues de la base de données IAPS (de gauche à droite : classe "excité positif", classe "excité négatif", classe "calme").

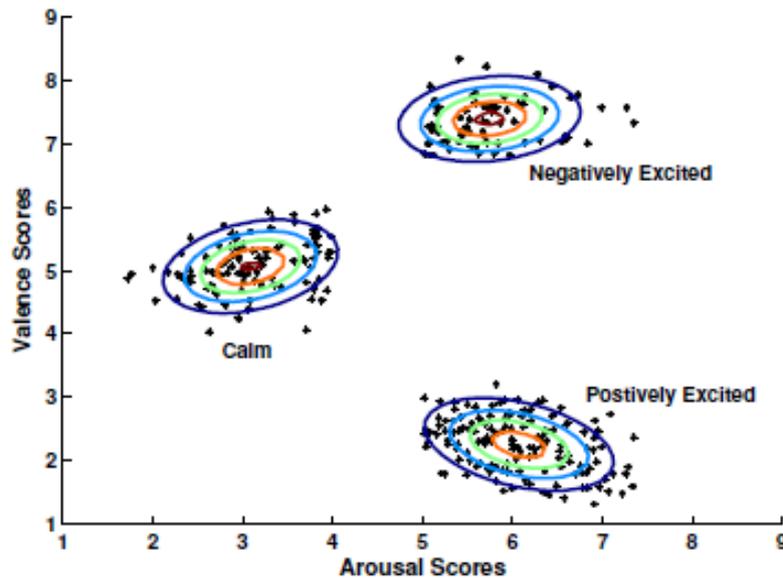


FIGURE 2.17 – Répartition des stimuli de la base ENTERFACE'06 dans le plan Valence/Activation [Xu and (Kostas) Plataniotis, 2012].

2.5.2 DEAP Dataset [Koelstra et al., 2012]

Le corpus DEAP s'est quant à lui orienté vers l'utilisation de stimuli musicaux et plus particulièrement de clips vidéos. Il est constitué de l'enregistrement de signaux physiologiques et EEG de 32 participants. Ce corpus a pour objectif de créer une base de données servant à la reconnaissance automatique des émotions véhiculées par des stimuli visuels et musicaux. Pour chaque session le participant visionne 40 extraits de clips musicaux d'une durée d'une minute. L'évaluation de l'émotion ressentie se fait sur une échelle continue allant de 1 à 9, sur les dimensions de valence, d'activation et de dominance. D'autres axes d'annotation sont utilisés, tels que l'appréciation de la vidéo et sa connaissance. Les stimuli sont répartis de telle sorte que soit présents 10 extraits dans chaque quadrant du plan valence/activation (Figure 2.18).

2.5.3 MAHNOB-HCI [Soleymani et al., 2012a]

Les expériences de cette base ont été menées dans le but d'acquérir des connaissances quant aux comportements naturels d'adultes en bonne santé lorsqu'ils sont confrontés à une interaction avec un ordinateur. Cette interaction est envisagée dans le cadre du visionnage de contenus multimédia, conçus pour provoquer des réactions affectives au contenu, comme l'amusement ou la répulsion, ainsi que son accord/désaccord avec le contenu fourni. Le comportement du participant est enregistré à l'aide de caméras, d'un microphone et d'un dispositif de suivi du regard. Plusieurs données physiologiques sont également enregistrées (Tableau 2.9). Les stimuli utilisés consistent en 20 vidéos d'une durée comprise entre 39,4 et 117 secondes qui rentrent dans l'une des catégories émotionnelles suivantes :

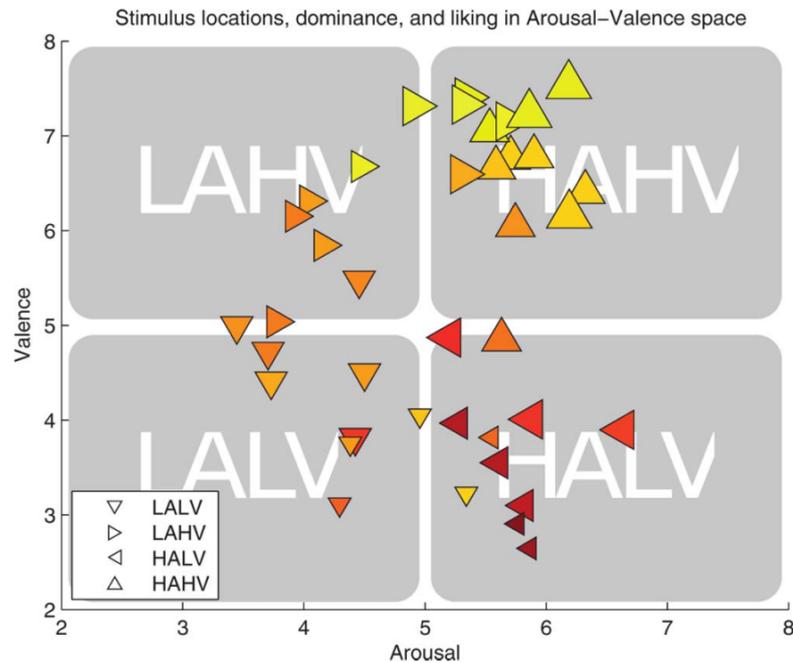


FIGURE 2.18 – Répartition des stimuli de la base DEAP dans le plan Valence/Activation pour les quatre conditions (LALV, HALV, LAHV et HAHV) [Koelstra et al., 2012]. La couleur traduit l'appréciation (le rouge foncé est une faible appréciation), et la taille traduit la dominance, de petite à élevée.

le dégoût, l'amusement, la joie, la peur, la tristesse et le neutre. Les données de 27 participants sont enregistrées. Le protocole se déroule de la manière suivante : le participant visionne une vidéo neutre sélectionnée au hasard dans une base fournie par la laboratoire de psychophysologie de Stanford [Rottenberg et al., 2007], puis visionne une des 20 vidéos à contenu émotionnel et effectue ensuite une auto-évaluation de l'émotion qu'il a ressenti en répondant aux cinq évaluations suivantes :

- le choix d'un label émotionnel parmi les labels suivants : neutre, anxiété, amusement, tristesse, joie, dégoût, colère, surprise et peur,
- l'activation,
- la valence,
- la dominance,
- et la prévisibilité.

Hormis le label émotionnel, les autres dimensions sont sur une échelle allant de 1 à 9. La diffusion d'une vidéo neutre avant chaque vidéo émotionnelle a pour objectif de réduire la probabilité d'un biais provoqué par un état émotionnel.

2.5.4 Discussion

Peu de corpus liés à la reconnaissance des émotions basée sur les signaux physiologiques dont l'EEG sont rendus accessibles. Une contrainte, directement liée à la complexité de la

mise en place du setup matériel d'une expérience utilisant un dispositif EEG, est le temps de l'expérience. Cela, rajouté à la charge cognitive importante demandée à un participant au cours d'une expérience cognitive, réduit le nombre de participants volontaires.

Les stratégies d'élicitation et d'annotation diffèrent d'un corpus à l'autre, de même que le choix effectué dans la sélection des stimuli dans le plan valence/activation. Concernant l'emploi de stimuli audio-visuels, seul un point de vue statique a été adopté pour l'annotation et il pourrait être pertinent de se pencher sur les variations de l'émotion ressentie qui ont eu lieu pendant le visionnage de ces stimuli.

2.6 Conclusion

Le phénomène émotionnel est complexe et l'informatique affective s'est dans un premier temps penchée sur l'étude de ses manifestations corporelles pour mettre en place des systèmes de reconnaissance des émotions.

La démocratisation des dispositifs d'enregistrement de signaux EEG a permis le développement de nombreuses études sur la reconnaissance des émotions basée sur l'activité cérébrale. Si toutefois ce moyen reste peu employé pour cette tâche dans le domaine des neurosciences, il fournit des informations utiles qui permettent son utilisation dans le domaine d'études cognitives.

Comme dans toute architecture de système de reconnaissance, l'utilisation de signaux EEG nécessite un traitement des données en amont du classifieur. La méthodologie à apporter est dessinée dans ses grandes lignes par l'étude des travaux existants.

Chapitre 3

EMOGEE : un nouveau corpus pour la reconnaissance des émotions basée sur l'EEG

Sommaire

3.1 Introduction	59
3.2 Stratégie d'élicitation : les stimuli	59
3.2.1 Dimensionnement des stimuli	60
3.2.2 Stimuli visuels statiques	60
3.2.3 Stimuli audio-visuels	61
Corpus SAFE	63
Vidéos liées aux phobies	64
3.3 L'installation expérimentale	65
3.3.1 Le système matériel	65
Capteurs physiologiques et EEG	66
Matériel Audio-Visuel et autres périphériques	67
Synchronisation	70
3.3.2 Le protocole expérimental	70
Protocole avec les stimuli visuels	71
Protocole avec les stimuli audio-visuels	72
Calibrage	72
3.4 Stratégie d'annotation	74
3.4.1 Annotation statique	74
3.4.2 Annotation dynamique	74
3.5 Les données enregistrées	76
3.6 Analyse statistique des annotations	77
La mesure de la fiabilité inter-annotateur	77
La mesure de la fiabilité intra-annotateur	78

3.6.1	Analyses des annotations sur les stimuli visuels	78
	Analyse des données de calibrage	79
	Analyse des blocs eNTERFACE'06	80
3.6.2	Analyse des annotations sur les stimuli audio-visuels	82
3.6.3	Bilan	84
3.7	Conclusion	84

Résumé

Nous proposons ici une approche novatrice qui repose sur l'aspect dynamique de l'émotion. Cette approche est élaborée à travers une stratégie d'élicitation de l'émotion qui repose sur deux types de stimuli : visuels et audio-visuels. Elle s'illustre également dans le choix de la stratégie d'annotation des stimuli audio-visuels qui sont évalués par le sujet sur trois segments. La recherche de la construction d'un corpus viable passe par l'enregistrement de données multimodales et un souci tout particulier à été porté à la bonne synchronisation de ces données.

L'analyse statistique des annotations du corpus EMOGEE met en évidence la variabilité des évaluations inter-annotateur qui peut impacter par la suite les choix à effectuer pour la mise en œuvre d'un système de reconnaissance des émotions.

3.1 Introduction

La reconnaissance des émotions en informatique affective passe par la possession de données de qualité. Parmi les quelques corpus accessibles, s'ils présentent chacun leurs qualités, il est à noter quelques inconvénients. Si certains sont inhérents au dispositif d'EEG assez lourd, provoquant un nombre limité de participants, d'autres résultent d'un parti pris par les expérimentateurs.

Cependant tous s'accordent sur une vision statique de l'évaluation de l'émotion. Or le phénomène émotionnel est rarement brutal et invariant. Il est alors intéressant de conserver une trace des variations émotionnelles subies en les intégrant dans la phase de retranscription. C'est un des aspects novateurs dans le corpus que nous proposons.

La mise en place d'un protocole expérimental pour un corpus relatif à des tâches cognitives est complexe, et ce d'autant plus lorsqu'il est constitué de données multimodales. Un tel protocole nécessite un travail en amont d'un point de vue pluri-disciplinaire sur des études précédentes afin de définir les stratégies appropriées tant sur le plan de l'élicitation de l'émotion que sur la manière dont cette dernière doit être retranscrite par le participant. Une fois ces stratégies définies, il convient également de mettre en œuvre les moyens nécessaires afin de gérer de manière fiable l'enregistrement des données multimodales.

Un des défis majeurs concernant les travaux de reconnaissance des émotions basée sur l'EEG est la disponibilité de données fiables. Différents corpus ont déjà été réalisés dans de précédents travaux, cependant peu d'entre eux ont été rendus accessibles. On peut citer parmi les corpus accessibles eNTERFACE'06 [Savran et al., 2006], et plus récemment DEAP [Koelstra et al., 2012] et MAHNOB-HCI [Soleymani et al., 2012a]. Ces corpus ont pour point commun d'être des corpus multimodaux contenant les enregistrements de signaux EEG et de différents signaux physiologiques annexes. Ces corpus possèdent cependant quelques faiblesses, parmi lesquelles :

- l'enregistrement d'un nombre limité de participants,
- un manque de variabilité dans les données,
- une absence de la prise en compte de la dynamique du phénomène émotionnel.

Afin de répondre au mieux à certaines faiblesses des corpus existants différentes stratégies ont été adoptées, notamment au niveau des stratégies de l'élicitation de l'émotion (Section 3.2), et de l'annotation de l'émotion ressentie (Section 3.4).

Nous proposons ici un corpus multi-modal destiné à l'analyse de l'état émotionnel, dont la configuration et le protocole expérimental seront commentés dans la Section 3.3, et dont les particularités sont notamment l'utilisation de stimuli visuels et audio-visuels, et la prise en compte de la dynamique dans l'émotion ressentie.

3.2 Stratégie d'élicitation : les stimuli

Dans notre corpus, deux types de stimuli sont utilisés : des stimuli visuels et des stimuli audio-visuels. Ces stimuli sont essentiellement extraits de deux bases existantes :

- l' *International Affective Picture System* (IAPS) [Lang et al., 2008] pour les stimuli visuels statiques,
- le corpus SAFE [Clavel et al., 2006] pour les stimuli vidéo, augmenté de vidéos liées à des phobies fréquentes.

Une sélection a dû être effectuée au sein de chacune de ces bases.

3.2.1 Dimensionnement des stimuli

Une des plus grandes contraintes expérimentales pour la création du corpus, est la gestion du niveau de fatigue du participant. Il faut donc veiller à ce que la demande de concentration demandée au participant ne soit pas trop importante mais il faut toutefois conserver un nombre de stimuli assez conséquent. En résumé, la sélection de l'ensemble des stimuli doit s'effectuer en respectant les conditions suivantes :

- la durée de diffusion d'un stimuli doit être suffisamment longue pour déclencher une émotion, à savoir au minimum une dizaine de secondes,
- cette durée est toutefois limitée afin de pouvoir diffuser un plus grand nombre de stimuli,
- le stimulus, de par sa durée et son contenu, doit être en mesure de faire apparaître une dynamique dans l'émotion ressentie.

La durée d'une session expérimentale, en dehors du temps requis pour l'installation du dispositif expérimental, a été fixée à une durée limite d'une heure et demie, période au delà de laquelle on estime que le sujet ne peut plus maintenir un niveau de concentration satisfaisant. Dans cette durée sont inclus les temps de repos et les temps alloués à la manipulation des logiciels de lancement des stimuli et d'enregistrement des signaux. Afin d'augmenter le nombre de stimuli visionnés pour un sujet, chaque participant passe deux sessions qui ont lieu à des jours différents, et dont les stimuli sont répartis de la manière suivante pour chaque type :

- environ 15 minutes pour les stimuli visuels,
- et environ 60 minutes pour les stimuli vidéos.

3.2.2 Stimuli visuels statiques

L'utilisation d'images comme stimuli peut représenter une contrainte lorsqu'on les utilise individuellement sur une durée de diffusion convenable : le biais introduit de manière consciente ou non par un souvenir du participant lors de l'observation de cette image pendant plus d'une dizaine de secondes. Le protocole utilisé dans le corpus eNTERFACE'06 [Savran et al., 2006] permet de réduire ce biais en considérant un stimulus comme un bloc de 5 images qui se succèdent. C'est la raison pour laquelle une partie du protocole de notre corpus reprend le protocole utilisé dans la base eNTERFACE'06. Les stimuli utilisés dans ce protocole sont des images de la base IAPS, regroupées en blocs de 5 images appartenant

au même état émotionnel. Les images défilent à un rythme de 2,5 secondes par image. On obtient alors une durée de diffusion des stimuli de 12,5 secondes par bloc.

Un des avantages de l'utilisation d'images provenant de la base IAPS est la solidité des annotations fournies. En effet, pour construire cette banque d'images il a été fait appel à une centaine d'annotateurs qui ont chacun annoté l'émotion que leur ont fait ressentir chaque image (voir Annexe A.2.2).

Une sélection est effectuée au sein des blocs du corpus eNTERFACE'06. Ils ont été sélectionnés de manière à présenter un équilibre dans la répartition des stimuli visuels présentés aussi bien sur la dimension de valence (valence positive et valence négative) et sur la dimension d'activation (faible activation et forte activation). Cet équilibre est recherché dans la perspective de permettre la construction d'un système de classification binaire sur l'un des deux axes au choix.

Parmi les 90 blocs constitués dans le protocole d'origine, sont conservés pour les deux sessions d'un participant (Tableau 3.1) :

- 22 blocs de la classe neutre,
- 14 blocs "excité positif",
- 14 blocs "excité négatif".

	Session 1	Session 2	Session 3
Calme	1, 7, 16, 21	3, 6, 14, 15, 17, 18, 21, 26	3, 7, 9, 10, 13, 16, 20, 22, 23, 26
Excité Positif	2, 8, 20, 23, 28	4, 10, 12, 16	4, 5, 18, 27, 28
Excité Négatif	4, 9, 22, 29	1, 19, 22, 27, 28	2, 6, 17, 19, 29

TABLEAU 3.1 – Références des blocs sélectionnés dans le protocole ENTERFACE'06 [Savran et al., 2006]

Cela aboutit à la sélection de 250 images, réparties de la manière suivante : 70 images appartenant à la classe excité négatif, 70 pour excité positif et 110 de la classe neutre.

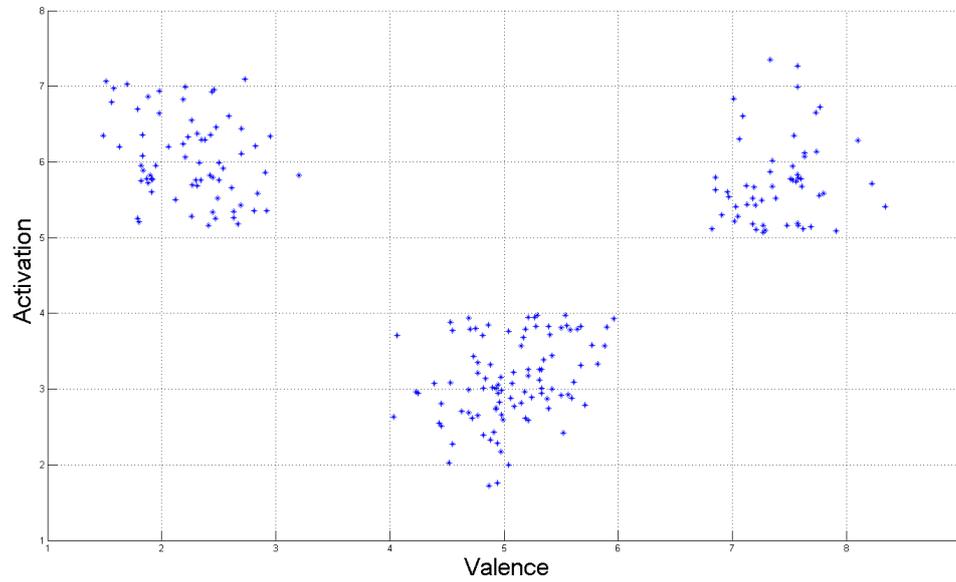
Les figures 3.1a et 3.1b illustrent la répartition des images sélectionnées dans l'espace Valence-Activation. Les détails des blocs se trouvent en annexe avec notamment les informations sur les images constituant chaque bloc, ainsi que les informations d'annotation sur les axes de valence et d'activation fournies par les annotateurs de la base IAPS.

3.2.3 Stimuli audio-visuels

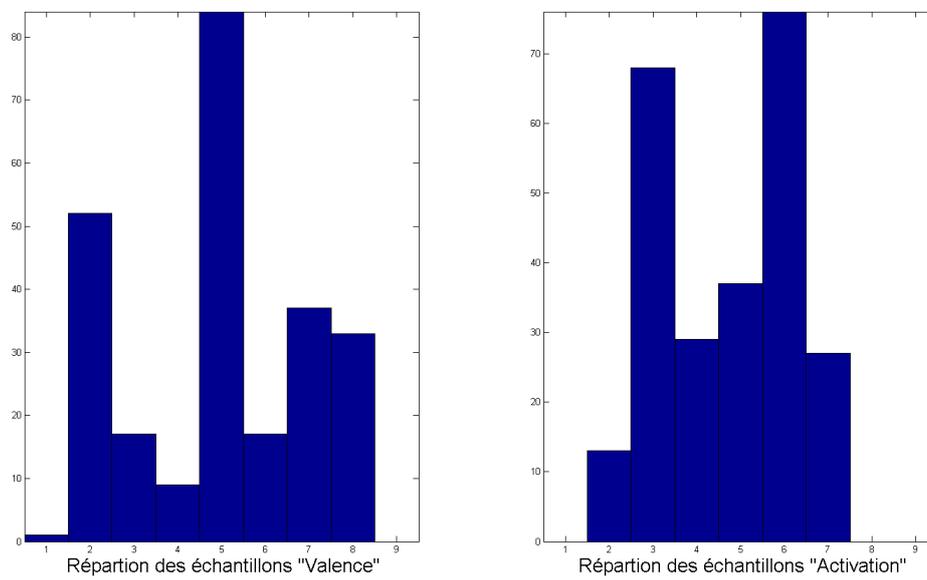
Le dernier point des contraintes du dimensionnement des stimuli (Section 3.2.1), implique une durée de diffusion des stimuli vidéos légèrement plus longues que pour les stimuli visuels, afin de s'assurer d'une dynamique de l'émotion, à savoir la présence de variations dans l'émotion ressentie. Au sein des stimuli audio-visuels, la répartition est la suivante pour une session :

- 42 vidéos issues du corpus SAFE (dont 6 neutres),
- 6 vidéos liées à des phobies (chiens, serpents, vertige) et 2 vidéos neutres.

Le volume sonore est normalisé pour l'ensemble des vidéos.



(a) Répartition des images sélectionnées dans le plan Valence/Activation.



(b) Répartition des échantillons suivants les dimensions de Valence et d'Activation.

FIGURE 3.1 – Répartition des stimuli visuels dans l'espace Valence/Activation.

Corpus SAFE

Le corpus est constitué de 401 extraits vidéos issus de 30 films. Chaque extrait est divisé en plusieurs segments de durées variables contenant soit une manifestation émotionnelle homogène ou alors un tour de parole. L'émotion ciblée est la peur, ce qui se conforme à notre objectif d'être en accord avec les scénarii proposés au sein du projet VERVE, qui pour rappel sont orientés vers l'utilisation d'émotions négatives.

Les annotations utilisées dans ce corpus fournissent des informations sur le *contexte situationnel* du segment [Clavel et al., 2011] (Figure 3.2), à savoir si une menace est présente associé aux indications suivantes :

- Le degré d'*imminence* de la menace : none(0), latente(1), potentielle/moyenne(2), immédiate(3), passée(4)
- et son *intensité* (ou gravité) : none(0), 1-3

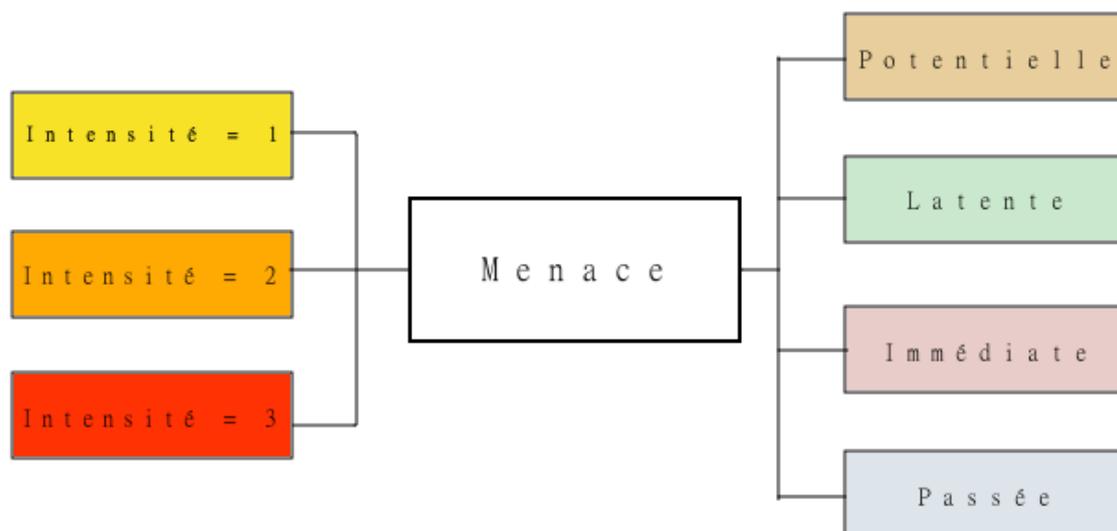


FIGURE 3.2 – Description du degré d'imminence et de la gravité dans l'annotation des segments du corpus SAFE.

On cherche à obtenir 84 vidéos d'une durée approximative de 15 secondes. Parmi ces 84 vidéos, 12 doivent avoir un contenu relativement neutre, les autres sont sélectionnées afin de présenter une dynamique de l'émotion. Les extraits d'une durée inférieure à 15 secondes sont éliminés car ils sont trop courts pour avoir la certitude qu'une émotion est élicitée chez le participant. Un aspect important recherché dans ces stimuli est une dynamique de l'émotion, c'est pourquoi les extraits présentant un écart-type nul sur les deux dimensions d'*Imminence* et d'*Intensité* sont écartés. Il ne reste donc que des extraits présentant une variation faible à avérée sur au moins l'un des deux axes. Les différents segments d'un même extrait sont regroupés en blocs d'au moins 15 secondes et dont la répartition des valeurs moyenne suivant les deux axes sont données dans la Figure 3.3. Une sélection est effectuée

au sein de ces regroupements après avoir éliminé les blocs d'écart-type nul sur *Imminence* et *Intensité*.

Les vidéos à contenu neutre sont sélectionnées parmi les extraits précédemment mis à l'écart.

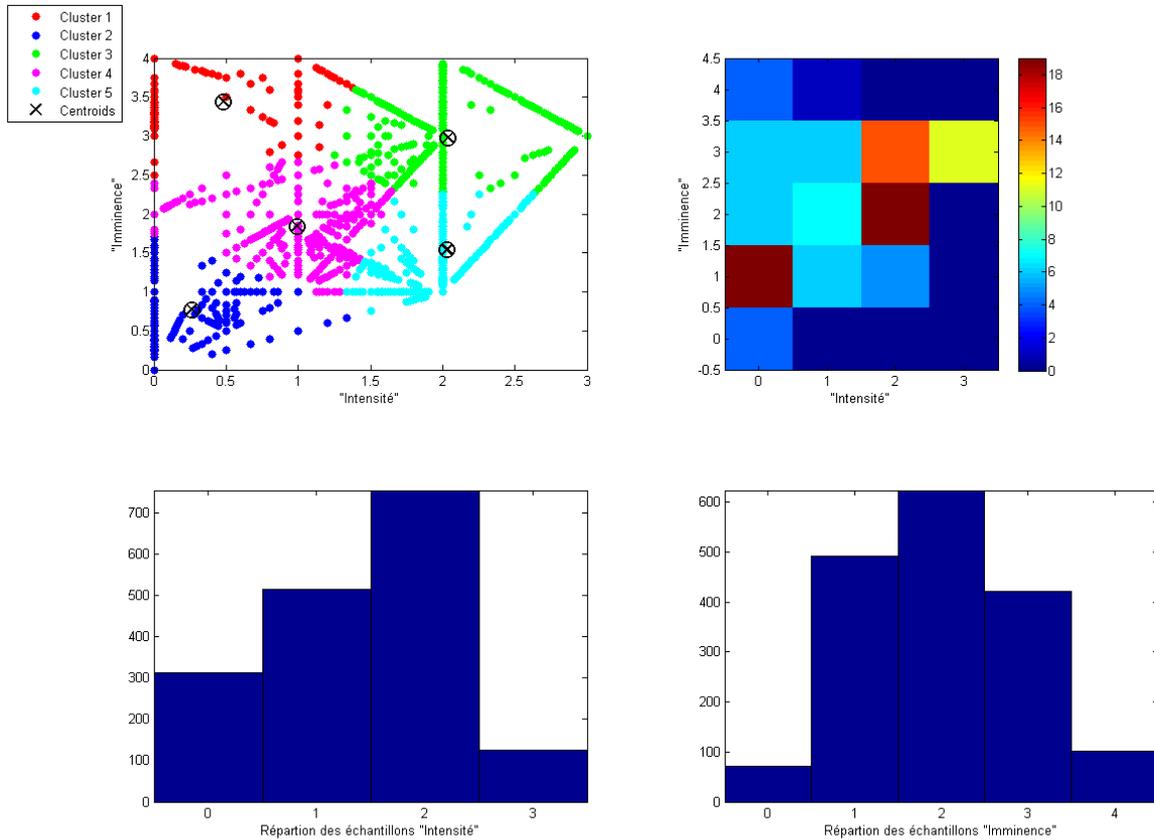


FIGURE 3.3 – Répartition de l'ensemble des regroupements sur les dimensions "Imminence" et "Intensité".

Vidéos liées aux phobies

A l'instar des vidéos extraites dans le corpus SAFE, les vidéos liées aux phobies sont construites afin de provoquer une dynamique émotionnelle chez le participant. Ces vidéos sont orientées vers les phobies suivantes : les chiens agressifs, les serpents et le vertige. Ces vidéos ont une durée de 15 secondes et consistent en :

- 4 vidéos contenant des serpents,
- 4 vidéos liées au vertige,
- 4 vidéos contenant des chiens agressifs,
- 4 vidéos neutres.

3.3 L'installation expérimentale

Le montage expérimental doit être réalisé dans une configuration aussi légère que possible afin de ne pas perturber le participant. La conception de ce montage expérimental a pris en compte la quantité de charge cognitive demandée au sujet mais également la mise en œuvre de bonnes conditions de concentration et d'immersion.

3.3.1 Le système matériel

Pendant l'expérience le sujet est assis face à un téléviseur HD 165cm, à une distance confortable, à savoir 2 mètres, qui correspond à une distance permettant à la fois d'avoir un sentiment d'immersion tout en diminuant les mouvements oculaires et de la tête de la part du participant. Il dispose sur un bureau devant lui d'une souris et d'un clavier, ce qui correspond au matériel minimum nécessaire au bon déroulement de l'expérience. L'expérimentateur se tient dans une pièce annexe (Figure 3.10). L'éclairage et le volume sonore sont contrôlés.

Le système matériel préparé comprend un ensemble de capteurs physiologiques périphériques et EEG, du matériel audio-visuel ainsi que d'autres éléments nécessaires à la synchronisation des signaux (Figure 3.4).

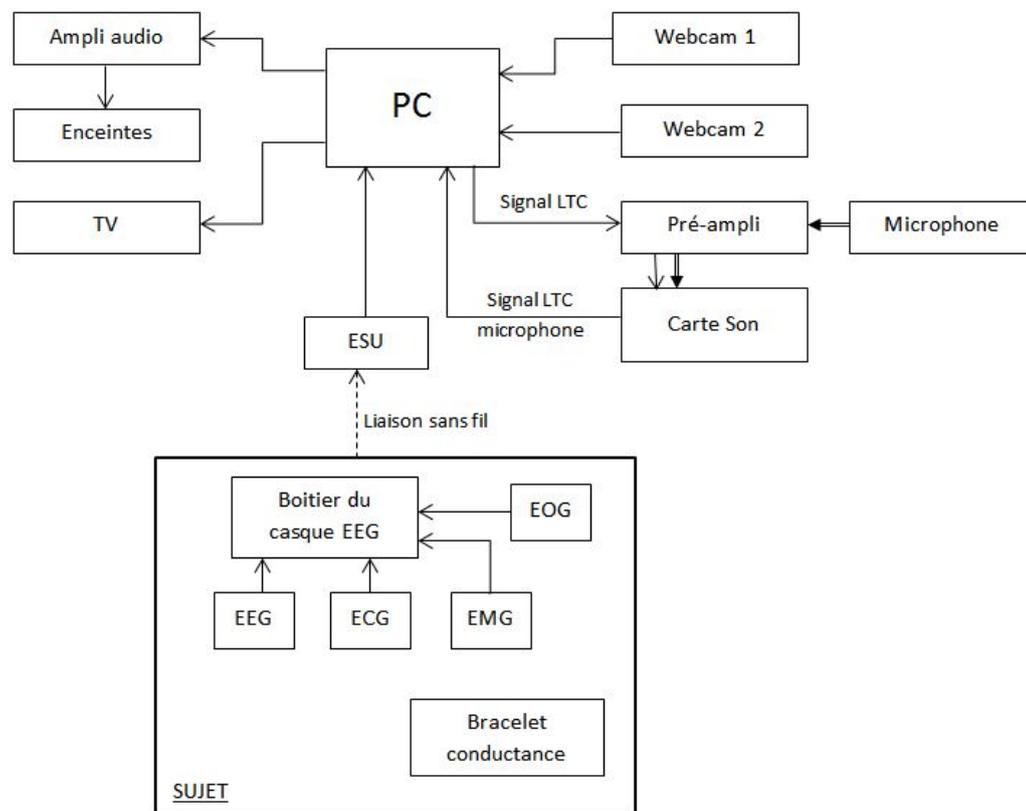


FIGURE 3.4 – Montage Expérimental pour l'enregistrement des données du corpus EMOGEE.

Capteurs physiologiques et EEG

Les signaux EEG sont enregistrés à l'aide d'un casque sans fil B-Alert X24¹ qui enregistre également en parallèle les signaux ECG, EOG et EMG. Ces signaux sont enregistrés à une fréquence d'échantillonnage de 256 Hz, et les 20 électrodes EEG sont disposées selon le système international 10-20 (Figure 3.5). En parallèle la conductance de la peau ainsi que la température sont enregistrées par un bracelet Affectiva² placé autour du poignet du participant.

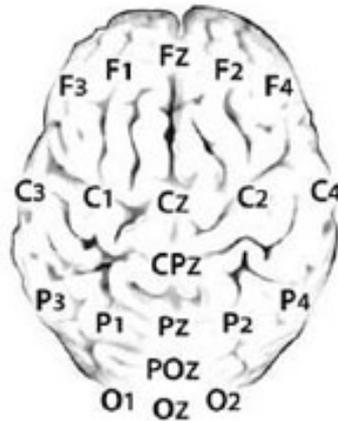


FIGURE 3.5 – Emplacements des 20 électrodes EEG du casque B-Alert X-24.

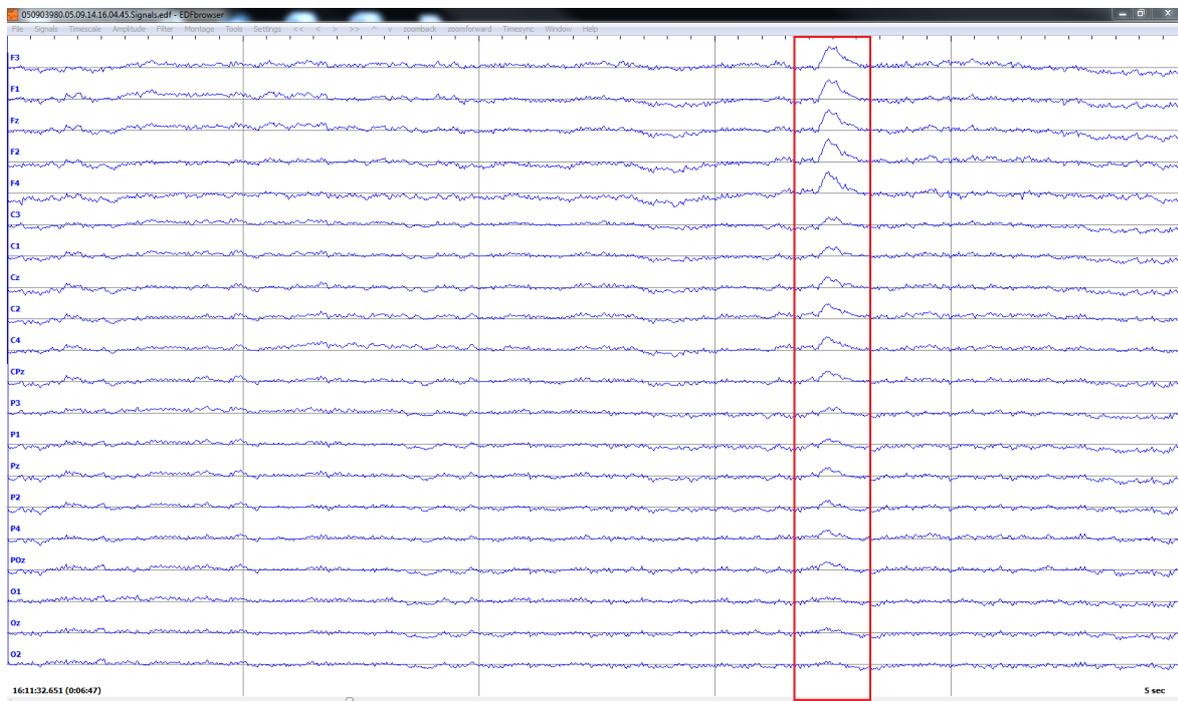


FIGURE 3.6 – Signaux EEG enregistrés par le système d'acquisition B-Alert X24 sur une durée de 5 secondes. Un artefact oculaire est présent dans cette portion de signaux (encadré rouge), de forte amplitude sur les électrodes frontales, il s'atténue à l'arrière du casque.

1. <http://www.advancedbrainmonitoring.com/xseries/x24/>
2. <http://qsensor-support.affectiva.com/>

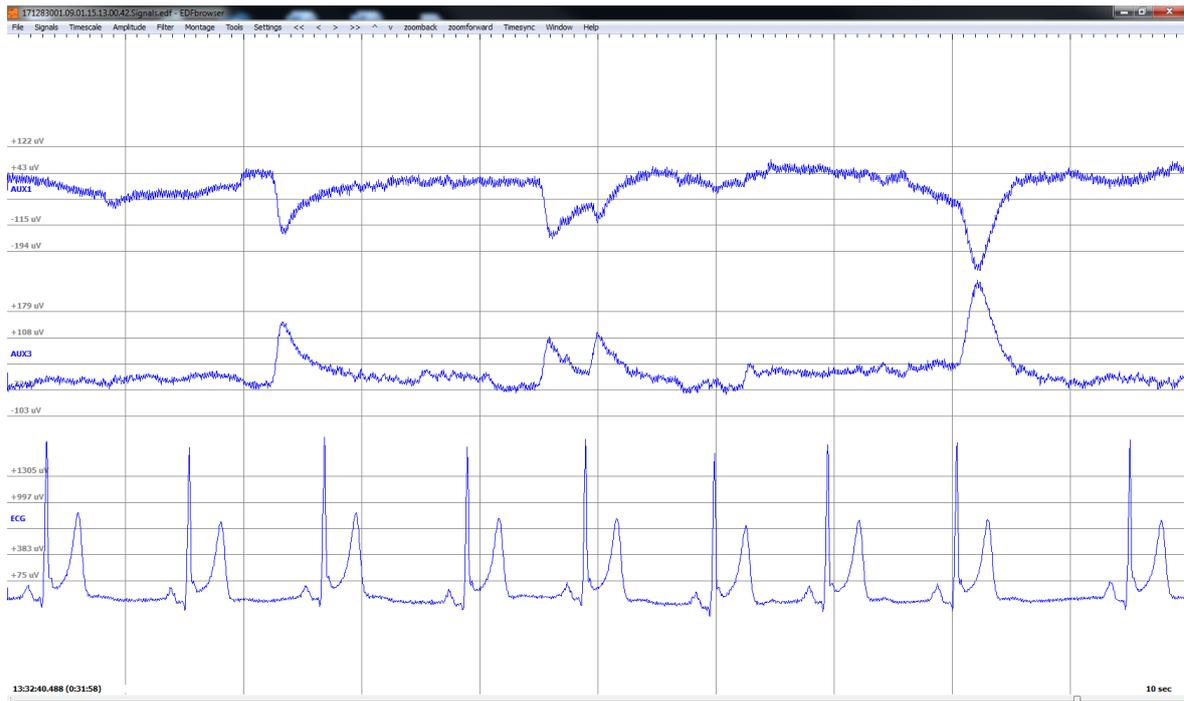


FIGURE 3.7 – Extraits de Signaux EOG et ECG.

L'utilisation d'électrodes EOG et EMG est faite pour répondre à deux objectifs :

- connaître l'activité oculaire et musculaire du participant dans le but de réaliser une réjection convenable des artefacts liés à ces activités,
- obtenir des informations pour une reconnaissance des émotions à partir de ces signaux physiologiques, méthode qui a fait ses preuves dans de précédentes études [Haag et al., 2004b; Jang et al., 2013; Kim and André, 2008].

En effet, l'activité des muscles du visage, bien que contraignant la qualité des signaux EEG, peut être révélatrice d'une émotion ressentie [Cheng and Liu, 2008]. Les électrodes sont placées de telle sorte qu'elles enregistrent l'activité de deux muscles essentiels du visage qui entrent en jeu dans l'expression des émotions : le grand zygomatique (dont l'activité peut être enregistrée à une position située à mi-distance entre le haut de l'oreille et le coin de la bouche), et l'orbiculaire palpébral inférieur qui ne peuvent être contractés simultanément que de manière involontaire [Darwin, 1872; Duchenne, 1862]. Activés de manière conjointe, ils sont révélateurs d'une émotion réelle et non actée [?]. La Figure 3.8 illustre l'emplacement des électrodes EOG et EMG.

L'ECG est également révélateur de l'émotion ressentie par le participant [Agrafiotti et al., 2012]. Il permet de plus à l'expérimentateur de pouvoir suivre la présence d'une émotion forte ressentie en visionnant les signaux ECG en temps réel au cours de l'expérience.

Matériel Audio-Visuel et autres périphériques

Les équipements audio-visuels utilisés pour l'expérience sont les suivants :

- une webcam HD placée discrètement en face du sujet (Logitech HD Pro Webcam C920),

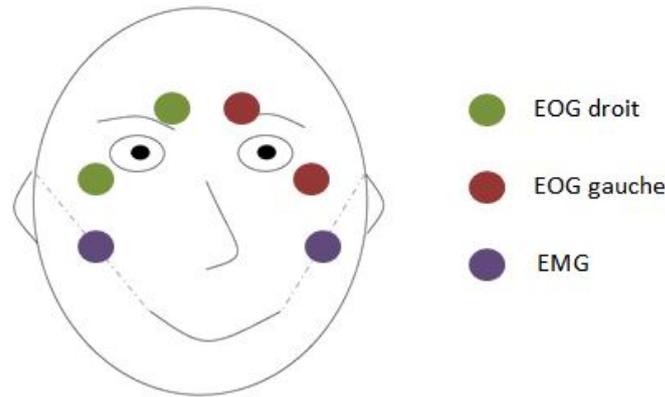


FIGURE 3.8 – Position des électrodes EOG et EMG. Quatre électrodes sont utilisées pour enregistrer l'EOG et deux pour l'EMG (Zygomatique majeur).

- un microphone qui est placé en direction du sujet et à proximité des enceintes,
- un boîtier de synchronisation ou *External Sync Unit (ESU)* qui récupère les données de l'ensemble des périphériques (physio+EEG) reliés au casque B-Alert X24 par Bluetooth,
- une TV Full HD (165cm) pour diffuser les stimuli,
- des enceintes, avec un volume réglé à l'identique pour tous les participants,
- une autre webcam HD de contrôle pour permettre à l'expérimentateur, qui se trouve dans une autre pièce, de vérifier le niveau d'attention du participant.

Un PC fonctionnant sous Windows 7 64-bits est également utilisé. Il enregistre l'ensemble des signaux et gère le déroulement du protocole.



FIGURE 3.9 – Participant lors d'une session de l'expérience.

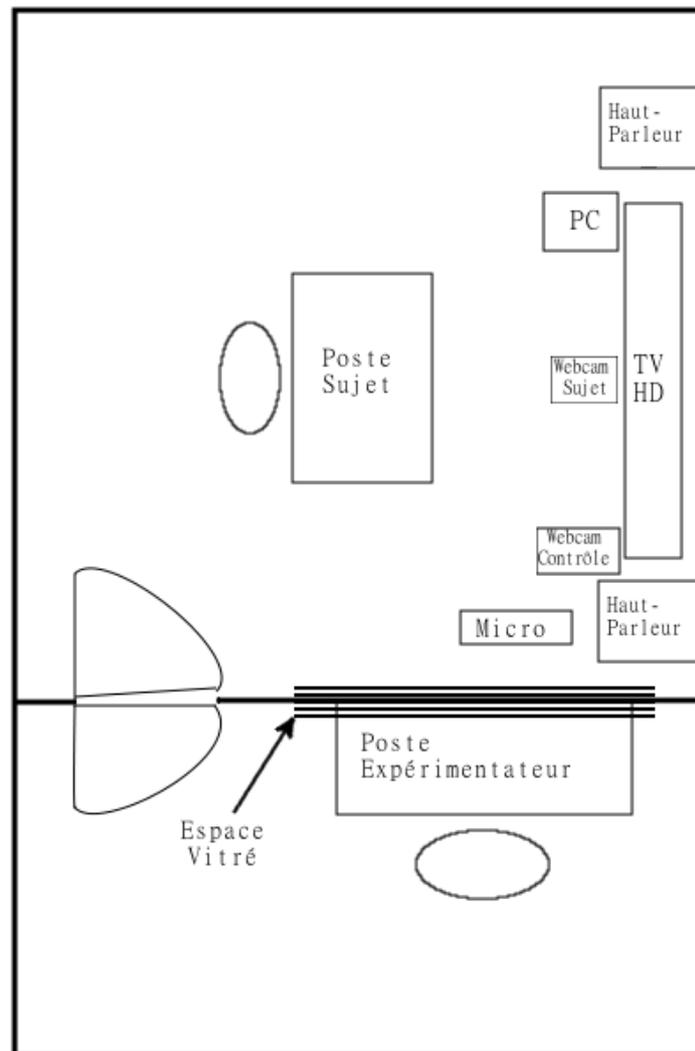


FIGURE 3.10 – Plan de la configuration de la salle lors de l'expérience.

Synchronisation

L'ensemble des périphériques est synchronisé grâce à deux mécanismes. Concernant le casque EEG et les périphériques pour l'enregistrement de signaux physiologiques, l'horodatage de l'ordinateur est enregistré en parallèle des données :

- Pour le bracelet Affectiva : le bracelet est synchronisé avec l'horodatage système de la machine avant le début de l'expérience. L'horodatage est ensuite enregistré en parallèle des données par le bracelet.
- Pour le casque EEG B-Alert X24 : le casque envoie les données par Bluetooth au boîtier ESU, qui possède son propre horodatage correspondant au temps écoulé depuis son branchement à la machine. Il enregistre l'horodatage système en parallèle des données qu'il reçoit.

Le flux vidéo est quant à lui enregistré via VLC³. Pour obtenir une information sur l'horodatage provenant du système au cours de la vidéo un Linear TimeCode⁴ (LTC) est enregistré sur un des canaux audio. L'initialisation de ce LTC est effectuée en début d'expérience en enregistrant l'horodatage système correspondant à son lancement. Lors de l'enregistrement vidéo, les canaux audio capturent alors les informations suivantes :

- pour le canal gauche le signal LTC, qui fonctionne en arrière-plan et qui code le temps écoulé depuis le début de l'expérience,
- pour le canal droit un microphone placé à proximité du sujet.

3.3.2 Le protocole expérimental

Le protocole expérimental du corpus proposé est dans un premier temps basé sur le pro-

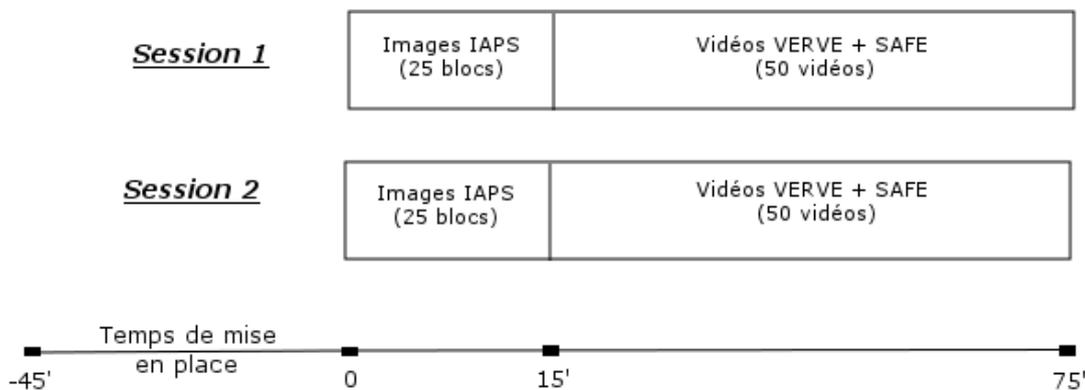


FIGURE 3.11 – Protocole Expérimental du corpus EMOGEE.

3. <http://www.videolan.org/vlc/>

4. http://en.wikipedia.org/wiki/Linear_timecode

tole de la base de données ENTERFACE'06, qui utilise comme stimuli des images issues de l'*International Affective Picture System* (IAPS), afin d'augmenter la taille de la base de données existante et de pouvoir mieux évaluer la validité statistique des résultats de classification. La deuxième partie du protocole repose sur des stimuli audio-visuels principalement orientés sur des émotions négatives telles que la peur, le stress ou l'anxiété. Les deux protocoles sont répartis de manière équivalente pour les deux sessions d'un sujet.

Chaque participant doit passer deux sessions d'enregistrement (Figure 3.11). Lors de la première session il lui est soumis 50 vidéos et 25 blocs du protocole ENTERFACE'06. La deuxième session est identique, seuls les stimuli présentés au participant sont modifiés. Les stimuli sont présentés à chaque participant dans un ordre aléatoire.

Quel que soit le stimulus utilisé, le protocole est construit de la manière suivante (Figure 3.12) :

- une croix noire sur écran blanc (3 secondes),
- diffusion des stimuli en plein écran,
- un écran blanc (10 secondes),
- suivi de l'auto-évaluation de l'émotion ressentie.

Ecran blanc avec une croix noire 3s	Stimuli	Ecran blanc 10s	Auto-évaluation
--	---------	--------------------	-----------------

FIGURE 3.12 – Description d'un essai correspondant à la diffusion d'un stimulus.

L'apparition de la croix noire sur l'écran pendant trois secondes permet de prévenir le participant de la diffusion proche du stimuli. Il peut ainsi se mettre en alerte pour rester concentré sur ce qui va suivre. Cette étape présente également l'avantage d'être une composante de neutralité avant la présentation du contenu émotionnel. A la suite du stimulus, et avant la phase d'auto-évaluation, un écran blanc apparaît pour une durée de dix secondes afin de conserver un maximum de trace émotive avant d'attaquer la réflexion de l'annotation.

Protocole avec les stimuli visuels

Trois états émotionnels sont ici considérés : calme, excité positif et excité négatif. Le protocole consiste en une succession de 25 blocs issus du protocole ENTERFACE'06. Les blocs sont répartis de la manière suivante :

- 11 blocs de la classe "calme",
- 7 blocs "excité positif",
- 7 blocs "excité négatif".

Un bloc correspond à une succession de 5 images appartenant à un seul état émotionnel. L'ensemble du protocole contient 25 blocs, à savoir 125 images. Chaque image est affichée pendant 2,5 secondes ce qui conduit à une durée totale de 12,5 secondes pour un bloc.

Un écran noir précède la phase d'annotation pendant 10 secondes. Puis il est demandé au participant de faire une auto-évaluation de l'émotion qu'il a ressenti en utilisant le Self-Assessment Manikin (SAM) avec un score entre 1 et 9 sur les composantes de valence et d'activation (Figure 3.13). Un écran blanc avec une croix noire au milieu précède l'arrivée de chaque bloc afin que le participant se prépare à porter son attention sur l'écran.

Ecran blanc avec une croix noire	Stimuli					Ecran blanc	Auto-évaluation
							
3 secondes	2,5 secondes	2,5 secondes	2,5 secondes	2,5 secondes	2,5 secondes	10 secondes	

FIGURE 3.13 – Description d'un essai dans le protocole ENTERFACE'06. la diffusion des stimuli est suivie d'une phase d'auto-évaluation *statique* (voir Section 3.4.1).

Protocole avec les stimuli audio-visuels

Les stimuli utilisés dans cette partie du protocole sont orientés sur les émotions négatives. Une session contient 50 vidéos, dont huit possèdent un contenu neutre, qui sont diffusées dans un ordre aléatoire. Elles ont chacune une durée comprise entre 15 et 20 secondes. A l'instar du protocole précédent, la diffusion d'une vidéo est précédée d'un écran noir avec une croix blanche pendant 3 secondes, et elle est suivie d'un écran noir pendant 10 secondes avant la phase d'annotation (Figure 3.14).

La phase d'annotation étant différente, un entrainement est réalisé en présence de l'expérimentateur avant le lancement de l'enregistrement de cette partie du protocole. Lorsque le participant se sent prêt, l'expérimentateur tamise la lumière pour renforcer le sentiment d'immersion, et quitte la pièce en laissant la porte très légèrement entrouverte afin d'isoler le participant.

Ecran blanc avec une croix noire	Stimuli Vidéo en plein écran	Ecran blanc	Auto-évaluation
			
3 secondes	Environ 15 secondes	10 secondes	

FIGURE 3.14 – Description du protocole pour une vidéo en plein écran. La diffusion du stimuli est suivie d'une phase d'auto-évaluation *dynamique* (voir Section 3.4.2)

Calibrage

Une phase de calibrage précède les deux parties du protocole. Elle a pour but d'aider le participant à se familiariser avec les composantes de valence et d'activation ainsi que de lui permettre d'évaluer ses limites sur chaque échelle. Neuf images issues de la base IAPS sont sélectionnées (Table 3.2) de la manière suivante : les valeurs moyennes des images doivent couvrir toute la dimension de l'espace valence/activation, puis les images possédant la plus faible valeur d'écart-type sont préférées, car elles illustrent un consensus dans l'annotation.

Les images sont disposées dans un ordre aléatoire pour chaque participant et numérotées de 1 à 9. Il lui est demandé d'annoter chaque image les unes par rapport aux autres sur la dimension de valence puis d'activation, selon son émotion ressentie (Figure 3.15). L'annotation est réalisée sur une échelle continue allant de 1 à 9 par un système de "glisser-déposer".

Afin de conserver une cohérence dans l'évaluation de l'émotion ressentie, ces points de repères apparaissent dans la suite du protocole sur l'ensemble des axes d'annotation, à l'exception des images placées aux extrémités des axes (voir Figure 3.16 et Figure 3.17). Ces images sont en effet écartées afin de ne pas perturber le participant dans l'annotation de son émotion ressentie.

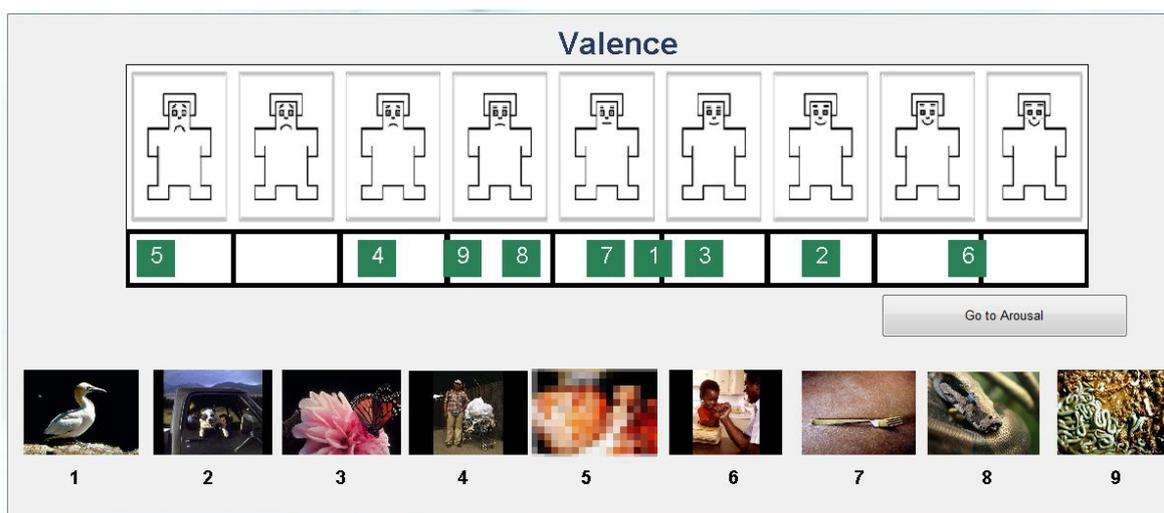


FIGURE 3.15 – Exemple d'annotation des 9 images de calibrage (1-9) sur la dimension de valence.

Référence IAPS	Description	Valence	Activation
1033	Serpent	3.87 (1.94)	6.13 (2.15)
1111	Serpents	3.25 (1.64)	5.20 (2.25)
1303	Chien	4.68 (2.11)	5.70 (2.04)
1450	Oiseau Marin	6.37 (1.62)	2.83 (1.87)
1604	Papillon	7.11 (1.41)	3.30 (2.17)
2154	Famille	8.03 (1.13)	4.48 (2.31)
3053	Victime de brûlures	1.31 (0.97)	6.91 (2.57)
7080	Fourchette	5.27 (1.09)	2.32 (1.84)
9331	Un sans-abri	2.87 (1.28)	3.85 (2.00)

TABEAU 3.2 – Moyenne (Écart-type) des images IAPS sélectionnées pour le protocole de calibrage

3.4 Stratégie d'annotation

Lors de la phase d'annotation, il est demandé au participant d'évaluer l'émotion qu'il a ressentie en visionnant les stimuli, et non l'émotion qu'il y perçoit.

Deux types d'annotations sont utilisées dans ce corpus :

- l'annotation statique ou globale,
- l'annotation dynamique.

Alors que seule l'annotation globale est utilisée pour les stimuli visuels (protocole ENTERFACE'06), les deux variantes d'annotation de l'émotion ressentie sont utilisées pour les stimuli audio-visuels. Toutefois, ces deux annotations reposent sur la technique du Self-Assessment Manikin (SAM) [Bradley and Lang, 1994], dont le principe repose sur le circumplex de Russell [Russell, 1980]. La représentation *Valence-Activation* [Bradley and Lang, 1994; Mehrabian and Russell, 1974] a été adoptée par les chercheurs se spécialisant sur une analyse basée sur l'EEG. Pour chacun des deux axes d'annotation, l'échelle utilisée est une échelle discrète allant de 1 à 9.

Pour l'ensemble des stimuli, l'annotation est effectuée suivant les composantes de valence et d'activation. Dans le cadre de l'annotation des stimuli vidéos, la plupart étant issus de films populaires, il est également demandé au participant d'indiquer s'il a déjà vu le film. Cette information nous permet de savoir s'il a eu une vision différente de l'extrait diffusé de par la connaissance de son contexte.

3.4.1 Annotation statique

L'annotation statique a pour objectif de permettre au participant d'évaluer l'émotion ressentie de manière globale. La technique de base du Self-Assessment Manikin est ici améliorée en ce sens que des images "repère" sont ajoutées aux illustrations d'origine (Figure 3.16). Ces images, identiques pour tous les participants, ont cependant une position unique sur chaque axe, c'est-à-dire qu'elles sont positionnées en haut de l'axe suivant l'évaluation qu'a faite le participant pendant la phase de calibrage. Afin de ne pas biaiser l'émotion ressentie par le sujet, les images qu'il a positionnées aux extrémités des axes pendant la phase de calibrage ne sont volontairement pas affichées sur les axes d'annotation. L'annotation statique est utilisée pour annoter de manière générale l'émotion ressentie après le passage d'un bloc d'images dans le protocole ENTERFACE'06. Cette annotation est également utilisée dans le protocole vidéo, pour l'annotation de l'émotion ressentie globalement pour chaque vidéo.

3.4.2 Annotation dynamique

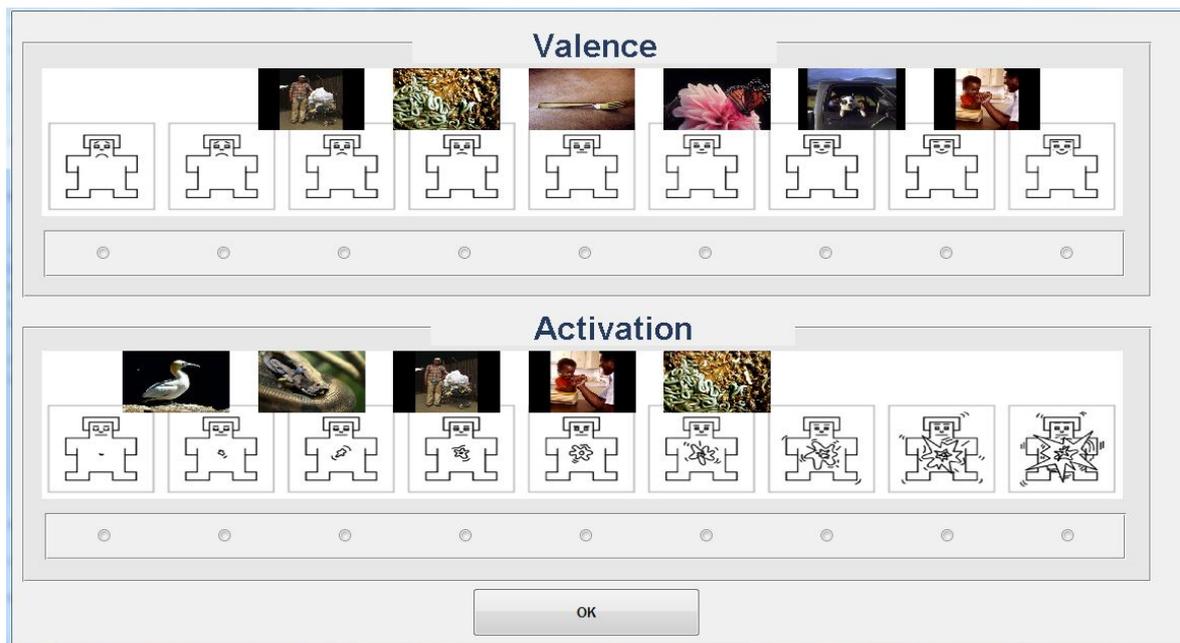


FIGURE 3.16 – Interface d'annotation globale

L'approche envisagée dans un premier temps pour annoter de manière dynamique l'émotion ressentie s'est orientée vers l'utilisation de la technique d'annotation par le logiciel GTrace⁵. Plusieurs obstacles sont apparus à l'intégration de cette stratégie dans notre protocole expérimental. Parmi ces obstacles, deux raisons qui influencent directement la qualité des données enregistrées peuvent être invoquées.

La première raison est liée au niveau de fatigue, due au temps et à la charge cognitive qu'implique l'utilisation d'une telle technique. En effet, le sujet doit faire l'annotation dynamique de chaque axe de manière synchrone avec la vidéo. On se retrouve donc avec un minimum de trente secondes de temps d'expérience supplémentaire par vidéo étant donné qu'il doit revoir la vidéo dans son intégralité pour l'annotation de chaque axe. Il est également demandé au participant d'annoter uniquement l'émotion qu'il a ressentie lorsqu'il a vu la vidéo pour la première fois. L'autre raison est que l'on craint donc ici un effet d'accoutumance au stimulus qui influencerait de manière volontaire/involontaire ou consciente/inconsciente l'annotation donnée par le sujet après trois visionnages de la vidéo.

Nous avons donc tenté de remédier à ces faiblesses en proposant une nouvelle stratégie d'annotation dynamique de l'émotion, moins coûteuse aussi bien en temps qu'en terme de charge cognitive, mais également soucieuse d'éviter tout problème d'accoutumance.

A l'instar de l'annotation statique effectuée dans le protocole des stimuli visuels, l'annotation dynamique s'effectue sur une seule fenêtre contenant les deux dimensions de valence et d'activation (Figure 3.17). Cette configuration permet tout d'abord d'avoir une réduction de la charge cognitive car le sujet peut annoter ces axes de manière quasi-simultanée. Cette configuration permet de diminuer le temps passé par le participant sur la phase d'an-

5. <https://sites.google.com/site/rodycowie/work-resources>

notation. Le participant, qui n'a plus la possibilité de visionner à nouveau la vidéo, doit faire appel à la mémoire de son ressenti, la réponse à ce stimulus est unique à son souvenir.

Le principe de l'annotation dynamique envisage un nouveau point de vue. L'annotation n'est plus synchrone sur la durée de la vidéo, mais elle est pensée comme la retranscription des variations majeures de l'émotion ressentie au cours de la vidéo. L'annotation se fait en trois segments qui peuvent être de durées différentes, ces durées sont laissées à l'appréciation du participant, pour lesquels il donne un score pouvant aller de 1 à 9 sur les dimensions de valence et d'activation. S'il n'a pas ressenti de variation dans son émotion, il a également la possibilité d'annoter ces trois segments de manière identique.

The screenshot shows a software interface for dynamic annotation. It is divided into two main columns: 'Valence' on the left and 'Activation' on the right. Each column has a header with a row of small video thumbnails. Below the thumbnails are two rows of nine icons each, representing a scale from 1 to 9. The 'Valence' scale is labeled 'extrêmement désagréable' on the left and 'extrêmement agréable' on the right, with 'neutre' in the middle. The 'Activation' scale is labeled 'calme' on the left and 'extrêmement activé' on the right. Below these scales are three rows of radio buttons, labeled 'Part 1', 'Part 2', and 'Part 3' on the left, and a 'Global' row at the bottom. At the bottom of the interface, there is a checkbox labeled 'Familiarité' with the text 'J'ai déjà vu le film' next to it, and an 'OK' button.

FIGURE 3.17 – Interface d'annotation dynamique

3.5 Les données enregistrées

En définitive, les données obtenues en fin d'expérience pour chaque sujet sont les suivantes :

- un fichier pour chaque tâche qui contient les enregistrements des signaux EEG et physiologiques annexes (ECG, EMG, EOG),
- un fichier de l'activité électrodermale (EDA) pour l'ensemble de la session qui contient les enregistrements de la conductance de la peau et de la température enregistrés par le bracelet Affectiva,
- un fichier vidéo pour le protocole IAPS et un autre pour le protocole vidéo, avec un canal audio pour le signal LTC et l'autre canal pour le microphone,
- un fichier *Sujet* contenant les informations utiles à la synchronisation avec les indications d'horodatage, et les informations relatives aux stimuli (l'ordre de passage et les

valeurs obtenues pour les annotations statiques et dynamiques sur les dimensions de valence et d'activation).

3.6 Analyse statistique des annotations

Chaque participant évalue l'émotion qu'il a ressentie selon son propre référentiel sur deux dimensions : la valence et l'activation. Bien qu'une réflexion ait été menée en amont sur la sélection de stimuli pertinents dans leur efficacité et dans leur contenu se répartissant dans le plan d'annotation, on peut toutefois s'interroger sur plusieurs points.

Dans un premier temps, il faut se pencher sur le poids de la subjectivité à la fois dans la manière d'appréhender les axes pour retranscrire l'émotion mais également observer les différences dans l'émotion ressentie par les participants. Le premier point va être examiné à travers les données de l'étape de calibrage, c'est-à-dire par une étude de la corrélation des évaluations fournies pour les images de référence. Il va s'en suivre une analyse statistique des évaluations pour les blocs d'eNTERFACE'06 qui vont être comparées aux évaluations fournies par ce corpus d'origine.

Enfin une partie des stimuli ne comportent pas de données terrain puisqu'ils ont été créés uniquement dans le cadre de ce corpus ou utiliser pour la première fois à des fins d'élicitation d'une émotion. Des analyses statistiques sont donc menées à partir des annotations fournies par les participants afin d'évaluer tout d'abord la capacité des vidéos à véhiculer une émotion pour enfin se pencher sur la pertinence d'une étude de la dynamique des émotions.

Plusieurs méthodes d'évaluations existent pour évaluer l'accord inter-annotateurs sur des échelles discrètes. L'annotation des stimuli s'effectue sur deux dimensions valence et activation sur une échelle allant de 1 à 9. Trois mesures sont considérées ici pour mesurer la fiabilité de l'annotation et du contenu.

La mesure de la fiabilité inter-annotateur

Le coefficient kappa [Cohen, 1960], noté ici κ , est défini comme le rapport suivant :

$$\kappa = \frac{\bar{p}_o - \bar{p}_e}{1 - \bar{p}_o}, \quad (3.1)$$

avec \bar{p}_o représentant la proportion d'accord observée entre les annotateurs et \bar{p}_e la proportion d'accord aléatoire, c'est-à-dire la probabilité que l'accord entre les annotateurs soit un hasard. Le coefficient κ est compris entre -1 et 1, qui correspond à un accord maximal. Si κ est égal à zéro, on considère que l'accord est obtenu par chance.

Une grille de lecture pour l'interprétation de ce coefficient est fournie par Landis and Koch [1977] et donnée dans le Tableau 3.3. Cette grille est arbitraire et doit être relativisée par le nombre de niveaux d'annotation disponibles.

Accord	Kappa
Excellent	$\geq 0,81$
Bon	0,80-0,61
Modéré	0,60-0,41
Médiocre	0,40-0,21
Mauvais	0,20-0
Très mauvais	< 0

TABLEAU 3.3 – Degré d'accord en fonction des valeurs de Kappa.

Nous faisons le choix d'une pondération quadratique pour le calcul du coefficient κ [Cohen, 1968] de Cohen car les échelles qui sont utilisées dans notre cas sont totalement ordonnées avec une graduation assez fine, et il convient de donner plus d'importance aux désaccords graves qu'aux désaccords légers qui correspondent à une différence d'un seul niveau.

Le coefficient κ de Cohen fonctionne pour mesurer l'accord entre deux annotateurs. Pour mesurer l'accord de plus de deux annotateurs on peut utiliser le coefficient de Fleiss κ_F qui repose sur la même interprétation. En partant de l'équation 3.1, et avec N_{ann} étant le nombre d'annotateurs et N_{st} le nombre de stimuli annotés, on calcule \bar{p}_o et \bar{p}_e de la manière suivante :

$$\bar{p}_o = \frac{1}{N_{ann}} \sum_{i=1}^{N_{st}} p_{st_i}, \quad (3.2)$$

où p_{st_i} représente la proportion d'accord d'annotateurs pour le i -ème stimuli par rapport au nombre total d'observateurs possibles, et

$$\bar{p}_e = \sum_{j=1}^k p_j^2, \quad (3.3)$$

où p_j correspond à la proportion des attributions au niveau j .

La mesure de la fiabilité intra-annotateur

Afin de mesurer l'accord des annotations au sein d'un même annotateur, on calcule le *coefficient de corrélation de Pearson*, noté r .

Cet indice permet d'évaluer l'intensité, à partir de sa valeur absolue $|r|$, et la relation inverse ou directe entre deux variables, à partir de son signe $sgn(r)$.

Si $r = 0$, il n'y a pas de relation linéaire entre les deux variables, tandis qu'une valeur de $|r|$ proche de 1 est le reflet d'une forte liaison entre les deux variables.

3.6.1 Analyses des annotations sur les stimuli visuels

L'avantage que procure les stimuli visuels utilisés dans notre protocole expérimental réside

	Session	Corrélation Valence	Corrélation Activation
Sujet 1	1	0,8273	0,5209
Sujet 2	1	0,8165	0,7477
Sujet 3	1	0,6497	0,1467
Sujet 4	1	0,7681	0,2785
Sujet 5	1	0,8889	0,5668
	2	0,8225	0,2487
Sujet 6	1	0,6904	0,6817
	2	0,7927	0,7975
Sujet 7	1	0,7796	0,7725
Sujet 8	1	0,8717	0,7706
	2	0,9004	0,7632

TABLEAU 3.4 – Coefficient de la corrélation de Pearson entre les scores IAPS et l'auto-évaluation par session de chaque sujet du corpus EMOGEE sur les images de calibrage.

dans une annotation de référence solide fournie par la base IAPS. Chaque image a été évaluée au moins une centaine de fois sur les dimensions de *valence*, *activation*, et *dominance*.

Dans un premier temps, les données fournies par la phase de calibrage vont être analysées en comparant les annotations des participants avec les données IAPS, puis pour les participants ayant réalisés les deux sessions de l'expérience, nous allons confronter les évaluations intra-annotateur.

Enfin, les données obtenues concernant la partie des blocs eNTERFACE'06 sont analysées d'un point de vue inter-annotateur et comparées aux annotations fournies dans la base eNTERFACE'06.

Analyse des données de calibrage

Dans la phase de calibrage il est demandé au participant de situer 9 images issues de la base IAPS sur l'axe de valence puis d'activation selon une échelle continue allant de 1 à 9 (voir Section 3.3.2). Afin d'établir une relation entre les évaluations de chaque participant et les scores moyens de chaque image fournis par la base IAPS (Tableau 3.2), nous calculons les coefficients de corrélation de Pearson pour les dimensions de valence et d'activation (Tableau 3.4).

Si les annotations de la dimension de valence sont en accord avec les scores d'IAPS, avec une moyenne de 0,80, l'accord semble moins évident sur la dimension d'activation, avec un coefficient de corrélation moyen de 0,57. Ces premiers résultats nous conforte sur le choix de l'intégration d'une phase de calibrage dans le protocole expérimental. Cette corrélation plus faible dans les annotations de la dimension de valence peut avoir deux origines : une réelle différence dans le ressenti de l'émotion véhiculée par les stimuli ou une appréhension différente de l'utilisation de cette dimension.

Dans ce corpus, trois participants ont réalisé les deux sessions de l'expérience. Les coefficients de Pearson pour les scores indiqués dans la phase de calibrage entre les deux sessions sont calculés (Tableau 3.5). Il est intéressant d'observer que pour un de ces sujet,

	Valence	Activation
Sujet 5	0,8586	0,3536
Sujet 6	0,8962	0,9392
Sujet 8	0,9611	0,9123

TABLEAU 3.5 – Coefficients de la corrélation de Pearson entre l'auto-évaluation de chaque session par sujet du corpus EMOGEE sur les images de calibrage.

à savoir le Sujet 4, le calibrage effectué a beaucoup varié entre les deux sessions et c'est surtout sa façon d'appréhender l'axe d'activation qui a affecté entre les deux sessions. Les corrélations des deux autres sujets sont quant à elles élevées, ce qui reflète une appréhension constante des axes d'annotations.

Analyse des blocs eNTERFACE'06

Les coefficients de Pearson entre les auto-évaluations et la moyenne des scores IAPS pour le corpus eNTERFACE'06 sont fournis par [Xu and \(Kostas\) Plataniotis \[2012\]](#) (Tableau 3.6). Nous réalisons la même démarche pour notre corpus (Tableau 3.7).

	Corrélation Valence	Corrélation Activation
Sujet 1	0,6994	0,9387
Sujet 2	0,6286	0,8628
Sujet 3	0,6550	0,6583
Sujet 4	0,9816	0,8533
Sujet 5	0,8057	0,7716

TABLEAU 3.6 – Coefficients de corrélation de Pearson entre les scores IAPS et les auto-évaluations par sujet de la base eNTERFACE'06 [[Xu and \(Kostas\) Plataniotis, 2012](#)].

Les coefficients de corrélations obtenus (en moyenne pour la valence et pour l'activation) nous indiquent que les annotations effectuées par les participants sur la dimension de valence sont proches du niveau de référence IAPS qui correspond à l'émotion que l'on cherche à véhiculer. L'évaluation sur l'axe d'activation est toutefois plus éloignée de la référence IAPS pour la plupart des sujets.

Le κ est calculé pour chaque paire d'annotateurs (Tableau 3.8). Comme en accord avec la phase de calibrage, l'accord des annotateurs est plus élevés sur la dimension de valence. Avec un κ moyen de 0,56 sur les deux axes, l'accord est modéré.

Lorsque l'on compare les annotations de plus de trois sujets (Tableau 3.9), l'accord sur la dimension de valence est toujours supérieure à celle de l'activation. Il est toutefois important de noter que pour le calcul de κ_F , les annotateurs ne sont pas forcément les mêmes pour l'ensemble des blocs. C'est le nombre d'annotations par bloc et non pas la référence de l'annotateur qui importe dans ce calcul.

	Valence	Activation
Sujet 1	0,6923	0,4847
Sujet 2	0,9529	0,8085
Sujet 3	0,8811	0,5357
Sujet 4	0,7678	0,4229
Sujet 5	0,9050	0,6804
Sujet 6	0,9586	0,7670
Sujet 7	0,9652	0,5646
Sujet 8	0,9349	0,7610

TABLEAU 3.7 – Coefficient de la corrélation de Pearson entre les scores IAPS et l'auto-évaluation des blocs eNTERFACE'06 par sujet du corpus EMOGEE.

Annotateur_1	Annotateur_2	Valence	Activation	Valence & Activation	Nbre de blocs
1	2	0,74	0,26	0,40	15
1	3	0,53	0,12	0,36	13
1	4	0,84	0,29	0,54	15
1	5	0,67	0,28	0,72	25
1	6	0,60	0,44	0,58	25
1	7	0,50	0,21	0,59	9
1	8	0,59	0,55	0,67	25
2	3	0,88	0,48	0,69	12
2	4	0,72	0,33	0,46	25
2	5	0,78	0,13	0,27	25
2	6	0,91	0,63	0,74	25
2	7	0,70	0,21	0,37	14
2	8	0,92	0,45	0,61	25
3	4	0,69	0,36	0,51	12
3	5	0,66	0,12	0,45	25
3	6	0,87	0,35	0,65	25
3	7	0,63	0,12	0,51	13
3	8	0,84	0,33	0,63	25
4	5	0,71	0,11	0,47	25
4	6	0,72	0,15	0,43	25
4	7	0,59	0,03	0,45	14
4	8	0,77	0,18	0,49	25
5	6	0,78	0,26	0,58	50
5	7	0,85	0,14	0,72	25
5	8	0,81	0,34	0,70	50
6	7	0,69	0,23	0,51	25
6	8	0,90	0,78	0,85	50
7	8	0,72	0,35	0,64	25

TABLEAU 3.8 – Calcul de κ pour chaque paire d'annotateurs pour les évaluations sur l'axe de valence et l'axe d'activation des stimuli visuels.

Nbre d'annotateurs	Nbre de blocs	Valence	Activation
3	50	0,21	0,05
4	45	0,21	0,05
5	37	0,22	0,05
6	26	0,26	0,05
7	15	0,17	0,05
8	2	0,08	-0,07

TABLEAU 3.9 – Coefficients κ_F en fonction du nombre d'annotateurs pour l'annotation selon les dimensions de valence et d'activation des stimuli visuels.

3.6.2 Analyse des annotations sur les stimuli audio-visuels

Les vidéos utilisées dans ce corpus ont la particularité d'être utilisées pour la première fois dans le but de provoquer une émotion chez son observateur. Elles ne possèdent donc pas de vérité terrain sur le plan de l'évaluation de l'émotion ressentie. Une analyse statistique des annotations obtenues est donc nécessaire pour situer de manière plus concrète la place de ces stimuli dans le plan valence/activation.

De même que pour l'analyse des annotations des stimuli visuels, le κ est calculé pour chaque paire d'annotateurs (Tableau 3.10). Les degrés d'accord sont très variables en fonction des paires d'annotateurs. Cela est confirmé par le calcul des coefficients κ_F (Tableau 3.11), qui mettent en lumière la faible cohérence dans l'annotation des stimuli vidéos.

En comparant les moyennes des coefficients κ entre l'annotation globale et dynamique (Tableau 3.12), on observe une subtile amélioration de l'accord inter-annotateur en adoptant l'annotation dynamique. Cette amélioration pourrait être due à une plus grande cohérence dans l'annotation des segments que sur la décision finale de l'annotation générale de la vidéo.

	Valence	Activation	Val. & Act.
Annotation Globale	0,12	0,08	0,09
Annotation Dynamique	0,19	0,08	0,11

TABLEAU 3.12 – Moyenne des coefficients κ pour l'annotation globale et dynamique des stimuli audio-visuels suivant les dimensions de valence et d'activation et leur regroupement.

Les résultats sur les degrés d'accord κ_F sont tout de même proches entre l'annotation globale et l'annotation dynamique (Tableaux 3.11 et 3.13). Cela est lié à une forte corrélation de ces annotations. L'observateur, après avoir évalué trois étapes dans son ressenti émotionnel, va conclure pour une annotation globale qui en est proche. Il pourrait être intéressant, pour l'étude de l'approche dynamique, de se focaliser sur les déplacements dans l'annotation des trois segments de chaque vidéo, plutôt que sur la valeur attribuée sur chacune des dimensions.

Annotateur_1	Annotateur_2	Valence	Activation	Valence & Activation	Nbre de vidéos
1	2	0,32	0,19	0,40	41
1	3	0,52	0,24	0,31	27
1	4	0,48	0,02	0,18	41
1	5	0,48	0,12	0,11	50
1	6	-0,13	-0,01	0,01	50
1	7	-0,13	-0,17	-0,16	25
1	8	0,34	0,28	0,26	50
2	3	0,02	0,19	-0,03	30
2	4	0,23	0,17	0,07	50
2	5	0,44	0,09	-0,06	50
2	6	-0,01	0	0,08	50
2	7	-0,57	-0,04	-0,22	25
2	8	0,54	0,27	0,31	50
3	4	0,23	0,27	0,29	30
3	5	0,34	0,18	0,37	50
3	6	-0,08	0,08	-0,01	50
3	7	0,09	-0,01	0,06	31
3	8	0,13	0,34	0,31	50
4	5	0,35	0,20	0,34	50
4	6	0,14	0,14	0,11	50
4	7	-0,17	0,13	0,02	25
4	8	0,29	0,21	0,25	50
5	6	-0,09	-0,02	-0,12	100
5	7	-0,28	-0,06	-0,07	50
5	8	0,52	0,25	0,41	100
6	7	-0,15	0,19	0,04	50
6	8	-0,03	-0,15	-0,12	100
7	8	-0,38	-0,24	-0,29	50

TABLEAU 3.10 – Calcul de κ pour chaque paire d'annotateurs pour les évaluations sur l'axe de valence et l'axe d'activation des stimuli audio-visuels.

Nbre d'annotateurs	Nbre de vidéos	Valence	Activation
3	100	0,07	0,02
4	87	0,07	0,02
5	67	0,05	0,02
6	45	0,05	0,01
7	36	0,02	0
8	15	0,02	-0,01

TABLEAU 3.11 – Coefficients κ_F en fonction du nombre d'annotateurs pour l'annotation globale selon les dimensions de valence et d'activation des stimuli audio-visuels.

Nbre d'annotateurs	Nbre de segments	Valence	Activation
3	300	0,08	0,01
4	261	0,08	0,01
5	201	0,08	0,01
6	135	0,09	0,01
7	108	0,05	0
8	45	0,03	0

TABLEAU 3.13 – Coefficients κ_F en fonction du nombre d'annotateurs pour l'annotation dynamique selon les dimensions de valence et d'activation des stimuli audio-visuels.

Stimuli	Valence	Activation	Valence & Activation
Visuels	0,75	0,30	0,57
Audio-Visuels (Annot. Globales + Dynamiques)	0,18	0,08	0,11
Visuels & Audio-Visuels (Annot. Globale)	0,49	0,18	0,32

TABLEAU 3.14 – Moyenne des coefficients κ en fonction du type de stimuli annotés.

3.6.3 Bilan

Les annotations fournies par les participants sont essentielles à l'expérience en ce sens qu'elles permettent une approche préliminaire des données enregistrées. Cette approche permet de mettre en lumière les comportements de chaque sujet face aux techniques d'élicitation.

Les grands écarts de valeurs entre κ et κ_F , sont dus au choix de pénalisation dans les écarts d'annotation, puisque κ_F est calculé sans pondération spécifique.

A travers les analyses menées nous pouvons nous attendre à connaître des obstacles dans la conception d'un système de reconnaissance qui serait appliqué à l'ensemble des sujets.

3.7 Conclusion

Ce corpus, destiné à l'analyse de l'émotion ressentie, a été créé en gardant plusieurs objectifs à l'esprit. En se reposant sur les travaux précédents dans le domaine de la reconnaissance des émotions, une approche multimodale a été privilégiée pour plusieurs raisons : tout d'abord une vérification de la présence de l'émotion par le biais de l'analyse de signaux physiologiques, et ainsi être en mesure de valider la solidité du corpus, mais également afin de connaître les actions menées par le participant pouvant mener à l'apparition d'artefacts dans les signaux EEG. Ce corpus a été réalisé pour garantir une utilisation optimale des signaux EEG dans le cadre de la reconnaissance automatique des émotions. Une approche novatrice dans l'annotation de l'émotion a également été mise en place, à savoir l'intégration d'une dynamique de l'émotion dans la phase d'auto-évaluation.

L'utilisation conjointe de ces modalités permet de proposer un corpus pouvant être utilisé dans divers domaines d'applications nécessitant un système de reconnaissance automatique des émotions.

Une des plus grandes difficultés a été l'ajustement d'un compromis entre un nombre suffisant de stimuli présentés au participant et la charge cognitive qui lui est imposée. En effet, l'installation du matériel, et tout particulièrement des dispositifs d'acquisition des différents signaux, est déjà relativement chronophage. Il faut optimiser au mieux le temps restant avant que le participant rentre dans un état de fatigue trop important et qu'il en résulte une annotation erronée et des signaux EEG peu révélateurs car noyés dans des ondes alpha.

L'utilité de ce corpus est dans un premier temps de pouvoir réaliser des études de reconnaissance automatique des émotions dans différentes configurations de modalités grâce à la précision de la synchronisation réalisée. Elle permet de vérifier la validité des différentes propositions de descripteurs pertinents dans les systèmes de classification, à la fois pour les signaux physiologiques et les signaux EEG.

Ce corpus n'a pas encore pu être exploité, puisque nous avons dans un premier temps porté nos travaux sur un autre corpus.

Chapitre 4

Etude comparative des caractéristiques

Sommaire

4.1 Introduction	88
4.2 Les descripteurs existants pour la reconnaissance des émotions	89
4.2.1 Les descripteurs temporels	89
4.2.2 Les descripteurs spectraux	90
4.3 Les descripteurs proposés	91
Les Moments Spectraux	91
Les descripteurs de la pente spectrale	92
Les mesures de platitude et de crête spectrale	92
Les autres descripteurs de la forme spectrale	93
4.4 Les paramètres d'extraction de caractéristiques	93
4.5 Données et Protocole d'évaluation	94
La validation croisée	95
4.6 Expériences et Résultats	97
4.6.1 Etude comparative des caractéristiques	97
4.6.2 Etudes des descripteurs proposés	99
Variabilité inter-sujet	99
Comparaison des caractéristiques	100
Capacité de généralisation	100
Configuration Mono-canal	102
4.7 Conclusion	102

Résumé

Dans ce chapitre nous exposons les études comparatives de descripteurs existants dans la littérature et de descripteurs que nous proposons.

Le système de reconnaissance repose sur une classification binaire au sein d'une dimension : valence positive/valence négative.

Nous étudions une problématique courante dans le cadre de traitement de données EEG : la variabilité inter-sujet.

Dans nos travaux nous allons explorer une grande gamme de caractéristiques différentes. Un certain nombre d'entre elles n'ont pas été utilisées auparavant dans le cadre de la reconnaissance des émotions. Ces caractéristiques, potentiellement pertinentes seront associées à des informations *temporelles*, *spectrales*, et *spatiales*. Une comparaison de ces caractéristiques sera effectuée par le biais d'une évaluation expérimentale rigoureuse.

En outre, nous allons étudier le comportement de ces caractéristiques à travers les différents sujets, et également sur plusieurs emplacements des électrodes EEG en vue d'examiner la possibilité d'exploiter ces caractéristiques dans une configuration d'EEG mono-canal. Ce dernier point est essentiel en ce sens qu'il correspond à une vraie tendance dans les configurations matérielles d'EEG à destination du grand public, dont on attend qu'ils soient le plus léger possible.

D'autre part, notre évaluation est basée sur un protocole strict qui expérimente les capacités de généralisation du système contrairement aux travaux précédents où l'on n'observe pas toujours une séparation claire entre les données utilisées pour la phase d'apprentissage et ceux utilisés pour la phase de test. Pour finir, nous essayons plusieurs combinaisons de différents descripteurs avec pour objectif d'optimiser la précision de la reconnaissance.

4.1 Introduction

La reconnaissance des émotions est habituellement envisagée sous l'angle d'un problème de classification où le choix de caractéristiques appropriées est essentiel dans le but de s'assurer une précision de reconnaissance satisfaisante. Une de nos problématiques repose sur le fait que, dans le cadre des caractéristiques EEG, un consensus n'a toujours pas été réalisé sur un ensemble standard de caractéristiques qui permettrait de garantir une distinction performante des émotions d'un sujet humain.

Il est toutefois reconnu dans le domaine des neurosciences qu'une part importante d'informations pertinentes est véhiculée par les propriétés spectrales des signaux EEG, où différents types d'activités humaines donnent lieu à différentes formes spectrales apparaissant dans des bandes de fréquences bien spécifiées.

Nous explorons une grande variété de caractéristiques temporelles, spectrales et spatiales pouvant être potentiellement utiles dans le cadre de la reconnaissance des émotions et nous les comparons à d'autres, exposées dans de précédents travaux, en utilisant un protocole expérimental rigoureux. Nous évaluons plus particulièrement l'efficacité de plu-

sieurs caractéristiques spectrales, n'ayant pas été précédemment proposées pour le problème de la classification, en deux classes : les émotions de valence positive versus négative.

4.2 Les descripteurs existants pour la reconnaissance des émotions

Deux types de caractéristiques peuvent être distinguées ici : celles utilisées dans de précédents travaux utilisant l'EEG comme une modalité pour la reconnaissance des émotions et celles exploitées dans d'autres applications EEG telles que les interfaces Homme-Machine.

Parmi ces dernières, le type de caractéristiques le plus couramment utilisé sont les caractéristiques dites standards [Picard et al., 2001]. Celles-ci peuvent être considérées comme limitées en ce sens qu'elles ne saisissent pas de manière efficace les informations utiles sur les caractéristiques spectrales des signaux. Par conséquent, les chercheurs ont étudié l'utilisation des caractéristiques contenant des informations spectrales. Elles sont essentiellement obtenues à partir du spectrogramme du signal de chaque d'électrode, en calculant les valeurs de puissance de sous-bande dans les bandes α , β , θ , γ . Les caractéristiques spatiales ont été également largement utilisées, en particulier les Common Spatial Patterns (CSP).

4.2.1 Les descripteurs temporels

Les descripteurs temporels les plus couramment utilisés dans le cadre de la reconnaissance des émotions basée sur l'EEG, et dont on emploie le terme de *standard* pour les désigner, sont les suivants :

- la moyenne des signaux bruts :

$$\mu_X = \frac{1}{N} \sum_{n=1}^N X_n, \quad (4.1)$$

- l'écart-type des signaux bruts :

$$\sigma_X = \left(\frac{1}{N-1} \sum_{n=1}^N (X_n - \mu_X)^2 \right)^{1/2}, \quad (4.2)$$

- la moyenne des valeurs absolues des premières différences des signaux bruts :

$$\delta_X = \frac{1}{N-1} \sum_{n=1}^{N-1} |X_{n+1} - X_n|, \quad (4.3)$$

- la moyenne des valeurs absolues des premières différences des signaux normalisés :

$$\tilde{\delta}_X = \frac{1}{N-1} \sum_{n=1}^{N-1} |\tilde{X}_{n+1} - \tilde{X}_n| = \frac{\delta_X}{\sigma_X}, \quad (4.4)$$

— la moyenne des valeurs absolues des secondes différences des signaux bruts :

$$\gamma_X = \frac{1}{N-2} \sum_{n=1}^{N-2} |X_{n+2} - X_n|, \quad (4.5)$$

— la moyenne des valeurs absolues des secondes différences des signaux normalisés :

$$\tilde{\gamma}_X = \frac{1}{N-2} \sum_{n=1}^{N-2} |\tilde{X}_{n+2} - \tilde{X}_n| = \frac{\gamma_X}{\sigma_X}, \quad (4.6)$$

Les descripteurs finaux correspondent à ces statistiques calculées à travers toutes les électrodes, ce qui permet de donner une idée de l'activité de différentes zones du cerveau.

4.2.2 Les descripteurs spectraux

Les caractéristiques fréquentielles ont également été étudiées dans des travaux précédents, et les *Common Spatial Patterns* (CSP) [Blankertz et al., 2008; Koelstra et al., 2010a; Li and Lu, 2009; Winkler et al., 2010] figurent parmi les plus utilisées dans le cadre de données multi-canaux. Le principe des CSP repose sur la décomposition d'un signal multi-canaux en sous-composantes qui présentent un contraste de variance. Ces caractéristiques ont également la propriété de pouvoir fournir des informations spatiales. La méthode utilisée pour les CSP est une technique d'analyse des données multi-canaux qui a pour objectif de trouver des filtres spatiaux pour une classification binaire supervisée. Les CSP produisent une décomposition supervisée du signal par une matrice $W \in \mathbb{R}^{C \times C}$, avec C correspondant au nombre de canaux, qui projette le signal $x(t) \in \mathbb{R}^C$ de l'espace d'origine des capteurs dans $x_{\text{CSP}}(t) \in \mathbb{R}^C$ qui correspond à l'espace de capteurs substitutif défini comme suit :

$$x_{\text{CSP}}(t) = W^T x(t). \quad (4.7)$$

Les signaux de $x_{\text{CSP}}(t)$ sont décorrélés et on peut réécrire l'équation ci-dessus :

$$x_{\text{CSP}j}(t) = w_j^T x(t), \quad (4.8)$$

avec $j = 1, 2, \dots, C$, et où w_j est un vecteur propre généralisé qui satisfait les propriétés suivantes :

$$w_j^T \Sigma^{(+)} w_j = \lambda_j^{(+)}, w_j^T \Sigma^{(-)} w_j = \lambda_j^{(-)}, \lambda_j^{(+)} + \lambda_j^{(-)} = 1, \quad (4.9)$$

où $\Sigma^{(+)}$ et $\Sigma^{(-)} \in \mathbb{R}^{C \times C}$ correspondent aux matrices de covariance de deux classes de signaux EEG comprenant C canaux. Il découle de la dernière contrainte énoncée ci-dessus, que w_j possède une forte variance du signal pour une classe et une faible variance pour l'autre classe. C'est cet effet contraire qui sert à discriminer les deux classes. Dans le cadre de la classification basée sur les CSP, on travaille donc habituellement par paires opposées.

Des variantes de CSP ont également été proposées dans de précédents travaux. Ces variantes reposent sur différentes décompositions fréquentielles. Un des paramètres principaux dans l'utilisation de CSP pour la classification est le nombre de paires de filtre spatiaux

à considérer.

Les CSP sont extraits sur l'ensemble du domaine fréquentiel, mais également sur des signaux EEG filtrés par des filtres passe-bande pour chaque bande fréquentielle associée aux ondes EEG.

4.3 Les descripteurs proposés

Parmi nos caractéristiques proposées, décrites dans [Essid \[2005\]](#); [Peeters \[2004\]](#) et qui n'ont pas été utilisées auparavant dans l'analyse des signaux EEG, nous pouvons distinguer deux sous-ensembles : les descripteurs temporels et les descripteurs fréquentiels.

Le descripteur temporel que l'on considère dans nos travaux correspond à des coefficients d'autocorrélation. Ce descripteur est obtenu en conservant les premiers coefficients de la transformée de Fourier inverse du périodogramme du signal, qui approxime sa densité spectrale de puissance (DSP).

Les Moments Spectraux

On considère le spectre comme une variable aléatoire dont les valeurs sont les fréquences. Les moments μ_i sont alors définis par : $\mu_i = \frac{\sum_{k=1}^K (f_k)^i a_k}{\sum_{k=1}^K a_k}$; a_k étant l'amplitude de la $k^{\text{ème}}$ composante de la transformée de Fourier du signal avec une fréquence $f_k = \frac{k}{N}$. Les Moments Spectraux, sont un sous-ensemble de caractéristiques qui reposent sur les quatre premiers moments statistiques des spectres de magnitude du signal EEG [[Gillet and Richard, 2004](#); [Peeters, 2004](#)] :

— *Le centroïde spectral* :

$$S_c = \mu_1, \quad (4.10)$$

On considère le spectre comme une variable aléatoire dont les valeurs sont les fréquences, et les probabilités de les observer correspondent à l'amplitude normalisée.

— *la largeur spectrale* :

$$S_w = \sqrt{\mu_2 - \mu_1^2}, \quad (4.11)$$

elle représente l'étendue du spectre autour de sa moyenne, c'est-à-dire la variance de la distribution définie ci-dessus.

— *l'asymétrie spectrale*, définie à partir de la *skewness* :

$$S_a = \frac{2(\mu_1)^3 - 3\mu_1\mu_2 + \mu_3}{S_w^3}, \quad (4.12)$$

elle donne une mesure de l'asymétrie du spectre autour de sa moyenne :

- $S_a = 0$: la distribution est symétrique
- $S_a < 0$: plus d'énergie sur la droite
- $S_a > 0$: plus d'énergie sur la gauche

— la *platitude spectrale*, définie elle à partir du *kurtosis* :

$$S_k = \frac{-3\mu_1^4 - 6\mu_1\mu_2 - 4\mu_1\mu_3 + \mu_4}{S_w^4} - 3, \quad (4.13)$$

Elle donne une mesure de la platitude d'une distribution autour de sa moyenne. S_k est d'autant plus grande que son spectre est "piqué" autour de sa moyenne :

- $S_k = 0$: distribution normale (gaussienne)
- $S_k < 0$: distribution plus plate
- $S_k > 0$: distribution plus "piquée"

Les descripteurs de la pente spectrale

— la *pente spectrale* mesurant le taux de décroissance spectrale :

$$S_s = \frac{K \sum_{k=1}^K f_k a_k - \sum_{k=1}^K f_k \sum_{k=1}^K a_k}{K \sum_{k=1}^K f_k^2 - (\sum_{k=1}^K a_k)^2}. \quad (4.14)$$

Il s'agit d'une régression linéaire du spectre d'amplitude, et qui est dépendant linéairement du centroïde spectral [Peeters et al., 2011].

— la *décroissance spectrale* donnée par Krimphoff [1993] :

$$S_d = \frac{1}{\sum_{k=2}^K a_k} \sum_{k=2}^K \frac{a_k - a_1}{k-1}, \quad (4.15)$$

et qui représente la quantité de décroissance d'un spectre d'amplitude.

Les mesures de platitude et de crête spectrale

— une alternative à la description de la platitude spectrale par kurtosis [Dubnov, 2004; Johnston, 1988] est donnée par :

$$SF = \frac{\prod_k a_k^{\frac{1}{K}}}{\frac{1}{K} \sum_k a_k}. \quad (4.16)$$

Le facteur de crête spectrale (Spectral Crest Factors - SCF) est défini pour chaque sous-bande sb par :

$$SCF(sb) = \frac{\max_{k \in sb} a_k}{\frac{1}{K} \sum_{k \in sb} a_k}, \quad (4.17)$$

Les sous-bandes utilisées pour extraire ces descripteurs correspondent aux bandes fréquentielles des *ondes* EEG, à savoir : θ (4-8 Hz), α (8-12 Hz), β (12-30 Hz), γ (30-100 Hz).

Les autres descripteurs de la forme spectrale

- La variation temporelle du spectre aussi connue sous le nom de "*flux spectral*" [Krimphoff et al., 1994] :

$$S_v = \frac{\sum_{k=1}^K a_k(t-1)a_k(t)}{\sqrt{\sum_{k=1}^K a_k(t-1)^2} \sqrt{\sum_{k=1}^K a_k(t)^2}}, \quad (4.18)$$

Elle correspond à une corrélation normalisée entre deux spectres d'amplitude successifs. C'est un descripteur spectro-temporel :

- proche de 0 si les spectres successifs sont similaires,
- proche de 1 si les spectres successifs sont très dissemblables.

Elle caractérise la variation dynamique de l'information spectrale.

- La *fréquence de coupure* calculée comme la fréquence en-dessous de laquelle 99% de l'énergie spectrale est prise en compte.
- Les coefficients LPC (Linear Prediction Coding) correspondant aux premiers coefficients du filtre obtenu par une analyse Auto-Régressive (AR) à l'ordre 2 du signal [Makhoul, 1975].

Descripteur	Taille	Description
$S_x = [S_c, S_w, S_a, S_k]$	10	Moments spectraux.
$SF = [SF1, \dots, SF4]$	4	Platitude spectrale.
$SCF = [SCF1, \dots, SCF4]$	4	Facteur de crête spectrale.
$AR = [AR1, AR2]$	2	Coefficients LPC.
$[S_s, S_d, S_v, S_o, F_c]$	5	Pente, décroissance, variation temporelle, platitude du spectre, fréquence de coupure.
$AC = [AC1, \dots, AC49]$	49	Coefficients d'autocorrelation.

TABLEAU 4.1 – Liste des descripteurs proposés dans nos travaux. Ces descripteurs sont extraits pour chaque signal d'électrode EEG.

4.4 Les paramètres d'extraction de caractéristiques

La mise en œuvre du système de classification demande dans un premier temps un paramétrage dans l'extraction des descripteurs étudiés. Ces paramètres correspondent à des représentations de la signification des descripteurs que l'on souhaite extraire.

Les paramètres de taille de fenêtre temporelle choisis pour chaque descripteur sont données dans le Tableau 4.2. Les fréquences très basses sont caractéristiques d'artefacts perturbant les signaux EEG, on applique donc un filtre passe-haut à 1 Hz sur les signaux avant d'en extraire les caractéristiques.

Les bandes fréquentielles sur lesquelles sont définies les ondes EEG (voir Annexe A.1.2) sont les suivantes :

- θ : 4-8 Hz,

Domaine	Caractéristiques	Description	Fenêtre [s]	Recouvrement [%]
Temporel	“Standard” [Picard et al., 2001]	min, max, skewness, kurtosis, moyenne, écart-type, médiane, moyenne et max des valeurs absolues des premières différences, moyenne et max des valeurs absolues des secondes différences	0.5	50
Fréquentiel	Spectrogramme		0.5	50
Spatial	Common Spatial Pattern (CSP) [Li and Lu, 2009]	bande θ , bande α , bande γ , bande β et l'ensemble des fréquences	0.25	-

TABLEAU 4.2 – Caractéristiques extraites pour la reconnaissance des émotions basée sur l’EEG.

- α : 8-12 Hz,
- β : 12-30 Hz,
- γ : 30-100 Hz.

On effectue une intégration précoce des descripteurs extraits en optant pour prendre comme taille de référence la plus grande fenêtre temporelle et on effectue un moyennage des sous-segments pour former la nouvelle fenêtre temporelle correspondante.

4.5 Données et Protocole d’évaluation

La base de données retenue pour notre étude expérimentale est la base ENTERFACE'06 qui est constituée d’enregistrements de signaux EEG et physiologiques pour 5 sujets masculins qui observent des images véhiculant un contenu émotionnel varié. Trois états émotionnels d’intérêt sont considérés : calme, excité positivement et excité négativement. Chaque enregistrement consiste en une session de 15 minutes et chaque session contient 30 blocs, 10 par état émotionnel. Un bloc est une succession de 5 images correspondant à un seul état émotionnel. 3 sessions sont réalisées pour chaque sujet. L’ensemble de la base de données comprend 90 blocs, à savoir 450 images.

Après le passage d’un bloc de 5 images, il est demandé au sujet d’auto-évaluer son état émotionnel en utilisant la technique d’auto-évaluation Manikin (Self-Assessment Manikin-SAM) [Bradley and Lang, 1994], en attribuant une note comprise entre 1 et 5 pour les composantes de valence et d’éveil (“arousal”).

Les images constituant chaque bloc sont issues d’un *dataset* de référence : *International Affective Picture System* (IAPS) [Lang et al., 2008], qui est un ensemble de 1196 images avec un contenu émotionnel varié. Pour les enregistrements de ENTERFACE'06, 3 sous-ensembles d’images ont été sélectionnés, un pour chaque état émotionnel, suivant des seuils spécifiques le long des axes de valence et arousal.

- calme : 4 < valence < 6 ; arousal < 4
- **excité positif** : valence > 6.8 ; arousal > 5
- **excité négatif** : valence < 3 ; arousal > 5

Cela a donné lieu à une sélection de respectivement 106, 71, et 50 images pour les 3 classes d’émotion considérées. Les signaux EEG sont enregistrés sur 54 électrodes à une fréquence d’échantillonnage de 1024 Hz.

Les caractéristiques sont évaluées à travers les performances des classificateurs qui les exploitent, où la performance est mesurée selon le taux de bonne reconnaissance de la *valence*, correspondant dans la base ENTERFACE'06 à une discrimination entre les classes excité positif et excité négatif. De plus, nous utilisons une procédure de validation croisée conçue de telle manière à évaluer la capacité de généralisation du système de reconnaissance. Xu *et al.* [Xu and (Kostas) Plataniotis, 2012] utilisent un processus de validation croisée *5-fold*, utilisant de façon aléatoire 80% des échantillons pour l'apprentissage et 20% pour le test. Cette approche a l'inconvénient de ne pas séparer clairement les enregistrements pour apprendre le classificateur et ceux utilisés pour fournir les données de test, limitant ainsi la validité des résultats. C'est pourquoi nous considérons un protocole *leave-one-bloc out* où les données correspondant à un bloc de ENTERFACE'06 sont utilisées pour l'ensemble de test et le reste des données est utilisé pour l'apprentissage.

Deux schémas alternatifs sont alors évalués : le schéma inter-session où les données de l'ensemble d'apprentissage sont issues de l'ensemble des sessions, y compris celle dont est issue le bloc de l'ensemble de test ; et le schéma intra-session où les blocs utilisés pour l'ensemble d'apprentissage proviennent uniquement de la même session que celle de l'ensemble de données test.

La validation croisée

La mise en place d'une méthode de validation croisée est une tentative de réponse à la faible quantité de données. Il s'agit en effet d'un problème récurrent lorsque l'on traite des données EEG dans le domaine de l'informatique affective.

On considère les données par blocs, qui pour rappel correspondent chacun à la succession de cinq images appartenant à une même classe. Les données d'un bloc vont être conservées afin de tester le classifieur, et selon la configuration utilisée, tout ou partie des données restantes sont utilisées pour l'apprentissage. On réitère l'opération en isolant à chaque fois un bloc différent pour le test, et ceci pour tous les blocs disponibles, c'est-à-dire dont les images appartiennent aux classes que l'on souhaite discriminer.

Plusieurs options dans le modèle de validation croisée ont été étudiées, à savoir un schéma inter-session et un schéma intra-session :

Schéma Inter-session

Pour réaliser une validation croisée sur la base du schéma inter-session, aucune distinction n'est effectuée entre les sessions. Plus précisément, cela revient à effectuer l'apprentissage du classifieur sur l'ensemble des données des trois sessions, hormis pour les données d'un bloc qui sont utilisées pour la phase de test (voir Figure 4.1).

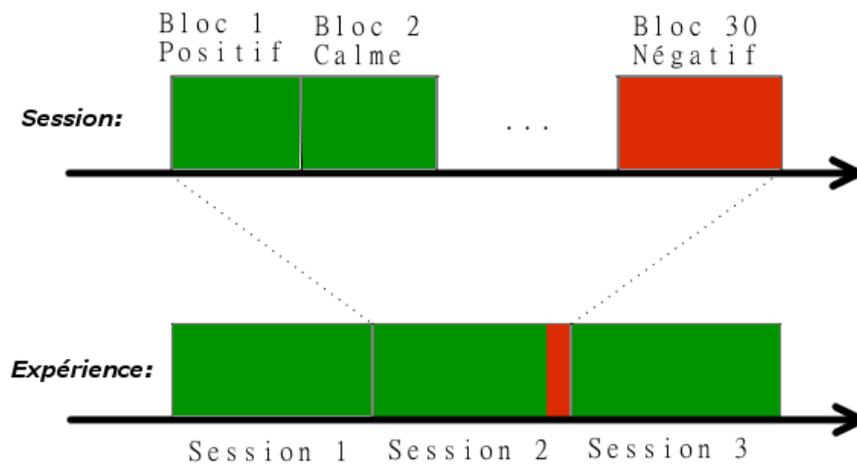


FIGURE 4.1 – Schéma de validation croisée inter-session. Les données de la phase d'apprentissage (i.e. un bloc) sont en vert et les données de la phase de test sont en rouge.

Schéma Intra-session

A l'instar du schéma inter-session, la configuration de la validation croisée sur le schéma intra-session n'utilise que les données d'un bloc pour la phase de test. Cependant, la différence réside dans les données utilisées dans la phase d'apprentissage. Seules les données appartenant à la même session que le bloc de test sont conservées pour réaliser l'apprentissage du classifieur (voir Figure 4.2).

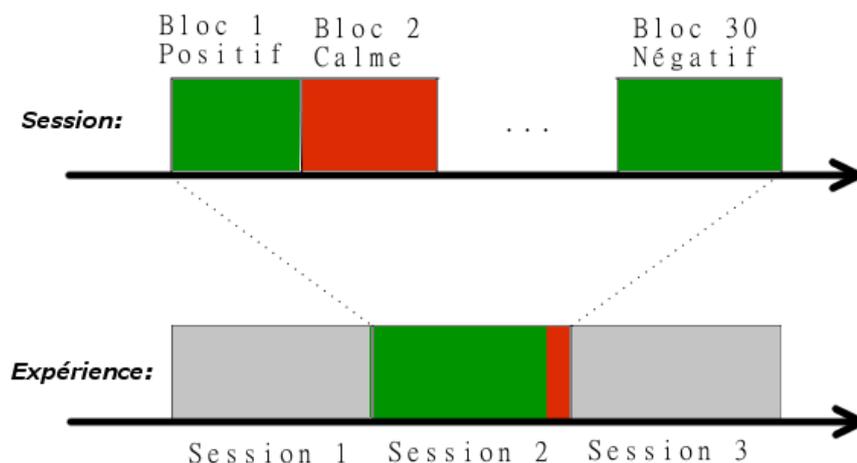


FIGURE 4.2 – Schéma de validation croisée intra-session. Les données de la phase d'apprentissage (i.e. un bloc) sont en vert et les données de la phase de test sont en rouge.

Les résultats correspondant à chacune de ces situations seront discutés ci-après.

4.6 Expériences et Résultats

Notre étude se penche sur l'exploration de nouvelles caractéristiques, mais en prenant une direction originale, qui envisage des caractéristiques qui se prêtent à une application mono-canal de la reconnaissance des émotions basée sur l'EEG (essentiellement des alternatives de caractéristiques spectrales), tout en les comparant, grâce à une évaluation expérimentale rigoureuse, à un grand nombre de caractéristiques existantes issues de travaux antérieurs (notamment vues dans les études envisageant la réduction du nombre d'électrodes) [Jenke et al., 2013; Xu and (Kostas) Plataniotis, 2012]. De plus, nous étudions le comportement de ces caractéristiques par le biais de différents sujets et de différentes positions des électrodes EEG, et nous examinons la possibilité d'exploiter ces caractéristiques dans le cadre d'une configuration EEG mono-canal. Ce dernier point est important car il correspond à une tendance dans les configurations de l'équipement EEG pour les applications grand public, qui cherchent à rester le moins encombrant possible. Enfin, nous testons plusieurs combinaisons de différents groupes de caractéristiques dans le but d'optimiser la précision de reconnaissance.

Notre évaluation est basée sur un protocole strict qui teste les capacités de généralisation de système contrairement à de précédents travaux où une séparation nette des enregistrements utilisés pour l'ensemble d'apprentissage et ceux utilisés dans l'ensemble de test n'a pas toujours été observée [Xu and (Kostas) Plataniotis, 2012].

La classification est réalisée à l'aide de SVM linéaire suivant l'axe de Valence des données du dataset eNTERFACE'06 [Savran et al., 2006] (positif vs. négatif). Les principaux résultats obtenus pour les différents descripteurs et pour chaque sujet sont données dans le Tableau 4.3. Les résultats obtenus sont en accord avec les résultats fournis par la littérature (Tableau 2.6).

4.6.1 Etude comparative des caractéristiques

		Sujet 1	Sujet 2	Sujet 3	Sujet 4	Sujet 5
Standards		30%	75%	58%	50%	67%
Spectrogramme	Theta	52%	48%	55%	43%	63%
	Alpha	65%	50%	47%	42%	70%
	Beta	48%	52%	48%	37%	55%
	Gamma	48%	50%	48%	43%	57%
	Tout	55%	58%	50%	45%	50%

TABLEAU 4.3 – Résultats obtenus pour chaque sujet en fonction de l'utilisation de différentes caractéristiques. Les meilleurs scores de chaque sujet sont indiqués en gras.

Concernant les descripteurs temporels, ils sont utilisés conjointement dans le classifieur. Ces descripteurs apportent une grande variabilité entre les sujets dans les résultats, allant de 30% de taux de bonne reconnaissance pour le Sujet 1 jusqu'à 75% pour le Sujet 2.

Les premiers résultats obtenus avec les CSP sur le corpus ENTERFACE'06 sont prometteurs. Plusieurs variantes sont étudiées avec l'extraction de caractéristiques suivant les différentes bandes de fréquences correspondant aux ondes EEG, et l'utilisation de nombres

différents de composantes de CSP. Les résultats sont obtenus pour les différentes bandes séparément (Tableau 4.4) puis en réalisant plusieurs combinaisons par paires de bandes fréquentielles (Tableau 4.5).

	Alpha	Beta	Gamma	Theta
Sujet 1	38%	37%	45%	30%
Sujet 2	37%	48%	55%	62%
Sujet 3	60%	62%	55%	30%
Sujet 4	40%	60%	38%	42%
Sujet 5	58%	47%	53%	67%

TABLEAU 4.4 – Résultats obtenus avec les CSP pour 20 composantes. Les scores les plus importants sont surlignés en vert.

	Alpha/Gamma	Theta/Alpha	Beta/Gamma	Theta/Gamma	Beta/Alpha	Beta/Theta
Sujet 1	45%	33%	52%	40%	42%	43%
Sujet 2	58%	53%	58%	60%	46%	55%
Sujet 3	60%	38%	60%	47%	57%	40%
Sujet 4	47%	48%	55%	45%	50%	52%
Sujet 5	57%	60%	42%	62%	52%	68%

TABLEAU 4.5 – Résultats obtenus avec les combinaisons de CSP pour 20 composantes. Les scores les plus importants sont surlignés en vert.

	Alpha/Gamma	Theta/Alpha	Beta/Gamma	Theta/Gamma	Beta/Alpha	Beta/Theta
Sujet 1	45%	43%	37%	55%	45%	42%
Sujet 2	48%	50%	58%	50%	51%	52%
Sujet 3	52%	40%	47%	43%	53%	42%
Sujet 4	53%	58%	57%	57%	52%	47%
Sujet 5	50%	48%	42%	50%	50%	43%

TABLEAU 4.6 – Résultats obtenus avec les combinaisons de CSP pour 40 composantes. Les scores les plus importants sont surlignés en vert.

On peut toutefois noter qu'un travail réalisé à partir de validation croisée sur les CSP, visant une sélection de bandes fréquentielles plus fine au sein de la bande γ et d'un nombre adéquat de composantes CSP pour chaque sujet, a permis à Li and Lu [2009] l'obtention de meilleurs résultats pour une classification binaire de l'émotion (joie vs. tristesse). Si l'on compare les résultats donnés par le Tableau 4.5 et le Tableau 4.6, dont la seule différence est le nombre de composantes de CSP utilisées pour la classification, on observe en effet que chaque sujet répond différemment suivant le nombre de composantes choisies, le Sujet 4 obtient un meilleur résultat avec 40 composantes (58% pour la combinaison θ et γ) qu'avec 20 composantes (48% pour la même combinaison), tout comme les Sujets 1 et 3. Alors qu'à l'inverse, pour la même combinaison, le Sujet 2 et le Sujet 5 obtiennent de moins bons résultats avec l'utilisation de 40 composantes de CSP.

Domaine	Caractéristiques	Description	Fenêtre [s]	Recouvrement [%]
Temporel	“Standard” [Picard et al., 2001]	min, max, skewness, kurtosis, moyenne, écart-type, médiane, moyenne et max des valeurs absolues des premières différences, moyenne et max des valeurs absolues des secondes différences	0.5	50
Fréquentiel	Spectrogramme		0.5	50
	Spectrales	Moments, SE, SCF, Fréquence de coupure, Autocorrélation, Pente, Variation, Platitude, LPC	0.25	50
Spatial	Common Spatial Pattern (CSP) [Li and Lu, 2009]	bande θ , bande α , bande γ , bande β et l'ensemble des fréquences	0.25	-

TABLEAU 4.7 – Caractéristiques extraites pour la reconnaissance des émotions basée sur l’EEG. Les caractéristiques que nous proposons sont en gras.

	Toutes les électrodes						Electrode Cz						CSP		
	Inter-session			Intra-session			Inter-session			Intra-Session			Inter-session		
	Mts	SF	SCF	Mts	SF	SCF	Mts	SF	SCF	Mts	SF	SCF	α	γ	$\alpha + \gamma$
S1	30 %	45 %	48 %	38 %	33 %	32 %	53 %	40 %	28 %	52 %	58 %	35 %	38 %	45 %	45 %
S2	48 %	62 %	65 %	53 %	53 %	52 %	55 %	52 %	60 %	55 %	42 %	35 %	37 %	55 %	58 %
S3	62 %	53 %	48 %	37 %	48 %	53 %	52 %	63 %	70 %	37 %	35 %	33 %	60 %	55 %	60 %
S4	58 %	30 %	45 %	35 %	37 %	45 %	48 %	38 %	50 %	40 %	37 %	35 %	40 %	38 %	47 %
S5	70 %	70 %	73 %	62 %	60 %	58 %	58 %	67 %	57 %	57 %	50 %	57 %	58 %	53 %	57 %

TABLEAU 4.8 – Taux de bonne reconnaissance (selon la dimension de valence) pour un sous-ensemble (le meilleur) des caractéristiques proposées : Moments (Mts), SE, SCF, et CSP pour une dimension de $D_{CSP} = 20$ pour chacun des 5 sujets (S1 à S2). Les meilleures précisions sont en caractère gras.

4.6.2 Etudes des descripteurs proposés

Dans cette section nous nous concentrons sur les différentes expériences menées afin d’étudier les caractéristiques présentées ci-dessus. Pour l’ensemble de nos expériences nous utilisons des machines à vecteurs de support (Support Vector Machine - SVM) comme classificateurs. Par ailleurs, une intégration temporelle précoce est utilisée, dans laquelle les caractéristiques extraites sur des fenêtres de signal locales (de taille donnée dans le Tableau 4.7) sont moyennées suivant la durée des blocs. Le processus de classification, comme montré dans la Figure 2.7 (Chapitre 2), est alors effectué séparément pour chaque sujet en utilisant la procédure de validation croisée décrite ci-dessus.

Les caractéristiques sont extraites pour chaque sujet suivant l’ensemble des emplacements des 54 électrodes. Le Tableau 4.8 nous fournit, pour chaque sujet, les taux de bonne reconnaissance obtenues pour chaque groupe de caractéristiques, en utilisant les caractéristiques de l’ensemble des électrodes ou juste d’une seule, à savoir l’électrode centrale Cz, moins affectée par les artefacts. Il est important de noter que les précisions inférieures à 50% correspondent à des systèmes de reconnaissance inutiles qui ne permettent pas de faire mieux que le hasard.

Variabilité inter-sujet

Conformément aux conclusions des travaux antérieurs, on constate que les résultats varient considérablement selon les différents sujets (naturellement, les individus réagissent différemment aux stimuli subjectifs). De façon intéressante, deux sujets semblent se comporter dans deux directions extrêmes opposées : le sujet 5 dont les valeurs de valence sont prédites avec une grande précision (jusqu’à 73%) et le sujet 1 dont les états émotionnels ne peuvent pas être correctement caractérisés, ce qui résulte en de moins bons scores que le

hasard dans la majorité des configurations de caractéristiques et ne dépassant jamais 58%.

En comparant les étiquettes résultant de l'auto-évaluation des sujets avec les étiquettes de la réalité terrain des stimuli d'image correspondants (tel que prévu dans l'ensemble de données IAPS), nous constatons qu'elles étaient cohérentes pour le sujet 5, mais pas pour le sujet 1. Cela tend à confirmer que les réactions affectives de ces derniers sont tout à fait inhabituelles, donc probablement difficiles à prévoir.

Comparaison des caractéristiques

Dans un but de synthèse, seule une sélection de caractéristiques résultant des meilleurs résultats globaux de classification est conservée dans le Tableau 4.8. Par exemple, les caractéristiques CSP ont été extraites à la fois sur les signaux EEG en pleine bande et sur les signaux filtrés dans le répertoire des bandes theta (4-8 Hz), alpha(8-12 Hz), beta(12-30 Hz) et gamma(30-+ Hz). Aussi pour chacune de ces versions, la dimension réduite de CSP a varié dans l'ensemble {2, 4, 20, 40}. Les meilleurs résultats ont ainsi obtenus avec les bandes alpha et gamma à l'aide de 20 coefficients de CSP. Nous constatons à nouveau un comportement variable suivant les sujets car différentes bandes (parmi α et γ) semblent préférables pour des sujets différents. En outre, la combinaison de CSP par bandes ne conduit pas à de meilleurs scores, de même qu'un plus grand choix de colonnes ne profite pas aux résultats

En comparant nos nouvelles propositions de caractéristiques spectrales aux caractéristiques précédemment exploitées, y compris les différentes variantes de CSP, nous trouvons que ces premières ont de meilleures performances que ces dernières dans la plus grande majorité des cas. Les moments spectraux, aussi bien que les caractéristiques SF et SCF se révèlent très utiles.

Dans une seconde expérience, nous avons regroupé les différentes paires d'ensembles de caractéristiques présentées ci-dessus. Les résultats sont donnés dans le tableau où seules les combinaisons les plus efficaces sont présentées. Lorsqu'aucune combinaison ne permettait d'obtenir de meilleurs résultats qu'un groupe unique de caractéristiques, nous avons gardé ce dernier. Il se trouve que les ensembles de caractéristiques gagnantes ont tendance à inclure soit des attributs spectraux ou CSP, ou les deux. En examinant les résultats détaillés afin de mieux comprendre les interactions entre les différentes caractéristiques, nous remarquons que le sujet 2 répond très bien à un schéma spécifique qui consiste en une combinaison de 20 caractéristiques de CSP avec les caractéristiques Standards, aboutissant à des scores compris entre 76 % à 78 %. Les meilleurs systèmes de reconnaissance des sujets 3, 4 et 5 reposent aussi essentiellement sur des combinaisons avec CSP mais avec respectivement des coefficients LPC, des moments spectraux, et soit SF ou SCF.

Capacité de généralisation

Afin d'évaluer les capacités de généralisation des systèmes de reconnaissance étudiés, nous examinons les résultats obtenus en utilisant le protocole intra-session par rapport au protocole inter-session (voir la Section 4.5). Les résultats montrent des résultats similaires dans les protocoles intra-session et inter-session, mis à part pour le sujet 3, pour lequel les

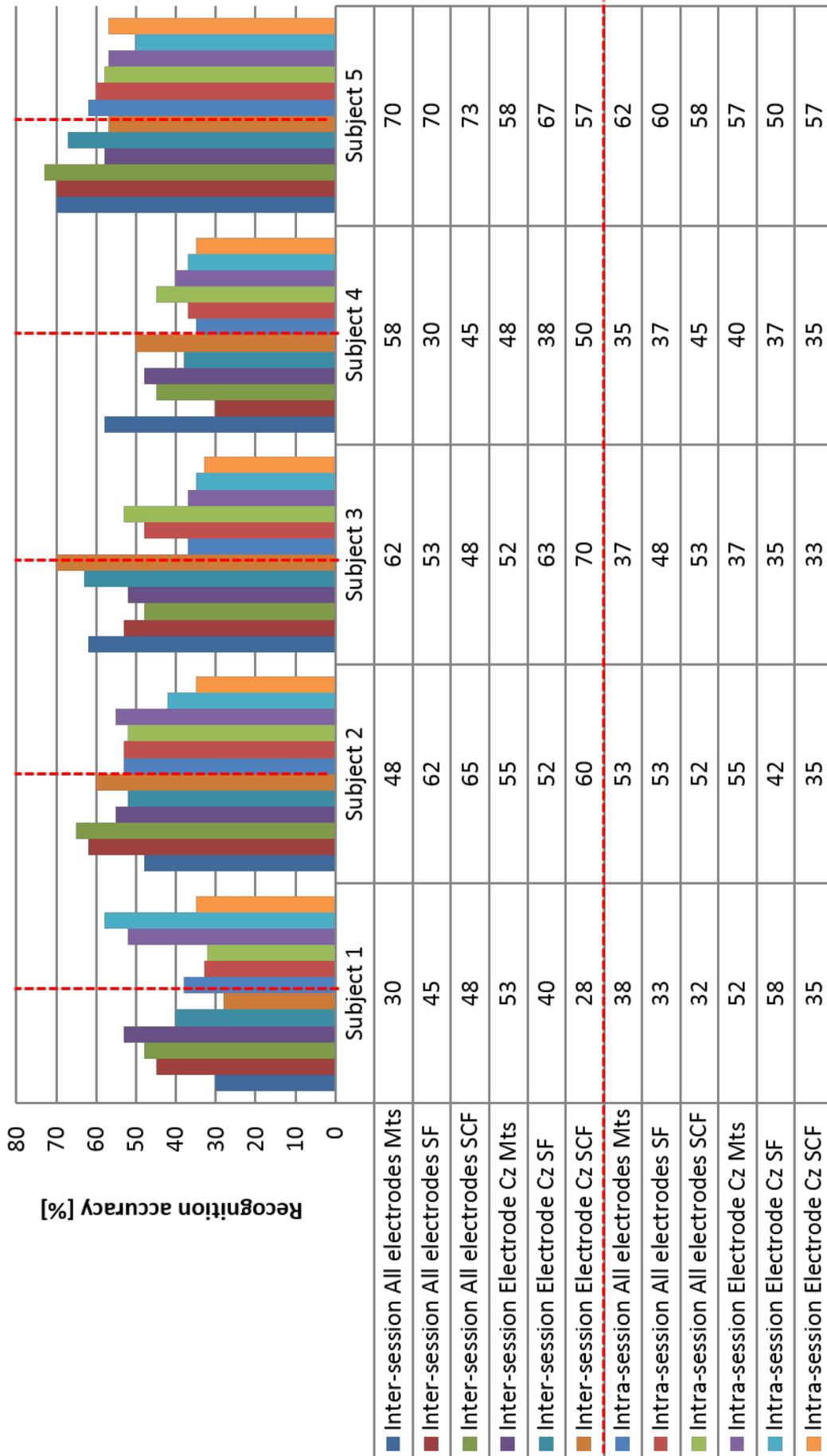


FIGURE 4.3 – Taux de bonnes reconnaissance obtenus pour chaque sujet en fonction de la configuration adoptée pour le système de classification.

résultats sont améliorés en considérant toutes les sessions dans le processus de validation croisée. Tous les sujets réagissent différemment, et les résultats qui sont améliorés par le schéma intra-session schéma différent d'un sujet à un autre, le sujet 5 obtient de meilleurs résultats dans le schéma inter-session, ainsi bien que d'une caractéristique à l'autre, le sujet 2 atteignant 65% avec SCF et 48% pour les moments spectraux lorsque l'on tient compte de toutes les sessions, et obtient respectivement 52% et 53% en ne conservant qu'une seule session.

Configuration Mono-canal

Dans une dernière expérience, nous explorons séparément les différentes caractéristiques dans le cadre d'une configuration EEG mono-canal. Pour ce faire, nous sélectionnons une électrode centrale, Cz, étant moins affectée par des artefacts. Les résultats correspondants pour les mêmes caractéristiques spectrales sont présentés dans le tableau 4.8. Il est intéressant de remarquer qu'avec cette configuration, nous avons des résultats similaires à ceux obtenus avec toutes les électrodes, atteignant même une précision de 70% avec les SCF pour le sujet 3. La configuration mono-canal est concevable car la plupart des informations nécessaires sont retrouvées en gardant uniquement une des électrodes.

	Score	Caractéristiques	Paramètres
Sujet 1	65 %	Puissance Spectrale dans la bande α	moyenne
Sujet 2	78 %	CSP β , α + Standards	$D_{CSP} = 20$
Sujet 3	70 %	SCF	Electrode Centrale (Cz)
Sujet 4	67 %	Variation Spectrale	
Sujet 5	77 %	CSP γ + SF	$D_{CSP} = 20$ or 40

TABLEAU 4.9 – Meilleurs scores avec leurs caractéristiques associées pour chaque sujet pour une classification binaire suivant la dimension de Valence.

	Score	Descripteurs	Paramètres
Sujet 1	65%	Puissance spectrale dans la bande α	Moyenne
Sujet 2	71%	Cutoff	Elec. centrale (Cz)
Sujet 3	64%	Spectral slope	Elec. centrale (Cz)
Sujet 4	73%	Spectral Flatness ou Cutoff	Elec. centrale (Cz)
Sujet 5	70%	Coefficients LPC	

TABLEAU 4.10 – Meilleurs scores avec leurs caractéristiques associées pour chaque sujet pour une classification binaire suivant la dimension d'Activation.

4.7 Conclusion

L'intérêt porté à la reconnaissance automatique des émotions a augmenté au cours de ces dernières années. Cela a permis la validation dans l'utilisation de descripteurs utilisés çà l'origine pour les signaux physiologiques, mais également l'émergence de nouveaux descripteurs qui s'orientent plus vers l'extraction d'informations fréquentielles et l'utilisation de séparation de source.

Un grand nombre d'expériences a été mené afin d'étudier l'utilité d'un éventail de caractéristiques EEG dans le cadre d'une classification binaire des émotions selon la dimension de valence. Une attention particulière a été apportée à l'utilisation d'un protocole de validation croisée approprié et conçu pour évaluer les capacités de généralisation des systèmes de reconnaissance testés.

Nos résultats montrent que les nouvelles caractéristiques spectrales que nous proposons sont compétitives comparées à celles précédemment utilisées. Elles nous amènent de plus vers une configuration mono-canal performante de la reconnaissance des émotions, ce qui implique un fort potentiel pour les applications destinées au grand public.

Bien que les résultats apportés soient corrects, on peut espérer trouver de nouvelles caractéristiques plus performantes dans ce cadre d'application. On pourra notamment trouver de nouveaux descripteurs pertinents dans un cadre fréquentiel.

Chapitre 5

Conclusions et Perspectives

Sommaire

5.1 Apports	107
Le corpus EMOGEE	107
Le système de reconnaissance des émotions	107
5.2 Perspectives	108
5.2.1 Perspectives à court terme	108
5.2.2 Perspectives à long terme	108

5.1 Apports

Les apports de cette thèse portent sur deux éléments principaux, à savoir la création d'un corpus multimodal destiné à la reconnaissance des émotions et l'étude de caractéristiques par un système de reconnaissance des émotions basé sur l'EEG.

Le corpus EMOGEE

Un effort important a été consacré au cours de cette thèse à la construction d'un corpus multimodal destiné à la reconnaissance des émotions. Notre corpus a la particularité d'utiliser à la fois des stimuli statiques (images) issus d'une base couramment utilisée dans le cadre de l'informatique affectif, et des stimuli dynamiques (vidéos) utilisés pour la première fois dans le cadre de l'élicitation d'émotions.

L'originalité de ce corpus réside dans la considération d'une dynamique de l'émotion ressentie. Nous proposons un nouveau procédé d'annotation, qui en plus d'une évaluation globale du stimulus, distingue les variations émotionnelles subies par le sujet. Il intègre ce caractère de la dynamique émotionnelle dans le cadre de l'évaluation de stimuli audiovisuels.

Notre corpus propose un référentiel d'annotation novateur afin de tenir compte de la subjectivité par le biais d'une phase de calibrage qui intervient en début d'expérience. C'est une nouvelle approche dans le cadre de la stratégie d'annotation de stimuli à contenu émotionnel, qui nous permet d'obtenir des informations subjectives sur la façon dont chaque participant appréhende l'annotation suivant chacun des axes en plus de guider l'annotateur sur un référentiel qui lui est propre tout au long de l'expérience. Les premières évaluations qui ont été menées sur ces informations de calibrage nous ont permis de mettre en évidence la pertinence de cette phase de calibrage qui réside dans un faible accord inter-annotateur.

Un effort a également été mené sur l'adaptation du système matériel et du protocole expérimental dans le but de fournir une synchronisation des données multimodales enregistrées afin de permettre un usage séparé ou combiné de l'ensemble des modalités.

Le système de reconnaissance des émotions

Nos travaux présentent de nouvelles approches pour la reconnaissance des émotions basée sur l'EEG. Les stratégies adoptées portent aussi bien sur le traitement des données, avec la proposition de nouveaux descripteurs, et sur les paramètres de l'architecture de classification. Le système de classification repose sur des machines à vecteurs support avec l'utilisation de descripteurs intégrés de façon précoce par un moyennage en fixant une fenêtre temporelle de référence.

Nous nous sommes penchés sur la représentation des données EEG sous la forme de descripteurs spectraux n'ayant pas été utilisés auparavant dans ce domaine. Nous avons comparé ces nouveaux descripteurs proposés avec les descripteurs habituellement employés dans le cadre de la reconnaissance des émotions basée sur l'EEG. Nous avons ainsi

pu mettre en évidence l'importance de l'information fréquentielle contenue dans les signaux EEG pour l'étude de tâches cognitives.

Une des originalités de la méthode de classification est de se positionner sur une configuration mono-canal. Nos travaux ont montré la possibilité de la mise en place d'une telle configuration pour les systèmes de reconnaissance des émotions basés sur l'EEG portée par des résultats performants.

5.2 Perspectives

Ces travaux ouvrent la voie à d'autres pistes d'études et des améliorations peuvent être apportées aux différents travaux menés. Nous distinguons ici les travaux pouvant être menés à court terme, qui s'orientent essentiellement vers des améliorations, et ceux pouvant être abordés à plus long terme et qui représentent plus de nouvelles pistes d'études.

5.2.1 Perspectives à court terme

Nos travaux sont orientés vers un système de classification binaire sur un des axes descripteur de l'émotion, à savoir la valence ou l'activation. Cependant, comme le prouvent les études théoriques, le phénomène émotionnel se révèle être plus complexe. Une extension de ce système à un modèle multi-classes permettrait une analyse plus fine dans la discrimination des catégories émotionnelles et présenterait un appui nécessaire à la réalisation d'un système de classification dynamique. Ce système pourrait être mis en place soit par une segmentation plus fine au sein des dimensions utilisées dans nos travaux soit par une catégorisation conjointe sur les deux axes.

Une des spécificités de l'étude de signaux d'EEG est de pouvoir s'inscrire dans plusieurs domaines : temporel, fréquentiel et spatial. Cela a été en partie exploité avec l'utilisation des CSP, qui permettent d'employer conjointement des informations spatiales et fréquentielles. Les travaux présentés dans ce manuscrit peuvent être consolidés en appliquant le modèle du système de classification avec de nouveaux descripteurs qui pourraient prendre en compte, par exemple, une dynamique temporelle et/ou spatiale de l'activité cérébrale.

Enfin, l'ajout de nouveaux descripteurs renforce le besoin de mettre en place une méthode de sélection de caractéristique adaptée. De part le grand nombre d'électrodes utilisées dans les systèmes d'acquisition d'EEG, les dimensions des descripteurs s'en trouvent aisément accrues. Il paraît alors plus adéquat d'être en mesure d'établir un assortiment pertinent de ces descripteurs.

Ces différentes perspectives, bien que complexes, méritent d'être traitées prioritairement dans la mesure où elles permettent d'ouvrir la voie à des perspectives de recherche plus larges.

5.2.2 Perspectives à long terme

L'émotion ressentie peut varier du fait du stimuli présenté au participant. Il est en effet difficile de maintenir un sujet dans un état émotionnel constant, cela résultant d'une

simple accoutumance au stimuli présenté. C'est la raison pour laquelle les stimuli présentés varient au bout de quelques secondes, par une succession d'images ou la diffusion d'une vidéo, bien qu'ils appartiennent à la même catégorie d'émotion. La dimension dynamique du phénomène émotionnel n'est cependant pas prise en compte dans nos travaux. Il serait donc intéressant de se pencher sur l'étude de la variation de l'état émotionnel expérimenté par le sujet.

Un des défis de l'informatique affective est l'utilisation de stimuli appropriés. Il serait donc bénéfique de réaliser l'annotation des stimuli vidéo utilisés dans notre corpus par un plus grand nombre de participants afin d'avoir une vérité terrain plus fiable. Une plateforme d'annotation pourrait être rendue disponible, sur une simple machine et dans un environnement exempt d'éventuelles distractions extérieures, afin de réaliser les tâches d'annotation.

La création d'un corpus multimodal permet d'élargir le domaine des études. Avec l'apport de l'information visuelle grâce à une vidéo du visage, il serait possible d'examiner les corrélations possibles entre les modalités vidéo et EEG. Il serait également possible de faire de même en se basant sur les autres signaux physiologiques enregistrés (ECG, EOG, EMG, conductance de la peau) grâce aux efforts portés sur la synchronisation des données.

Annexe A

Annexes

Sommaire

A.1 Les principes de l'électro-encéphalographie	I
A.1.1 Les bases neurophysiologiques, d'après Blinowska and Durka [2006]	I
A.1.2 Les rythmes EEG	II
A.2 Tableaux annexes	III
A.2.1 Extraits issus du corpus SAFE	III
A.2.2 Blocs d'images IAPS issus de ENTERFACE'06	V
A.3 Liste des acronymes	VIII

A.1 Les principes de l'électro-encéphalographie

Nous conseillons les papiers de [[Blinowska and Durka, 2006](#); [Bronzino, 2000](#); [Constant and Sabourdin, 2012](#); [Lux, 1981](#)] pour faire un tour d'horizon des principes de l'électro-encéphalographie.

A.1.1 Les bases neurophysiologiques, d'après [Blinowska and Durka \[2006\]](#)

Dans le cerveau on peut distinguer deux types de cellules :

- les *cellules nerveuses* ou neurones,
- et les *cellules gliales*.

Pour chacune d'entre elle, le potentiel au repos est d'environ -80 mV, l'intérieur des cellules étant négatif. La différence de potentiel à travers la membrane d'une cellule trouve son origine dans sa concentration en cations (K^+ , Na^+), en anions (Cl^-) et de larges anions organiques. Ca^{++} est moins important, mais joue tout de même un rôle de régulation important.

La différence de potentiel provient d'un transport actif des cations K^+ vers l'intérieur de la cellule et de Na^+ à l'extérieur.

L'activité électrique des neurones se manifeste par :

- les potentiels d'action qui apparaît lorsque l'excitation électrique dépasse un certain seuil,
- et les potentiels post-synaptiques (PSP) qui correspondent à un phénomène de sous-seuil.

La part de contribution des potentiels d'action dans le champ électrique qui est mesuré de manière extra-crânienne est négligeable. L'EEG provient de la somme de potentiels post-synaptiques générés de manière synchrone.

Une électrode capture l'activité de 10^8 neurones dont 1% sont synchrones. Leur contribution est 100 fois plus élevée que celles des neurones asynchrones.

A.1.2 Les rythmes EEG

- Delta : descripteur prédominant dans le sommeil profond. Ils ont habituellement une grande amplitude (75-200 μ V) et montrent une grande cohérence sur tout le scalp.
- Theta : peuvent apparaître dans le cas d'états émotionnels ou cognitifs.
- Alpha : prédominant dans l'éveil et prononcé dans les régions postérieures de la tête. On les observe le mieux les yeux fermés ou dans un état de relaxation.
 - ≠ mu : même bande de fréquence mais topographie et signification physiologique différentes (partie centrale de la tête, cortex moteur). Alpha est bloqué par l'attention et mu est bloqué par les fonctions motrices.
- Beta : caractérise les états de vigilance élevée et l'attention focalisée.
- Gamma : traitement de l'information(par exemple la reconnaissance de stimuli sensoriels) et le déclenchement des mouvements volontaires.

θ	4 – 8 Hz
α	8 – 12 Hz
β	12 – 30 Hz
γ	30 – 100 Hz

FIGURE A.1 – Bandes fréquentielles des principaux rythmes EEG.

A.2 Tableaux annexes

A.2.1 Extraits issus du corpus SAFE

ID	Film	Séquence
5	Shining	shining_c2s1
6	Mulholand Drive	Mulho_c28s1
7	Marathon man	marathon_c8s1
8	Hannibal	hannibal_c11s1
9	Dogville	dogville_c4s1
10	Dancer in the Dark	dancer_c5s2
17	Cape Fear	cape_c2s1
19	Speed	speed_c2s1
21	The Blair witch project	Blairwitch_c2s2
22	The Sixth Sense	6sens_c2s1
23	The silence of the Lambs	silence_c30s2
24	The Blair witch project	Blairwitch_c2s1
25	The silence of the Lambs	silence_c15s1
26	The silence of the Lambs	silence_c21s1
29	The silence of the Lambs	silence_c29s1
31	Seven	Seven_c27s1
32	Seven	Seven_c27s1
34	Schindler's List	schindler_c14s2
35	Schindler's List	schindler_c14s2
36	Schindler's List	schindler_c14s1
37	Schindler's List	schindler_c5s1
38	The Pianist	pianist_c4s4
39	The Pianist	pianist_c4s1
40	The Pianist	pianist_c4s1
41	Narc	narc_c14s3
42	Narc	narc_c13s2
43	Narc	narc_c12s4
44	The Fugitive	fugitif_c19s1
45	Hannibal	hannibal_c20s2
46	A Lesson Before Dying	homme_c4s1
47	Marathon Man	marathon_c32s1
48	Narc	narc_c5s1
49	The Pianist	pianist_c3s3
50	The silence of the Lambs	silence_c30s1
51	Dancer in the Dark	dancer_c8s2

52	Dancer in the Dark	dancer_c8s2
53	Goodfellas	affranchis_c16s1
54	The Pianist	pianist_c3s1
55	The Pianist	pianist_c3s1
56	The Sixth Sense	6sens_c2s2
57	The Sixth Sense	6sens_c2s2
58	The Sixth Sense	6sens_c9s1
59	Goodfellas	affranchis_c13s3
60	Goodfellas	affranchis_c19s1
61	The Blair witch project	Blairwitch_c5s1
62	Cape Fear	Cape_c7s1
63	Cape Fear	Cape_c16s1
64	Cape Fear	Cape_c17s1
65	Dancer in the Dark	dancer_c12s2
66	Dancer in the Dark	dancer_c17s1
67	Dogville	Dogville_c13s1
68	Dogville	Dogville_c19s2
69	Dogville	Dogville_c19s2
70	Dogville	Dogville_c19s2
71	Hannibal	hannibal_c4s1
72	Hannibal	hannibal_c20s2
73	Hannibal	hannibal_c29s1
74	Hannibal	hannibal_c29s1
75	A lesson before dying	homme_c4s1
76	Identity	identity_c2s1
77	Identity	identity_c2s4
78	Identity	identity_c11s1
79	Marathon man	marathon_c25s1
80	Marathon man	marathon_c28s1
81	Marathon man	marathon_c28s1
82	Marathon man	marathon_c33s1
83	Narc	narc_c5s1
84	Narc	narc_c8s1
85	Narc	narc_c11s2
86	Gangs of New York	gangs_c2s1
87	Hannibal	hannibal_c29s1
88	The Pianist	pianiste_c11s1
89	The Blair witch project	Blairwitch_c8s1
90	Dancer in the Dark	dancer_c3s3
91	Dancer in the Dark	dancer_c15s1

92	Dancer in the Dark	dancer_c17s1
93	Identity	identity_c2s4
94	Identity	identity_c11s1
95	Marathon	marathon_c32s1
96	The Pianist	pianiste_c11s1
97	The Pianist	pianist_c3s3
98	Schindler's List	schindler_c14s1
99	Schindler's List	schindler_c14s1
100	Shining	Shining_c34s1

TABLEAU A.1 – Séquences du corpus SAFE utilisées comme stimuli audio-visuels du corpus.

A.2.2 Blocs d'images IAPS issus de ENTERFACE'06

Session	Référence du bloc	Moyenne sur la valence	Ecart-type sur la valence	Moyenne sur l'activation	Ecart-type sur l'activation
1	1	4,98	0,29	2,80	0,34
	7	5,26	0,18	3,11	0,22
	16	4,88	0,38	3,09	0,64
	21	5,26	0,44	3,12	0,70
2	3	5,13	0,13	3,18	0,83
	6	5,21	0,49	2,89	0,51
	14	4,86	0,34	2,68	0,61
	15	5,20	0,36	3,59	0,34
	17	5,11	0,36	3,21	0,45
	18	5,00	0,57	2,92	0,71
	21	5,07	0,15	3,22	0,80
	26	5,13	0,63	3,51	0,32
3	3	5,00	0,30	3,02	0,52
	7	4,83	0,73	3,23	0,51
	9	5,21	0,53	3,15	0,52
	10	4,72	0,54	3,23	0,63
	13	4,92	0,37	2,86	0,16
	16	5,17	0,30	3,16	0,40
	20	5,05	0,41	2,87	0,66
	22	5,10	0,53	2,93	0,42
	23	4,97	0,43	3,10	0,53
	26	5,24	0,43	3,43	0,61

TABLEAU A.2 – Blocs *Neutre*

Session	Référence du bloc	Moyenne sur la valence	Ecart-type sur la valence	Moyenne sur l'activation	Ecart-type sur l'activation
1	2	7,63	0,43	5,49	0,35
	8	7,30	0,28	6,13	0,84
	20	7,12	0,28	5,79	0,64
	23	7,70	0,10	5,81	0,56
	28	7,34	0,26	5,83	0,71
2	4	7,29	0,24	5,50	0,26
	10	7,44	0,17	6,16	0,73
	12	7,28	0,38	6,01	0,50
	16	7,52	0,33	5,41	0,67
3	4	7,27	0,28	5,62	0,34
	5	7,77	0,20	5,88	0,68
	18	7,31	0,38	6,45	0,57
	27	7,39	0,53	5,54	0,20
	28	7,25	0,21	5,48	0,26

TABLEAU A.3 – Blocs *Excité positif*

Session	Référence du bloc	Moyenne sur la valence	Ecart-type sur la valence	Moyenne sur l'activation	Ecart-type sur l'activation
1	4	2,10	0,46	6,40	0,48
	9	2,04	0,44	5,62	0,42
	22	2,29	0,36	5,74	0,53
	29	1,96	0,21	6,45	0,54
2	1	2,33	0,55	6,12	0,73
	19	2,45	0,36	5,69	0,24
	22	2,44	0,44	5,83	0,69
	27	2,32	0,59	6,11	0,63
	28	2,05	0,40	5,81	0,31
3	2	2,43	0,19	6,02	0,59
	6	2,18	0,35	6,53	0,40
	17	2,34	0,27	6,01	0,64
	19	2,29	0,51	6,35	0,50
	29	2,35	0,33	5,98	0,31

TABLEAU A.4 – Blocs *Excité négatif*

Session	Référence de bloc	Image 1	Image 2	Image 3	Image 4	Image 5
1	1	7090	7006	7030	5530	9700
	2	1710	7270	8210	8350	2160
	4	3191	6243	9252	6550	3063
	7	7052	6150	7710	7179	5500
	8	8400	8170	8185	4641	8503
	9	3016	6830	2095	3301	3101
	16	7057	9210	7037	7031	5471
	20	2216	7260	7570	7501	8186
	21	5533	7183	5740	7110	7285
	22	3140	2751	9301	9910	9423
	23	8080	5600	5910	8420	8499
	28	4610	5460	8034	2352.1	8370
29	3110	9420	9405	3071	6313	
2	1	6570	3266	6540	6825	7359
	3	1616	7058	7002	7004	5510
	4	4599	8540	2346	5700	5623
	6	8311	7056	7020	7140	7025
	10	8380	8200	4640	2216	8030
	12	8496	8370	8340	8034	4623
	14	7055	7010	7040	7060	7057
	15	7710	2445	2446	7190	7034
	16	8400	2058	7330	1811	5660
	17	7179	7036	5390	7052	7044
	18	7700	1670	7009	7207	7175
	19	9925	3350	3160	9500	9400
	21	7950	7550	2840	2749	7043
	22	9560	3500	9340	9901	6200
	26	7242	2518	5120	7192	7595
	27	6210	6242	9410	3180	3550
28	3062	3005.1	3017	3225	9920	
3	2	3150	3220	2053	6370	6360
	3	5740	7050	7150	7180	7546
	4	2208	5270	5470	8496	8371
	5	8470	8080	5600	7330	8190
	6	6315	6821	3130	6260	6021
	7	7035	7161	7234	7495	2206
	9	7547	5471	7100	8465	5130
	10	7096	9360	7705	7590	9210
	13	7059	7038	5731	2221	7233
	16	2880	2890	7235	7285	6150
	17	9430	2703	9006	3550.1	6510
	18	8340	5621	8370	8186	5460
	19	3069	9433	6213	6300	9050
	20	7205	6570.2	7041	5520	7031
	22	2580	5534	2980	7110	7130
	23	7283	7224	7491	9700	7500
26	2745.1	5532	9401	7236	7490	
27	4640	4626	8531	7220	5833	
28	4623	8500	8502	7230	5660	
29	9911	3261	9620	9600	6838	

A.3 Liste des acronymes

BCI : Interface Cerveau-Machine (*Brain-Computer Interface*)
BP/BVP : Volume Sanguin Impulsionnel (*Blood Volume Pulse*)
CSP : *Common Spatial Pattern*
ECG : Electro-cardiographie
EDA : Activité électro-dermale (*Electro Dermal Activity*)
EEG : Electro-encéphalographie
EMG : Electromyogramme
EOG : Electro-oculographie
ESU : Unité de synchronisation externe (*External Sync Unit*)
GSR : Réflexe cutané galvanique (*Galvanic Skin Response*)
HMI : Interface Homme-Machine (*Human-machine Interface*)
IAPS : International Affective Picture System
IRM : Imagerie par résonance magnétique
LDA : Analyse Discriminante Linéaire (*Linear Discriminant Analysis*)
LTC : Code temporel linéaire (*Linear Timecode*)
MEG : Magnétoencéphalographie
PAD : Plaisir, Activation, Dominance
PET : Tomographie par émission de positons (*Positron Emission Tomography*)
SAM : Self-Assessment Manikin
SNA : Système nerveux autonome
SNC : Système nerveux central
SNP : Système nerveux périphérique
SNS : Système nerveux somatique
SVM : Machine à vecteurs de support (*Support Vector Machine*)
VERVE : Vanquishing fear and apathy through E-inclusion

Liste des figures

2.1	Le processus émotionnel selon le <i>Sens Commun</i>	13
2.2	Les principales théories des émotions	15
2.3	Le processus émotionnel dans la prise de décision : l'hypothèse des marqueurs somatiques formulée par Damasio [2008].	17
2.4	La métaphore de l'Iceberg de Freud appliquée au processus émotionnel. La force de flottabilité est liée à la masse de liquide déplacée par l'objet.	19
2.5	Schéma du cerveau.	20
2.6	Exemple d'enregistrements de signaux périphériques EOG, EMG et ECG sur une durée de 10 secondes.	25
2.7	Diagramme du système de classification.	32
2.8	Les trois modes de représentation des changements dans les composantes de l'émotion : l'inconscience, la conscience et la verbalisation, tiré de Scherer [2005a].	37
2.9	Le cône des émotions de Plutchik [Plutchik, 1984]. Les 8 émotions élémentaires sont placées par opposition au centre de la roue.	39
2.10	<i>Circumplex</i> de Russell and Barrett [1999], modifié à partir de la version de Feldman Barrett and Russell [1998]. Le disque interne représente un schéma de l'affect central (<i>core affect</i>), et le disque extérieur représente plusieurs émotions prototypiques liées à leurs caractéristiques affectives.	43
2.11	Exemple d'une représentation des axes de valence et d'activation à l'aide du Self-Assessment Manikin.	44
2.12	Le système d'annotation bi-dimensionnel de Feeltrace.	46
2.13	Le système d'annotation de GTrace.	46
2.14	Caractéristiques de trois types d'affects proposés par Scherer [2005a]. (VL : très faible, L : faible, M : Moyen, H : élevé, VH : très élevé) [Scherer, 2004].	48
2.15	Description du protocole ENTERFACE'06 [Savran et al., 2006].	52
2.16	Images issues de la base de données IAPS (de gauche à droite : classe "excité positif", classe "excité négatif", classe "calme").	52
2.17	Répartition des stimuli de la base ENTERFACE'06 dans le plan Valence/Activation [Xu and (Kostas) Plataniotis, 2012].	53

2.18 Répartition des stimuli de la base DEAP dans le plan Valence/Activation pour les quatre conditions (LALV, HALV, LAHV et HAHV) [Koelstra et al., 2012]. La couleur traduit l'appréciation (le rouge foncé est une faible appréciation), et la taille traduit la dominance, de petite à élevée.	54
3.1 Répartition des stimuli visuels dans l'espace Valence/Activation.	62
3.2 Description du degré d'imminence et de la gravité dans l'annotation des segments du corpus SAFE.	63
3.3 Répartition de l'ensemble des regroupements sur les dimensions "Imminence" et "Intensité".	64
3.4 Montage Expérimental pour l'enregistrement des données du corpus EMOGEE.	65
3.5 Emplacements des 20 électrodes EEG du casque B-Alert X-24.	66
3.6 Signaux EEG enregistrés par le système d'acquisition B-Alert X24 sur une durée de 5 secondes. Un artefact oculaire est présent dans cette portion de signaux (encadré rouge), de forte amplitude sur les électrodes frontales, il s'atténue à l'arrière du casque.	66
3.7 Extraits de Signaux EOG et ECG.	67
3.8 Position des électrodes EOG et EMG. Quatre électrodes sont utilisées pour enregistrer l'EOG et deux pour l'EMG (Zygomatique majeur).	68
3.9 Participant lors d'une session de l'expérience.	68
3.10 Plan de la configuration de la salle lors de l'expérience.	69
3.11 Protocole Expérimental du corpus EMOGEE.	70
3.12 Description d'un essai correspondant à la diffusion d'un stimulus.	71
3.13 Description d'un essai dans le protocole ENTERFACE'06. la diffusion des stimuli est suivie d'une phase d'auto-évaluation <i>statique</i> (voir Section 3.4.1).	72
3.14 Description du protocole pour une vidéo en plein écran. La diffusion du stimulus est suivie d'une phase d'auto-évaluation <i>dynamique</i> (voir Section 3.4.2)	72
3.15 Exemple d'annotation des 9 images de calibration (1-9) sur la dimension de valence.	73
3.16 Interface d'annotation globale	75
3.17 Interface d'annotation dynamique	76
4.1 Schéma de validation croisée inter-session. Les données de la phase d'apprentissage (i.e. un bloc) sont en vert et les données de la phase de test sont en rouge.	96
4.2 Schéma de validation croisée intra-session. Les données de la phase d'apprentissage (i.e. un bloc) sont en vert et les données de la phase de test sont en rouge.	96
4.3 Taux de bonnes reconnaissances obtenus pour chaque sujet en fonction de la configuration adoptée pour le système de classification.	101
A.1 Bandes fréquentielles des principaux rythmes EEG.	II

Liste des tableaux

2.1	Les principaux signaux périphériques utilisés dans le cadre de la reconnaissance de l'émotion en informatique affective.	23
2.2	Liste et description de publications sur la reconnaissance de l'émotion utilisant les signaux physiologiques périphériques.	26
2.3	Liste et description de publications sur la reconnaissance de l'émotion utilisant les signaux périphériques et l'EEG.	28
2.4	Différentes combinaisons de modalités.	30
2.5	Liste et description de publications sur la reconnaissance de l'émotion utilisant des systèmes multi-modaux.	31
2.6	Principaux résultats dans la littérature sur la reconnaissance des émotions basée sur l'EEG (PSD : Power Spectral Density, CSP : Common Spatial Pattern, AI : Asymmetry Index; SVM : Support Vector Machine, QDA : Quadratic Discriminant Analysis, KNN : K-Nearest Neighbor).	35
2.7	Les principales théories des émotions <i>basiques</i>	41
2.8	Les bases de données validées de stimuli à contenu émotionnel (Val. : Valence, Act. : Activation, Dom. : Dominance, App.-Av. : Approach-Avoidance).	49
2.9	Comparatif des bases de données existantes et accessibles	51
3.1	Références des blocs sélectionnés dans le protocole ENTERFACE'06 [Savran et al., 2006]	61
3.2	Moyenne (Écart-type) des images IAPS sélectionnées pour le protocole de calibrage	73
3.3	Degré d'accord en fonction des valeurs de Kappa.	78
3.4	Coefficient de la corrélation de Pearson entre les scores IAPS et l'auto-évaluation par session de chaque sujet du corpus EMOGEE sur les images de calibrage.	79
3.5	Coefficients de la corrélation de Pearson entre l'auto-évaluation de chaque session par sujet du corpus EMOGEE sur les images de calibrage.	80
3.6	Coefficients de corrélation de Pearson entre les scores IAPS et les auto-évaluations par sujet de la base eNTERFACE'06 [Xu and (Kostas) Plataniotis, 2012].	80
3.7	Coefficient de la corrélation de Pearson entre les scores IAPS et l'auto-évaluation des blocs eNTERFACE'06 par sujet du corpus EMOGEE.	81
3.8	Calcul de κ pour chaque paire d'annotateurs pour les évaluations sur l'axe de valence et l'axe d'activation des stimuli visuels.	81

3.9	Coefficients κ_F en fonction du nombre d'annotateurs pour l'annotation selon les dimensions de valence et d'activation des stimuli visuels.	82
3.12	Moyenne des coefficients κ pour l'annotation globale et dynamique des stimuli audio-visuels suivant les dimensions de valence et d'activation et leur regroupement.	82
3.10	Calcul de κ pour chaque paire d'annotateurs pour les évaluations sur l'axe de valence et l'axe d'activation des stimuli audio-visuels.	83
3.11	Coefficients κ_F en fonction du nombre d'annotateurs pour l'annotation globale selon les dimensions de valence et d'activation des stimuli audio-visuels.	83
3.13	Coefficients κ_F en fonction du nombre d'annotateurs pour l'annotation dynamique selon les dimensions de valence et d'activation des stimuli audio-visuels.	84
3.14	Moyenne des coefficients κ en fonction du type de stimuli annotés.	84
4.1	Liste des descripteurs proposés dans nos travaux. Ces descripteurs sont extraits pour chaque signal d'électrode EEG.	93
4.2	Caractéristiques extraites pour la reconnaissance des émotions basée sur l'EEG.	94
4.3	Résultats obtenus pour chaque sujet en fonction de l'utilisation de différentes caractéristiques. Les meilleurs scores de chaque sujet sont indiqués en gras.	97
4.4	Résultats obtenus avec les CSP pour 20 composantes. Les scores les plus importants sont surlignés en vert.	98
4.5	Résultats obtenus avec les combinaisons de CSP pour 20 composantes. Les scores les plus importants sont surlignés en vert.	98
4.6	Résultats obtenus avec les combinaisons de CSP pour 40 composantes. Les scores les plus importants sont surlignés en vert.	98
4.7	Caractéristiques extraites pour la reconnaissance des émotions basée sur l'EEG. Les caractéristiques que nous proposons sont en gras.	99
4.8	Taux de bonne reconnaissance (selon la dimension de valence) pour un sous-ensemble (le meilleur) des caractéristiques proposées : Moments (Mts), SF, SCE, et CSP pour une dimension de $D_{CSP} = 20$ pour chacun des 5 sujets (S1 à S2). Les meilleures précisions sont en caractère gras.	99
4.9	Meilleurs scores avec leurs caractéristiques associées pour chaque sujet pour une classification binaire suivant la dimension de Valence.	102
4.10	Meilleurs scores avec leurs caractéristiques associées pour chaque sujet pour une classification binaire suivant la dimension d'Activation.	102
A.1	Séquences du corpus SAFE utilisées comme stimuli audio-visuels du corpus.	V
A.2	Blocs <i>Neutre</i>	V
A.3	Blocs <i>Excité positif</i>	VI
A.4	Blocs <i>Excité négatif</i>	VI
A.5	Composition des blocs sélectionnés parmi ceux du corpus eINTERFACE'06	VII

Annexe B

Bibliographie

B.1 Publications de l'auteur

Anne-Claire Conneau, Slim Essid. "Assessment of new spectral features for eeg-based emotion recognition." IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2014.

Anne-Claire Conneau, Ayoub Hajlaoui, Mohamed Chetouani, Slim Essid, "EMOGEE : A new multimodal dataset for dynamic EEG-based emotion recognition with audio-visual stimuli." *in preparation*

Bibliographie

Ralph Adolphs. Recognizing Emotion From Facial Expressions : Psychological and Neurological Mechanisms. *Behavioral and Cognitive Neuroscience Reviews*, 1(1) :21–62, 2002. [25](#)

Foteini Agrafioti, Dimitris Hatzinakos, and Adam K. Anderson. ECG pattern analysis for emotion detection. *IEEE Transactions on Affective Computing*, 3(1) :102–115, 2012. [24](#), [67](#)

O. Alaoui-Ismaïli, O Robin, H Rada, A Dittmar, and E Vernet-Maury. Basic emotions evoked by odorants : Comparison between autonomic responses and self-evaluation. *Physiology and Behavior*, 62(4) :713–720, 1997. [49](#)

Omar Alzoubi, Rafael A. Calvo, and Ronald H. Stevens. Classification of EEG for Affect Recognition : An Adaptive Approach. In *AI 2009 : Advances in Artificial Intelligence*, pages 52–61. Springer Berlin Heidelberg, 2009. [35](#), [49](#)

Jenni Anttonen and Veikko Surakka. Emotions and heart rate while sitting on a chair. In *Proceedings CHI 2005*, pages 491–499, Portland, Oregon, USA, 2005. [24](#)

Magda B. Arnold. Emotion and personality. *New York Columbia University Press Aubergé V*, 1 :430, 1960. [18](#)

- Albert F. Ax. The physiological differentiation between fear and anger in humans. *Psychosomatic Medicine*, 15(5) :433–442, 1953. [14](#), [23](#)
- Jeremy N. Bailenson, Emmanuel D. Pontikakis, Iris B. Mauss, James J. Gross, Maria E. Jabon, Cendri A.C. Hutcherson, Clifford Nass, and Oliver John. Real-time classification of evoked emotions using facial feature tracking and physiological responses. *International Journal of Human Computer Studies*, 66(5) :303–317, may 2008. [30](#), [31](#)
- E.Y. Bann and J.J. Bryson. The Conceptualisation of Emotion Qualia : Semantic Clustering of Emotional Tweets. In *Proceedings of the 13th Neural Computation and Psychology Workshop (NCPW13)*, number July, pages 1–15, 2012. [41](#), [42](#)
- Philip Bard. A diencephalic mechanism for the expression of rage with special reference to the sympathetic nervous system. *American Journal of Physiology*, 1928. [14](#)
- Lisa Feldman Barrett. Discrete Emotions or Dimensions? The Role of Valence Focus and Arousal Focus. *Cognition & Emotion*, 12(4) :579–599, 1998. [42](#), [45](#)
- Teodiano Freire Bastos-Filho, Andre Ferreira, Anibal Cotrina Atencio, Sridhar Arjunan, and Dinesh Kumar. Evaluation of feature extraction techniques in emotional state recognition. In *IEEE Proceedings of 4th International Conference on Intelligent Human Computer Interaction (IHCI)*, pages 1–6, Kharagpur, India, 2012. [35](#)
- Anirban Basu and Anisha Halder. Facial expression and EEG signal based classification of emotion. *International Conference on Electronics, Communication and Instrumentation (ICECI)*, pages 1–4, 2014. [29](#)
- Thomas Baumgartner, Michaela Esslen, and Lutz Jäncke. From emotion perception to emotion experience : Emotions evoked by pictures and classical music. *International Journal of Psychophysiology*, 60(1) :34–43, apr 2006a. [49](#)
- Thomas Baumgartner, Kai Lutz, Conny F Schmidt, and Lutz Jäncke. The emotional power of music : how music enhances the feeling of affective pictures. *Brain Research*, 1075(1) : 151–64, feb 2006b. [49](#)
- Kent C. Berridge and Morten L. Kringelbach. Neuroscience of affect : brain mechanisms of pleasure and displeasure. *Current opinion in neurobiology*, 23(3) :294–303, jun 2013. [19](#)
- Kent C Berridge and Piotr Winkielman. What is an unconscious emotion?(The case for unconscious "liking"). *Cognition & Emotion*, 17(2) :181–211, 2003. [48](#)
- Benjamin Blankertz, Ryota Tomioka, Steven Lemm, Motoaki Kawanabe, and Klaus-Robert Müller. Optimizing spatial filters for robust EEG single-trial analysis. *IEEE Signal Processing Magazine*, 20 :1–12, 2008. [90](#)
- Katarzyna Blinowska and Piotr Durka. Electroencephalography (eeg). *Wiley Encyclopedia of Biomedical Engineering*, (5), 2006. [2](#), [I](#)

- Susana Bloch, Madeleine Lemeignan, and T. N. Aguilera. Specific respiratory patterns distinguish among human basic emotions. *International Journal of Psychophysiology*, 11(2) :141–154, 1991. [24](#)
- Frans A Boiten, Nico H Fridja, and Cornelis J.E. Wientjes. Emotions and respiratory patterns : Review and critical analysis. *International Journal of Psychophysiology*, 17(94) : 103–128, 1994. [24](#)
- Danny Oude Bos. EEG-based Emotion Recognition. *The Influence of Visual and Auditory Stimuli*, pages 1–17, 2006. [49](#)
- Margaret M. Bradley and Peter J. Lang. Measuring emotion : The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, 25(1) :49–59, mar 1994. [44](#), [74](#), [94](#)
- Margaret M. Bradley and Peter J. Lang. Affective Norms for English Words (ANEW) : Instruction Manual and Affective Ratings. *Psychology*, Technical(C-1), 1999. [38](#), [41](#), [49](#)
- Margaret M. Bradley and Peter J. Lang. The International Affective Digitized Sounds (2nd Edition ; IADS-2) : Affective Ratings of Sounds and Instruction Manual. Technical Report 2, 2007. [49](#)
- Margaret. BrBradley. Emotion and motivation. In *Handbook of psychophysiology*, chapter 22, pages 602–642. 2000. [24](#)
- Ed Joseph D Bronzino. Principles of Electroencephalography. *The Biomedical Engineering Handbook : Second Edition*, 2000. [I](#)
- Tobias Brosch and David Sander. Comment : The Appraising Brain : Towards a Neuro-Cognitive Model of Appraisal Processes in Emotion. *Emotion Review*, 5(2) :163–168, mar 2013. ISSN 1754-0739. [20](#)
- Jing Cai, Guangyuan Liu, and Min Hao. The Research on Emotion Recognition from ECG Signal. *2009 International Conference on Information Technology and Computer Science*, pages 497–500, jul 2009. [24](#)
- Walter B Cannon. The James-Lange theory of emotions : a critical examination and an alternative theory. *The American journal of psychology*, 39(1-4) :106–124, 1927. [14](#)
- George Caridakis and Ginevra Castellano. Multimodal emotion recognition from expressive faces, body gestures and speech. *Artificial intelligence and ...*, 247 :375–388, 2007. [31](#)
- Sandra Carvalho, Jorge Leite, Santiago Galdo-Álvarez, and Oscar F Gonçalves. The Emotional Movie Database (EMDB) : a self-report and psychophysiological study. *Applied psychophysiology and biofeedback*, 37(4) :279–94, dec 2012. [49](#)
- Ginevra Castellano, Santiago D Villalba, and Antonio Camurri. Recognising Human Emotions from Body Movement and Gesture Dynamics. *Affective Computing and Intelligent Interaction*, 4738 :71_82, 2007. [25](#)

- Ginevra Castellano, Loic Kessous, and George Caridakis. Emotion recognition through multiple modalities : Face, body gesture, speech. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 4868 LNCS :92–103, 2008. [29](#), [30](#)
- Guillaume Chanel. *Emotion assessment for affective computing based on brain and peripheral signals*. PhD thesis, Université Genève, 2009. [22](#)
- Guillaume Chanel, Julien Kronegg, Didier Grandjean, and Thierry Pun. Emotion assessment : Arousal evaluation using EEGs and peripheral physiological signals. *Multimedia Content Representation, Classification and Security*, pages 530–537, 2006. [27](#), [28](#), [30](#), [31](#), [34](#)
- Guillaume Chanel, K Ansari-Asl, and Thierry Pun. Valence-arousal evaluation using physiological signals in an emotion recall paradigm. *IEEE International Conference on Systems, Man and Cybernetics*, 41(22) :2662–2667, 2007. [26](#), [30](#), [34](#), [49](#)
- Chuan Yu Chang, Jeng Shiun Tsai, Chi Jane Wang, and Pau Choo Chung. Emotion recognition with consideration of facial expression and physiological signals. In *2009 IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology, CIBCB 2009 - Proceedings*, pages 278–283, 2009. ISBN 9781424427567. [30](#), [31](#)
- Denise Chen and Pamela Dalton. The effect of emotion and personality on olfactory perception. *Chemical senses*, 30(4) :345–51, may 2005. [49](#)
- L.S. Chen, H. Tao, T.S. Huang, T. Miyasato, and R. Nakatsu. Emotion recognition from audiovisual information. *IEEE Second Workshop on Multimedia Signal Processing*, pages 83–88, 1998. [31](#)
- Bo Cheng and Gy Liu. Emotion recognition from surface EMG signal using wavelet transform and neural network. In *Proceedings of The 2nd International Conference on Bioinformatics and Biomedical Engineering (ICBBE)*, pages 2–7, 2008. ISBN 9781424417483 (ISBN). [24](#), [67](#)
- Chloé Clavel. Fear-type emotions of the SAFE Corpus : annotation issues. *Proc. of LREC*, pages 1099–1104, 2006. [27](#), [29](#)
- Chloé Clavel, Ioana Vasilescu, Laurence Devillers, Gaël Richard, Thibaut Ehrette, and Célestin Sedogbo. The SAFE Corpus : illustrating extreme emotions in dynamic situations. *First International Workshop on Emotion : Corpora for Research on Emotion and Affect (International conference on Language Resources and Evaluation (LREC 2006))*, pages 76–79, 2006. [60](#)
- Chloé Clavel, Laurence Devillers, Gaël Richard, Ioana Vasilescu, and Thibaut Ehrette. Detection and analysis of abnormal situations through fear-type acoustic manifestations. In *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, volume 4, pages 21–24, 2007. ISBN 1424407281. [27](#)

- Chloé Clavel, Ioana Vasilescu, and Laurence Devillers. Fiction support for realistic portrayals of fear-type emotional manifestations. *Computer Speech and Language*, 25(1) : 63–83, jan 2011. [63](#)
- James A. Coan and John J B Allen, Allen. *Handbook of emotion elicitation and assessment*. 2007. [49](#)
- Jacob Cohen. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20(1) :37–46, 1960. [77](#)
- Jacob Cohen. Weighted kappa : Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological bulletin*, 70(4) :213, 1968. [78](#)
- Isabelle Constant and Nada Sabourdin. The EEG signal : A window on the cortical brain activity. *Pediatric Anesthesia*, 22(6) :539–552, jun 2012. [21](#), [I](#)
- Géraldine Coppin and David Sander. Contemporary Theories and Concepts in the Psychology of Emotions. *Emotion-oriented Systems*, pages 3–32, 2012. [18](#), [40](#)
- John L. Cotton. A review of research on Schachter’s theory of emotion and the misattribution of arousal. *European Journal of Social Psychology*, 11(4) :365–397, 1981. [16](#)
- Roddy Cowie, Ellen Douglas-cowie, Susie Savvidou, Edelle McMahon, Martin Sawey, and Marc Schröder. ‘FEELTRACE’ : An instrument for recording perceived emotion in real time. In *ISCA Workshop on Speech & Emotion*, pages 19–24, 2000. [45](#)
- Ian Daly, Asad Malik, Faustina Hwang, Etienne Roesch, James Weaver, Alexis Kirke, Duncan Williams, Eduardo Miranda, and Slawomir J. Nasuto. Neural correlates of emotional responses to music : An EEG study. *Neuroscience Letters*, 573 :52–57, jun 2014. [21](#), [49](#)
- Antonio Damasio. *Descartes’ error : Emotion, reason and the human brain*. Random House, 2008. [17](#), [IX](#)
- Antonio R. Damasio. The Somatic Marker Hypothesis and the Possible Functions of the Prefrontal Cortex. *Philosophical Transactions of the Royal Society B : Biological Sciences*, 351 :1413–1420, 1996. [17](#), [18](#), [48](#)
- Elise S. Dan-Glauser and Klaus R. Scherer. The Geneva affective picture database (GAPED) : a new 730-picture database focusing on valence and normative significance. *Behavior Research Methods*, 43(2) :468–477, jun 2011. [49](#)
- Charles Darwin. The expression of the emotions in man and animals. *The American Journal of the Medical Sciences*, 1872. [13](#), [19](#), [25](#), [67](#)
- Richard J. Davidson. Affect, cognition, and hemispheric specialization. In *Emotions, cognition, and behavior*, volume 1, pages 320–365. 1984. [20](#)

- Hallowell Davis, Truman Mast, Nobuo Yoshie, and Stanley Zerlin. The slow response of the human cortex to auditory stimuli : recovery process. *Electroencephalography and clinical neurophysiology*, 21(2) :105–113, 1966. [19](#), [21](#)
- Abhinav Dhall, Roland Goecke, Jyoti Joshi, Michael Wagner, and Tom Gedeon. Emotion recognition in the wild challenge 2013. *International Conference On Multimodal Interaction*, pages 509–516, 2013. [27](#)
- Sidney K D’mello and Jacqueline Kory. A review and meta-analysis of multimodal affect detection systems. *ACM Computing Surveys (CSUR)*, 47(3) :43, 2015. [29](#)
- Simon Dobrišek, Rok Gajšek, France Mihelič, Nikola Pavešić, and Vitomir Štruc. Towards efficient multi-modal emotion recognition. *International Journal of Advanced Robotic Systems*, 10 :1–10, 2013. [31](#)
- Ruoyu Du and Hyo Jong Lee. Power spectral performance analysis of EEG during emotional auditory experiment. In *ICALIP 2014 - 2014 International Conference on Audio, Language and Image Processing, Proceedings*, pages 64–68, 2015. [49](#)
- Shlomo Dubnov. Generalization of spectral flatness measure for non-gaussian linear processes. *Signal Processing Letters, IEEE*, (1), 2004. [92](#)
- Guillaume Benjamin Duchenne. Mécanisme de la physionomie humaine, ou analyse électro-physiologique de l’expression des passions, 1862. [67](#)
- Paul Ekman. Expression and the nature of emotion. *Approaches to emotion*, (1963) :319–344, 1984. [50](#)
- Paul Ekman. An argument for basic emotions. 6(3) :169–200, 1992. [20](#), [25](#), [38](#)
- Paul Ekman. Basic emotions. In *Handbook of cognition and emotion*, pages 45–60. John Wiley & Sons, 1999. [14](#), [25](#), [38](#), [41](#)
- Paul Ekman. *Darwin and facial expression : A century of research in review*. Ishk, 2006. [25](#)
- Paul Ekman and Wallace V. Friesen. Measuring facial movement. *Environmental Psychology and Nonverbal Behavior*, 1(1) :56–75, 1976. [20](#)
- Paul Ekman, Robert W. Levenson, and Wallace V. Friesen. Autonomic nervous system activity distinguishes among emotions. *Science*, 221(4616) :1208–1210, 1983. [23](#)
- Moataz El Ayadi, Mohamed S. Kamel, and Fakhri Karray. Survey on speech emotion recognition : Features, classification schemes, and databases. *Pattern Recognition*, 44(3) : 572–587, mar 2011. [27](#)
- P. C. Ellsworth and Klaus R. Scherer. Appraisal processes in emotion. In *Handbook of affective sciences*, chapter 29, pages 572–595. 2003. [47](#)

- Slim Essid. *Classification automatique des signaux audio-frequences : reconnaissance des instruments de musique*. PhD thesis, 2005. [91](#)
- Beat Fasel and Juergen Luetttin. Automatic facial expression analysis : a survey. *Pattern Recognition*, 36(1) :259–275, 2003. [27](#)
- Lisa Feldman Barrett and James A Russell. Independence and Bipolarity in the Structure of Current Affect. *Journal of Personality and Social Psychology*, 74(4) :967–984, 1998. [43](#), [IX](#)
- Antoinette Feleky. The influence of the emotions on respiration. *Journal of Experimental Psychology*, 1916. [24](#)
- Gereon R. Fink, Richard S.J. Frackowiak, Uwe Pietrzyk, and Richard E. Passingham. Multiple nonprimary motor areas in the human cortex. *Journal of neurophysiology*, 77(4) :2164–74, 1997. [21](#)
- Johnny R.J. Fontaine, Klaus R. Scherer, Etienne B. Roesch, and Phoebe C. Ellsworth. The World of Emotions Is Not Two-Dimensional. *Psychological Science*, 18(12) :1050–1057, 2007. [47](#)
- Itzhak Fried, Catherine Mateer, George Ojemann, Richard Wohns, and Paul Fedio. Organization of visuospatial functions in human cortex. Evidence from electrical stimulation. *Brain : a journal of neurology*, 105(Pt 2) :349–71, 1982. [21](#)
- Nico H. Frijda. The emotions : Studies in emotion & social interaction. *Paris : Maison de Sciences de l'Homme*, 1986. [18](#)
- Nico H. Frijda. The Laws of Emotion. *American Psychologist*, 43(5) :349–358, 1988. [18](#)
- Dario Galati and Barbara Sini. Les structures sémantiques du lexique français des émotions. *Les émotions dans les interactions*, pages 75–87, 2000. [42](#)
- Antje B.M. Gerdes, Matthias J. Wieser, and Georg W. Alpers. Emotional pictures and sounds : a review of multimodal interactions of emotion cues in multiple domains. *Frontiers in psychology*, 5(December) :1351, jan 2014. [49](#)
- Astrid Gerrards-Hesse, Kordelia Spies, and Friedrich W. Hesse. Experimental inductions of emotional states and their effectiveness : A review. *British Journal of Psychology*, 85 : 55–78, 1994. [49](#)
- Olivier Gillet and Gaël Richard. Automatic transcription of drum loops. *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*, 4 :2–5, 2004. [91](#)
- Donald Glowinski, Antonio Camurri, Gualtiero Volpe, Nede Dael, and Klaus Scherer. Technique for automatic emotion recognition by body gesture analysis. In *2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pages 1608–1613, 2008. [25](#)
- Vannina Goossens. Les noms de sentiment. *Lidil*, 32, 2005. [42](#)

- James J. Gross and Robert W. Levenson. *Emotion Elicitation Using Films*, volume 9, pages 87–108. 1995. [49](#)
- Hatice Gunes and Maja Pantic. Automatic measurement of affect in dimensional and continuous spaces : why, what, and how?*. In *Proceedings of the 7th International Conference on Methods and Techniques in Behavioral Research (MB '10)*, pages 1–5, 2010. [47](#)
- Hatice Gunes and Massimo Piccardi. Bimodal face and body gesture database for automatic analysis of human nonverbal affective behavior. In *Proceedings - International Conference on Pattern Recognition*, volume 1, pages 1148–1153, 2006. [30](#)
- Hatice Gunes and Massimo Piccardi. Bi-modal emotion recognition from expressive face and body gestures. *Journal of Network and Computer Applications*, 30(4) :1334–1345, 2007. [29](#)
- Hatice Gunes, Massimo Piccardi, and Maja Pantic. From the Lab to the Real World : Affect Recognition Using Multiple Cues and Modalities. In *Affective Computing : Focus on Emotion Expression, Synthesis, and Recognition*, chapter 10, pages 185–218. 2008. [13](#)
- Shilpi Gupta, Anu Mehra, and Namrata Gandhi. Recognizing Emotions using Speech and Blood Volume Pulse. *Proc. Int. Conf. on Computational Intelligence and Information Technology*, pages 399–402, 2012. [24](#)
- Andreas Haag, Silke Goronzy, Peter Schaich, and Jason Williams. Emotion recognition using bio-sensors : First steps towards an automatic system. *Affective dialogue systems*, 2004a. [26](#)
- Andreas Haag, Silke Goronzy, Peter Schaich, and Jason Williams. Emotion Recognition Using Bio-sensors : First Steps towards an Automatic System. *Affective dialogue systems*, i :36–48, 2004b. [67](#)
- M. Haak, S. Bos, S. Panic, and Leon J. M. Rothkrantz. Detecting Stress Using Eye Blinks And Brain Activity From EEG Signals. In *Proceeding of the 1st driver car interaction and interface (DCII 2008)*, volume 10th Inter, pages 75–82, 2009. [24](#), [30](#)
- Dirk Hagemann, Shari R. Waldstein, and Julian F. Thayer. Central and autonomic nervous system integration in emotion. *Brain and Cognition*, 52(1) :79–87, jun 2003. [29](#)
- Samuel Haq and Philip J.B. Jackson. Multimodal emotion recognition. In *Machine audition : principles, algorithms and systems*, IGI Global, Hershey, pages 398–423. 2010. [29](#)
- William J. Havlena and Morris B. Holbrook. The varieties of consumption experience : comparing two typologies of emotion in consumer behavior. *Journal of Consumer Research*, 13(3) :394–404, 1986. [44](#)
- Alicia Heraz, Ryad Razaki, and C Frasson. Using machine learning to predict learner emotional state from brainwaves. *2007 Seventh IEEE International Conference on Advanced Learning Technologies (ICALT 2007)*, (Table 1) :853–857, 2007. [28](#)

- Ursula Hess and Pascal Thibault. Darwin and emotion expression. *The American psychologist*, 64(2) :120–128, 2009. [14](#)
- Bo Hjorth. EEG analysis based on time domain properties. *Electroencephalography and Clinical Neurophysiology*, 29(3) :306–310, 1970. [34](#)
- Robert Horlings, Dragos Datcu, and L.J.M. Rothkrantz. Emotion recognition using brain activity. In *Proceedings of the 9th International Conference on Computer Systems and Technologies and Workshop for PhD Students in Computing*, page 6, 2008. [6](#), [28](#), [34](#), [49](#)
- Eva Hudlicka and Hatice Gunes. Benefits and Limitations of Continuous Representations of Emotions in Affective Computing : Introduction to the Special Issue. *International Journal of Synthetic Emotions*, 3(1) :i–vi, 2012. [47](#)
- J. McV. Hunt, Wakeman Marie-Louise Cole, and Eva E.S. Reis. Situational Cues Distinguishing Anger, Fear, and Sorrow. *The American Journal of Psychology*, 71(1) :136–151, 1958. [16](#)
- Carroll E. Izard. Basic emotions, relations among emotions, and emotion-cognition relations. *Psychological Review*, 99(3) :561–565, 1992. [41](#)
- Alejandro Jaimes and Nicu Sebe. Multimodal human–computer interaction : A survey. *Computer Vision and Image Understanding*, 108(1-2) :116–134, 2007. [29](#)
- William James. What is an emotion? *Mind*, 9 :188–205, 1884. [14](#), [23](#)
- Eun-hye Jang, Byoung-Jun Park, Sang-Hyeob Kim, Myoung Chung, Mi-Sook Park, and Jin-Hun Sohn. Classification of Human Emotions from Physiological signals using Machine Learning Algorithms. In *The Sixth International Conference on Advances in Computer-Human Interactions (ACHI 2013)*, pages 395–400, 2013. [25](#), [67](#)
- Noppadon Jatupaiboon, Setha Pan-ngum, and Pasin Israsena. Emotion classification using minimal EEG channels and frequency bands. *The 2013 10th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, pages 21–24, may 2013a. [49](#)
- Noppadon Jatupaiboon, Setha Pan-ngum, and Pasin Israsena. Real-time EEG-based happiness detection system. *The Scientific World Journal*, 2013 :618649, 2013b. [7](#)
- Robert Jenke, Angelika Peer, and Martin Buss. Effect-Size-Based Electrode and Feature Selection for Emotion Recognition from EEG. 2013. [34](#), [35](#), [97](#)
- Xiang Jie, Rui Cao, and Li Li. Emotion recognition based on the sample entropy of EEG. *Bio-medical materials and engineering*, 24(1) :1185–92, jan 2014. [35](#)
- Cyril Joder, Slim Essid, and Gaël Richard. Temporal integration for audio classification with application to musical instrument classification. *IEEE Transactions on Audio, Speech and Language Processing*, 17(1) :174–186, 2009. [36](#)

- P. N. Johnson-Laird and Keith Oatley. The language of emotions : An analysis of a semantic field. *Cognition & Emotion*, 3(2) :81–123, 1989. [41](#), [42](#)
- James D. Johnston. Transform coding of audio signals using perceptual noise criteria. *IEEE Journal on Selected Areas in Communications*, 6(2) :314–323, 1988. [92](#)
- Kari Kallinen. Emotion related psychophysiological responses to listening to music with eyes-open versus eyes-closed : Electrodermal (EDA), Electrocardiac (ECG), And Electromyographic (EMG) Measures. In *Proceedings of the 8th International Conference on Music Perception & Cognition (ICMPC8)*, pages 299–301, 2004. [24](#)
- Sabine Kastner and Leslie G. Ungerleider. Mechanisms of Visual Attention in the Human Cortex. *Annual Review of Neuroscience*, 23 :315–341, 2000. [21](#)
- Benjamin Kedem and Michael Frey. Time Series Analysis by Higher Order Crossings. *IEEE press Piscataway*, may 1994. [34](#)
- Dacher Keltner, Paul Ekman, Gian C. Gonzaga, and Jennifer Beer. Facial expression of emotion. *Oxford University Press*, 2003. [27](#)
- Loic Kessous, Ginevra Castellano, and George Caridakis. Multimodal emotion recognition in speech-based interaction using facial expression, body gesture and acoustic analysis. *Journal on Multimodal User Interfaces*, 3(1) :33–48, dec 2010. [29](#), [30](#)
- Stéphanie Khalifa, Isabelle Peretz, Jean-Pierre Blondin, and Manon Robert. Event-related skin conductance responses to musical emotions in humans. *Neuroscience Letters*, 328 (2) :145–149, 2002. [24](#)
- Zahra Khalili. Emotion recognition system using brain and peripheral signals : using correlation dimension to improve the results of EEG. *Proceedings of International Joint Conference on Neural Networks (IJCNN 2009)*., pages 1571–1575, 2009. [28](#), [34](#), [35](#)
- Zahra Khalili and M.H. Moradi. Emotion detection using brain and peripheral signals. In *2008 Cairo International Biomedical Engineering Conference (CIBEC'08)*, pages 2006–2009, 2008. [22](#)
- Masood Mehmood Khan, Michael Ingleby, and Robert D. Ward. Automated Facial Expression Classification and affect interpretation using infrared measurement of facial skin temperature variations. *ACM Transactions on Autonomous and Adaptive Systems*, 1(1) : 91–113, 2006. [24](#)
- Jonghwa Kim and Elisabeth Andre. Emotion recognition using physiological and speech signal in short-term observation. In *Perception and Interactive Technologies*, number 1, pages 53–64. 2006. [29](#), [30](#), [31](#)
- Jonghwa Kim and Elisabeth André. Emotion recognition based on physiological changes in music listening. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(12) : 2067–2083, 2008. [26](#), [49](#), [67](#)

- Kyung Hwan Kim, S. W. Bang, and S. R. Kim. Emotion recognition system using short-term monitoring of physiological signals. *Medical and Biological Engineering and Computing*, 42(3) :419–427, 2004. [26](#)
- Youngmoo E. Kim, Erik M. Schmidt, Raymond Migneco, Brandon G. Morton, Patrick Richardson, Jeffrey Scott, Jacquelin A. Speck, and Douglas Turnbull. Music Emotion Recognition : a State of the Art Review. *Information Retrieval*, (Ismir) :255–266, 2010. [27](#)
- Paul R. Kleinginna and Anne M. Kleinginna. A categorized list of emotion definitions, with suggestions for a consensual definition. *Motivation and Emotion*, 5(4) :345–379, 1981. [18](#)
- Andrea Kleinsmith and Nadia Bianchi-Berthouze. Affective body expression perception and recognition : A survey. *IEEE Transactions on Affective Computing*, 4(1) :15–33, 2013. [25](#)
- Stefan Koelsch. Towards a neural basis of music-evoked emotions. *Trends in Cognitive Sciences*, 14(3) :131–137, mar 2010. [48](#)
- Sander Koelstra, Ashkan Yazdani, Mohammad Soleymani, and C Mühl. Single trial classification of EEG and peripheral physiological signals for recognition of emotions induced by music videos. *Brain informatics*, 6334 :89–100, 2010a. [90](#)
- Sander Koelstra, Ashkan Yazdani, Mohammad Soleymani, Christian Mühl, Jonk-See Lee, Anton Nijholt, Thierry Pun, Touradj Ebrahimi, and Ioannis Patras. Single Trial Classification of EEG and Peripheral Physiological Signals for Recognition of Emotions Induced by Music Videos. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 6334 LNAI :89–100, 2010b. [30](#), [35](#)
- Sander Koelstra, Christian Mühl, Mohammad Soleymani, Jong-seok Lee, Ashkan Yazdani, Touradj Ebrahimi, Thierry Pun, Anton Nijholt, and Ioannis (Yiannis) Patras. Deap : A database for emotion analysis using physiological signals. *IEEE Transactions on Affective Computing*, 3(1) :18–31, 2012. [1](#), [6](#), [12](#), [32](#), [44](#), [49](#), [50](#), [53](#), [54](#), [59](#), [X](#)
- Sylvia D. Kreibig. Autonomic nervous system activity in emotion : a review. *Biological psychology*, 84(3) :394–421, jul 2010. [23](#), [24](#)
- Jochen Krimphoff. Analyse acoustique et perception du timbre [Acoustic analysis and perception of timbre](Unpublished master’s thesis). *Université du Maine, Le Mans, France*, 1993. [92](#)
- Jochen Krimphoff, S. McAdams, and S. Winsberg. Caractérisation du timbre de sons complexes : {II}. Analyses acoustiques et quantification psychophysique. *Journal de Physique*, 4 :625–628, 1994. [93](#)
- Ludmila I Kuncheva, Thomas Christy, Iestyn Pierce, and Sa’ad P. Mansoor. Multi-modal Biometric Emotion Recognition Using Classifier Ensembles. In *Modern Approaches in Applied Intelligence*, number 6703, pages 317–326. 2011. [26](#), [30](#), [31](#), [49](#)

- James D. Laird. The real role of facial response in the experience of emotion : A reply to Tourangeau and Ellsworth, and others. *Journal of Personality and Social Psychology*, 47 (4) :909–917, 1984. [49](#)
- James D. Laird and Sarah Strout. Emotional behaviors as Emotional Stimuli. *Handbook of Emotion Elicitation and Assessment*, pages 54–64, 2007. [49](#)
- John A. Lambie and Anthony J. Marcel. Consciousness and the varieties of emotion experience : a theoretical framework. *Psychological review*, 109(2) :219–259, 2002. [48](#)
- J. Richard Landis and Gary G. Koch. The measurement of observer agreement for categorical data. *Biometrics*, 33(1) :159–174, 1977. [77](#)
- Peter J. Lang. The cognitive psychophysiology of emotion : Fear and anxiety. *Anxiety and the anxiety disorders.*, 1985. [45](#)
- Peter J. Lang, Mark Greenwald, Margaret M. Bradley, and Alfons O. Hamm. Looking at pictures : affective, facial, visceral, and behavioral reactions. *Psychophysiology*, 30(3) :261–273, 1993. [49](#)
- Peter J. Lang, Margaret M. Bradley, and Bruce N. Cuthbert. International Affective Picture System (IAPS) : Technical Manual and Affective Ratings. *NIMH Center for the Study of Emotion and Attention*, pages 39–58, 1997. [49](#), [50](#)
- Peter J. Lang, Margaret M. Bradley, and Bruce N. Cuthbert. International affective picture system (IAPS) : Affective ratings of pictures and instruction manual. *Technical Report A-8*, 2008. [60](#), [94](#)
- Joseph LeDoux. The amygdala. *Current Biology*, 17(20) :868–874, 2007. [19](#)
- Joseph E. LeDoux. Emotion circuits in the brain. *Annual review of neuroscience*, 23 :155–84, jan 2000. [19](#), [20](#)
- Nicolas Leveau, Sandra Jhean-Larose, and Guy Denhière. EMOVAL : évaluation automatique de la valence et de l'activation émotionnelles des textes à l'aide d'une méta-norme de 5656 mots-racines. *Psychologie française*, 56(4) :209–221, 2011. [42](#)
- Robert W. Levenson. Emotion and the autonomic nervous system : a prospectus for research on autonomic specificity, 1988. [24](#), [50](#)
- Robert W. Levenson. Blood, Sweat, and Fears : The Autonomic Architecture of Emotion. *Annals of the New York Academy of Sciences*, 1000(1) :348–366, jan 2003. [24](#)
- Robert W. Levenson. The Autonomic Nervous System and emotion. *Emotion Review*, 6(2) : 100–112, mar 2014. [23](#), [24](#)
- Michael Lewis. The Self-Conscious Emotions. *Emotions*, pages 1–5, 2011. [48](#)

- Lan Li and Ji-hua Chen. Emotion recognition using physiological signals. *Advances in Artificial Reality and Tele-Existence*, pages 437–446, 2006. [26](#)
- Mu Li and Bao Liang Lu. Emotion classification based on gamma-band EEG. In *Proceedings of the 31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society : Engineering the Future of Biomedicine, EMBC 2009*, pages 1323–1326, 2009. [28](#), [34](#), [35](#), [90](#), [94](#), [98](#), [99](#)
- Yuan-Pin Lin, Chi-Hong Wang, Tien-Lin Wu, Shyh-Kang Jeng, and Jyh-Horng Chen. Support vector machine for EEG signal classification during listening to emotional music. In *Proceedings of the 2008 IEEE 10th Workshop on Multimedia Signal Processing (MMSP 2008)*, pages 127–130, 2008. [28](#), [49](#)
- Yuan-Pin Lin, Chi-Hong Wang, Tzyy-Ping Jung, Tien-Lin Wu, Shyh-Kang Jeng, Jeng-Ren Duann, and Jyh-Horng Chen. EEG-based emotion recognition in music listening. *IEEE Transactions on Biomedical Engineering*, 57(7) :1798–1806, 2010. [49](#)
- Shuang Liu, Jiayuan Meng, Di Zhang, Jiajia Yang, Xin Zhao, Feng He, Hongzhi Qi, and Dong Ming. Emotion recognition based on EEG changes in movie viewing. In *International IEEE/EMBS Conference on Neural Engineering, NER (2015)*, pages 1036–1039, 2015. [35](#), [49](#)
- H D Lux. Neurophysiological Basis of the EEG and DC Potentials. In *Electroencephalography : Basic Principles, Clinical Applications, and Related Fields*, pages 24–28. 1981. [1](#)
- Chang-wei Ma and Guang-yuan Liu. Feature Extraction, Feature Selection and Classification from Electrocardiography to Emotions. In *Proceedings of the 2009 International Conference on Computational Intelligence and Natural Computing, CINC 2009 (2009)*, pages 190–193. Ieee, jun 2009. [24](#)
- Choubeila Maaoui and Alain Pruski. Emotion recognition through physiological signals for human-machine communication. In *Cutting Edge Robotics*, pages 317–333. INTECH Open Access Publisher, 2010. [26](#)
- Paul D. MacLean. Psychosomatic disease and the visceral brain. *Psychosomatic Medicine*, XI(6) :338–353, 1949. [20](#)
- John Makhoul. Linear prediction : A tutorial review. *Proceedings of the IEEE*, 63(4) :561–580, 1975. [93](#)
- Artur Marchewka, Lukasz Zurawski, Katarzyna Jednoróg, and Anna Grabowska. The Nencki Affective Picture System (NAPS) : Introduction to a novel, standardized, wide-range, high-quality, realistic picture database. *Behavior Research Methods*, 46(2) :596–610, aug 2014. [49](#)
- Soroosh Mariooryad and Carlos Busso. Analysis and Compensation of the Reaction Lag of Evaluators in Continuous Emotional Annotations. In *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*, pages 85–90. Ieee, sep 2013. [47](#)

- Olivier Martin, Irene Kotsia, Benoit Macq, and Ioannis Pitas. The eNTERFACE'05 audio-visual emotion database. In *International Conference on Data Engineering Workshops*, number 1, pages 8–15, 2006. [31](#)
- Douglas S. Massey. A brief history of human society : The origin and role of emotion in social life. *American Sociological Review*, 67 :1–29, 2002. [25](#)
- Richard A. Mcfarland. Relationship of Skin Temperature Changes to the Emotions Accompanying Music. *Biofeedback and Self-Regulation*, 10(3) :255–267, 1985. [24](#)
- Albert Mehrabian. Comparison of the PAD and PANAS as Models for Describing Emotions and for Differentiating Anxiety from Depression. *Journal of Psychopathology and Behavioral Assessment*, 19(4) :331–357, 1997. [44](#)
- Albert Mehrabian and James A. Russell. *An approach to environmental psychology*. The MIT Press, 1974. [40](#), [42](#), [44](#), [74](#)
- Angeliki Metallinou and Shrikanth Narayanan. Annotation and processing of continuous emotional attributes : Challenges and opportunities. In *2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition, FG 2013*, number EmoSPACE, 2013. [47](#)
- Jon D. Morris. Observations : SAM : The Self-Assessment Manikin - An Efficient Cross-Cultural Measurement of Emotional Response. *Journal of Advertising Research*, 35(6) : 63–68, 1995. [45](#)
- Christian Mühl, Egon L. Van Den Broek, Anne Marie Brouwer, Femke Nijboer, Nelleke Van Wouwe, and Dirk Heylen. Multi-modal affect induction for affective brain-computer interfaces. In *Affective Computing and Intelligent Interaction*, pages 235–245. Springer Berlin Heidelberg, 2011. [49](#)
- Fionnuala C Murphy, Ian Nimmo-Smith, and Andrew D Lawrence. Functional neuroanatomy of emotions : a meta-analysis. *Cognitive, affective and behavioral neuroscience*, 3 (3) :207–33, sep 2003. [19](#), [20](#)
- Arturo Nakasone, Helmut Prendinger, and Mitsuru Ishizuka. Emotion Recognition from Electromyography and Skin Conductance. *The 5th International Workshop on Biosignal Interpretation*, pages 219–222, 2005. [24](#)
- Jérémie Nicolle, Vincent Rapp, Kévin Bailly, Lionel Prevost, and Mohamed Chetouani. Robust continuous prediction of human emotions using multiscale dynamic cues. In *Proceedings of the 14th ACM international conference on Multimodal interaction*, pages 501–508, 2012. [47](#)
- Dan Nie, XW Wang, and LC Shi. EEG-based emotion recognition during watching movies. *Neural Engineering (NER)*,, pages 667–670, 2011. [28](#)

- Tin Lay Nwe, Say Wei Foo, and Liyanage C. De Silva. Speech emotion recognition using hidden Markov models. *Speech Communication*, 41(4) :603–623, nov 2003. [27](#)
- Pierre-Yves Oudeyer. The production and recognition of emotions in speech : Features and algorithms. *International Journal of Human Computer Studies*, 59(1-2) :157–183, jul 2003. [27](#)
- Marco Paleari, Ryad Chellali, and Benoit Huet. Features for multimodal emotion recognition : An extensive study. In *2010 IEEE Conference on Cybernetics and Intelligent Systems (CIS)*, pages 90–95, 2010. [29](#), [31](#)
- M. Carmen Pastor, Margaret M. Bradley, Andreas Löw, Francesco Versace, Javier Moltó, and Peter J. Lang. Affective picture perception : emotion, context, and the late positive potential. *Brain research*, 1189 :145–51, 2008. [48](#)
- Silke Paulmann and Marc D. Pell. Is there an advantage for recognizing multi-modal emotional stimuli? *Motivation and Emotion*, 35(2) :192–201, apr 2011. [49](#)
- Geoffroy Peeters. A large set of audio features for sound description (similarity and classification) in the CUIDADO project. Technical report, IRCAM, 2004. [91](#)
- Geoffroy Peeters, Bruno L Giordano, Patrick Susini, Nicolas Misdariis, and Stephen McAdams. The Timbre Toolbox : extracting audio descriptors from musical signals. *The Journal of the Acoustical Society of America*, 130(5) :2902–16, nov 2011. [92](#)
- Catherine Pelachaud. *Emotion-Oriented Systems*. John Wiley and Sons, 2013. [22](#)
- Luiz Pessoa and Ralph Adolphs. Emotion processing and the amygdala : from a 'low road' to 'many roads' of evaluating biological significance. *Nature reviews neuroscience*, 11(11) : 773–783, 2010. [19](#)
- Tomas Pfister, Xiaobai Li, Guoying Zhao, and Matti Pietikainen. Recognizing Spontaneous Facial Micro-Expressions. In *IEEE International Conference on Computer Vision*, pages 1449–1456, 2011. [25](#)
- Pierre Philippot. Inducing and assessing differentiated emotion-feeling states in the laboratory. *Cognition & Emotion*, 1993. [48](#), [49](#)
- Rosalind W. Picard, Elias Vyzas, and Jennifer Healey. Toward machine emotional intelligence : Analysis of affective physiological state. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(10) :1175–1191, 2001. [26](#), [34](#), [89](#), [94](#), [99](#)
- Annie Piolat and Rachid Bannour. EMOTAIX : Un scénario de Tropes pour l'identification automatisée du lexique émotionnel et affectif. *Annee Psychologique*, 109(4) :655–698, 2009. [38](#)
- Robert Plutchik. Emotions : A general psychoevolutionary theory. *Approaches to emotion*, 1984 :197–219, 1984. [38](#), [39](#), [41](#), [IX](#)

- Senya Polikovsky, Yoshinari Kameda, and Yuichi Ohta. Facial micro-expressions recognition using high speed camera and 3D-gradient descriptor. In *3rd International Conference on Imaging for Crime Detection and Prevention (ICDP 2009)*, pages 1–6, 2009. 25
- Jonathan Posner, James A. Russell, and Bradley S. Peterson. The circumplex model of affect : an integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and Psychopathology*, 17(3) :715–734, 2005. 44
- Pierre Rainville, Antoine Bechara, Nasir Naqvi, and Antonio R. Damasio. Basic emotions are associated with distinct patterns of cardiorespiratory activity. *International Journal of Psychophysiology*, 61(1) :5–18, 2006. 24
- Srinivasan Ramakrishnan. *Speech Enhancement, Modeling And Recognition – Algorithms And Applications*. InTech, 2012. 27
- Pramila Rani, Changchun Liu, Nilanjan Sarkar, and Eric Vanman. An empirical study of machine learning techniques for affect recognition in human-robot interaction. *Pattern Analysis and Applications*, 9(1) :58–69, apr 2006. 24, 26
- Rainer Reisenzein. The Schachter Theory of Emotion : Two Decades Later. *Psychological Bulletin*, 94(2) :239–264, 1983. 16
- Bernard Rimé and Klaus R. Scherer. *Les émotions, textes de base en psychologie*, 1989. 14
- Sara E. Rimm-Kaufman and Jerome Kagan. The psychological significance of changes in skin temperature. *Motivation and Emotion*, 20(1) :63–78, 1996. 24
- Jonathan Rottenberg, Rebecca D. Ray, and James J. Gross. Emotion Elicitation Using Films. In *Handbook of Emotion Elicitation and Assessment*, chapter 1, pages 9–28. 2007. 49, 54
- James A. Russell. A circumplex model of affect. *Journal of Personality & Social Psychology*, 39(6) :1161–1178, 1980. 41, 42, 74
- James A. Russell and Lisa Feldman Barrett. Core affect, prototypical emotional episodes, and other things called emotion : dissecting the elephant. *Journal of personality and social psychology*, 76(5) :805–819, 1999. 43, 44, IX
- James A. Russell and Albert Mehrabian. Evidence for a three-factor theory of emotions. *Journal of Research in Personality*, 11(3) :273–294, 1977. 40
- James A. Russell and Geraldine Pratt. A description of the affective quality attributed to environments. *Journal of Personality and Social Psychology*, 38(2) :311–322, 1980. 42
- Michela Sarlo, Giulia Buodo, Silvia Poli, and Daniela Palomba. Changes in EEG alpha power to different disgust elicitors : The specificity of mutilations. *Neuroscience Letters*, 382(3) : 291–296, jul 2005. 29
- Jean-Paul Sartre. *Esquisse d'une théorie des émotions*. Hermann, 1965. 16, 17

- Arman Savran, Koray Ciftci, Guillame Chanel, Javier Cruz Mota, Luong Hong Viet, Bülent Sankur, Lale Akarun, Alice Caplier, and Michele Rombaut. Emotion Detection in the Loop from Brain Signals and Facial Images. *eNTERFACE*, 06 :69–80, 2006. [1](#), [12](#), [31](#), [32](#), [44](#), [49](#), [50](#), [52](#), [59](#), [60](#), [61](#), [97](#), [IX](#), [XI](#)
- Kristina Schaaff and Tanja Schultz. Towards emotion recognition from electroencephalographic signals. *3rd International Conference on Affective Computing and Intelligent Interaction and Workshops*, pages 1–6, sep 2009a. [35](#)
- Kristina Schaaff and Tanja Schultz. Towards an EEG-based emotion recognizer for humanoid robots. In *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication*, pages 792–796, 2009b. [6](#), [35](#)
- Joseph Schachter. Pain, Fear, and Anger in Hypertensives and Normotensives : A Psychophysiological Study. *Psychosomatic Medicine*, 19(1) :17–29, 1957. [14](#)
- Stanley Schachter and Jerome Singer. Cognitive, social, and physiological determinants of emotional state. *Psychological review*, 69(5) :379–399, 1962. [16](#), [23](#), [48](#)
- Klaus R. Scherer. Psychological models of emotion. *The neuropsychology of emotion*, 2000. [18](#)
- Klaus R. Scherer. Which Emotions Can be Induced by Music? What Are the Underlying Mechanisms? And How Can We Measure Them? *Journal of New Music Research*, 33(3) : 239–251, sep 2004. [18](#), [48](#), [IX](#)
- Klaus R. Scherer. Unconscious Processes in Emotion : The Bulk of the Iceberg. *The unconscious in emotion*, pages 312–334, 2005a. [18](#), [36](#), [37](#), [48](#), [IX](#)
- Klaus R. Scherer. What are emotions? And how can they be measured? *Social Science Information*, 44(4) :695–729, dec 2005b. [38](#), [47](#)
- Klaus R. Scherer, Angela Schorr, and Tom Johnstone. *Appraisal processes in emotion : Theory, methods, research*. Oxford University Press, 2001. [18](#), [20](#), [48](#)
- Klaus R Scherer, Tanja Bänziger, and Etienne Roesch. *A Blueprint for Affective Computing : A sourcebook and manual*. Oxford University Press, 2010. [18](#)
- Konrad Schindler, Luc Van Gool, and Beatrice de Gelder. Recognizing emotions expressed by body pose : A biologically inspired neural model. *Neural Networks*, 21(9) :1238–1246, nov 2008. [25](#)
- Alois Schlögl. The Electroencephalogram and the Adaptive Autoregressive Model : Theory and Applications. Technical report, 2000. [19](#)
- Björn Schuller, Manfred Lang, and Gerhard Rigoll. Multimodal emotion recognition in audiovisual communication. In *Proceedings - 2002 IEEE International Conference on Multimedia and Expo, ICME 2002*, volume 1, pages 745–748, 2002. [29](#), [30](#), [31](#)

- Björn Schuller, Anton Batliner, Stefan Steidl, and Dino Seppi. Recognising realistic emotions and affect in speech : State of the art and lessons learnt from the first challenge. *Speech Communication*, 53(9-10) :1062–1087, 2011. [27](#)
- John Paul Scott. Animal behaviour. *Animal Behaviour*, 1958. [38](#)
- Nicu Sebe, Ira Cohen, Theo Gevers, and Thomas S. Huang. Emotion recognition based on joint visual and audio cues. In *Proceedings - International Conference on Pattern Recognition*, volume 1, pages 1136–1139, 2006. [31](#)
- Matthew Shreve, Sridhar Godavarthy, Dmitry Goldgof, and Sudeep Sarkar. Macro- and micro-expression spotting in long videos using spatio-temporal strain. In *2011 IEEE International Conference on Automatic Face and Gesture Recognition and Workshops, FG 2011*, pages 51–56, 2011. [25](#)
- Matthias Siemer, Iris B. Mauss, and James J. Gross. Same situation - different emotions : how appraisals shape our emotions. *Emotion*, 7(3) :592–600, aug 2007. [48](#)
- Liyanage C. Silva, Tsutomu Miyasato, and Ryohei Nakatsu. Facial emotion recognition using multi-modal information. *Proceedings of IEEE International Conference on Information, Communications and Signal Processing (ICICS1997)*, 1 :397–401, 1997. [31](#)
- Harsimrat Singh, Markus Bauer, Wojtek Chowanski, Yi Sui, Douglas Atkinson, Sharon Baurley, Martin Fry, Joe Evans, and Nadia Bianchi-Berthouze. The brain's response to pleasant touch : an EEG investigation of tactile caressing. *Frontiers in human neuroscience*, 8 (November) :893, jan 2014. [49](#)
- Ahmad Tauseef Sohaib, Shahnawaz Qureshi, Johan Hagelbäck, Olle Hilborn, and Petar Jerčić. Evaluating classifiers for emotion recognition using EEG. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 8027 LNAI, pages 492–501, 2013. [34](#), [35](#)
- Mohammad Soleymani, Sander Koelstra, Ioannis Patras, and Thierry Pun. Continuous emotion detection in response to music videos. In *2011 IEEE International Conference on Automatic Face and Gesture Recognition and Workshops*, pages 803–808, 2011. [47](#)
- Mohammad Soleymani, Jeroen Lichtenauer, Thierry Pun, and Maja Pantic. A multimodal database for affect recognition and implicit tagging. *IEEE Transactions on Affective Computing*, 3(1) :42–55, 2012a. [1](#), [12](#), [44](#), [49](#), [50](#), [53](#), [59](#)
- Mohammad Soleymani, Maja Pantic, and Thierry Pun. Multimodal emotion recognition in response to videos. *IEEE Transactions on Affective Computing*, 3(2) :211–223, 2012b. [24](#), [30](#), [31](#)
- Larry R. Squire. Memory and the hippocampus : a synthesis from findings with rats, monkeys, and humans. *Psychological review*, 99(2) :195–231, 1992. [19](#)

- Gerhard Stemmler, Marcus Heldmann, Cornelia A. Pauls, and Thomas Scherer. Constraints for emotion specificity in fear and anger : the context counts. *Psychophysiology*, 38(2) : 275–291, 2001. [23](#), [24](#), [48](#)
- Chad L. Stephens, Israel C. Christie, and Bruce H. Friedman. Autonomic specificity of basic emotions : Evidence from pattern classification and cluster analysis. *Biological Psychology*, 84(3) :463–473, 2010. [23](#), [24](#)
- Kazuhiko Takahashi. Remarks on emotion recognition from multi-modal bio-potential signals. In *Proceedings of the 2004 IEEE International Workshop on Robot and Human Interactive Communication*, pages 95–100, 2004. [26](#)
- Kazuhiko Takahashi and Akinori Tsukaguchi. Remarks on Emotion Recognition from Multi-Modal Bio-Potential Signals. *2nd International Conference on Autonomous Robots and Agents*, 2004. [28](#)
- Silvan S. Tomkins and Robert Mc Carter. What and where are the primary affects? Some evidence for a theory. *Perceptual and Motor Skills*, 18 :119–158, 1964. [41](#)
- Khiet P Truong, David A van Leeuwen, and Mark A Neerincx. Unobtrusive multimodal emotion detection in adaptive interfaces : speech and facial expressions. In *Foundations of Augmented Cognition*, pages 354–363, 2007. [29](#)
- Naotsugu Tsuchiya and Ralph Adolphs. Emotion and consciousness. *Trends in cognitive sciences*, 11(4) :158–67, apr 2007. [48](#)
- Stefano Valenzi, Tanvir Islam, Peter Jurica, and Andrezej Cichocki. Individual Classification of Emotions Using EEG. *Journal of Biomedical Science and Engineering*, 7 :604–620, 2014. [34](#), [35](#)
- Jaram Van Hooff. A structural analysis of the social behavior of a semi-captive group of chimpanzees. *Social communication and movement : Studies of interaction and expression in man and chimpanzee*, pages 75–162, 1973. [38](#)
- Anne Vanpé. *Expressions et micro-expressions spontanées de la face et de la voix en Interaction Homme-Machine : esquisse d'un modèle du " Feeling of Thinking"*. PhD thesis, 2011. [25](#), [29](#)
- Dimitrios Ververidis and Constantine Kotropoulos. Emotional speech recognition : Resources, features, and methods. *Speech Communication*, 48(9) :1162–1181, sep 2006. [27](#)
- Johannes Wagner, Jonghwa Kim, and Elisabeth André. From physiological signals to emotions : Implementing and comparing selected methods for feature extraction and classification. *IEEE International Conference on Multimedia and Expo, ICME 2005*, pages 940–943, 2005. [26](#)

- Johannes Wagner, Florian Lingenfelser, Elisabeth André, and Jonghwa Kim. Exploring fusion methods for multimodal emotion recognition with missing data. *IEEE Transactions on Affective Computing*, 2(4) :206–218, 2011. [29](#), [30](#), [31](#)
- Brian A. Wandell, Serge O. Dumoulin, and Alyssa A. Brewer. Visual field maps in human cortex. *Neuron*, 56(2) :366–383, 2007. [21](#)
- Shangfei Wang, Yachen Zhu, Guobing Wu, and Qiang Ji. Hybrid video emotional tagging using users' EEG and video content. *Multimedia tools and applications*, pages 1257–1283, 2014a. [27](#)
- Xiao-Wei Wang, Dan Nie, and Bao-Liang Lu. EEG-based emotion recognition using frequency domain features and support vector machines. *ICONIP'11 Proceedings of the 18th international conference on Neural Information Processing*, pages 734–743, 2011. [29](#), [34](#), [35](#)
- Xiao-Wei Wang, Dan Nie, and Bao-Liang Lu. Emotional state classification from EEG data using machine learning approach. *Neurocomputing*, 129 :94–106, apr 2014b. [6](#), [7](#), [29](#), [49](#)
- Amy Beth Warriner, Victor Kuperman, and Marc Brysbaert. Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior research methods*, 45(4) :1191–207, 2013. [42](#)
- Edward L. White. *Cortical circuits : synaptic organization of the cerebral cortex : structure, function, and theory*. Birkhäuser Basel, 1989. [19](#)
- Irene Winkler, M Jäger, V Mihajlović, and T Tsoneva. Frontal EEG asymmetry based classification of emotional valence using common spatial patterns. *World Academy of Science, Engineering and Technology*, 45 :373–378, 2010. [90](#)
- Martin Wöllmer, Florian Eyben, Stephan Reiter, Björn Schuller, Cate Cox, Ellen Douglas-Cowie, and Roddy Cowie. Abandoning emotion classes - Towards continuous emotion recognition with modelling of long-range dependencies. In *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, pages 597–600, 2008. [47](#)
- Haiyan Xu and Konstantinos N. (Kostas) Plataniotis. Affect recognition using EEG signal. *IEEE 14th International Workshop on Multimedia Signal Processing (MMSP)*, pages 299–304, sep 2012. [34](#), [53](#), [80](#), [95](#), [97](#), [IX](#), [XI](#)
- Wen Jing Yan, Qi Wu, Jing Liang, Yu Hsin Chen, and Xiaolan Fu. How Fast are the Leaked Facial Expressions : The Duration of Micro-Expressions. *Journal of Nonverbal Behavior*, 37(4) :217–230, jul 2013. [25](#)
- Michelle Yik, James a Russell, and James H Steiger. A 12-point circumplex structure of core affect. *Emotion*, 11(4) :705–731, aug 2011. [44](#)

- Chai Tong Yuen, Woo San San, Mohamed Rizon, and Tan Ching Seong. Classification of Human Emotions from EEG Signals using Statistical Features and Neural Network. *International Journal of Integrated Engineering*, 1(3) :71–79, 2011. [29](#), [35](#), [49](#)
- Zhihong Zeng and Maja Pantic. A survey of affect recognition methods : Audio, visual, and spontaneous expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(1) :39–58, 2009. [27](#)
- Cong Zong and Mohamed Chetouani. Hilbert-Huang transform based physiological signals analysis for emotion recognition. In *IEEE Symposium on Signal Processing and Information Technology (ISSPIT)*, pages 334–339, 2009. [26](#)

RESUME : Alors que l'utilisation de l'électro-encéphalographie (EEG) a longtemps été confinée au domaine médical, l'intérêt dans les interfaces cerveau-machine basées sur ces signaux s'est développé au cours de ces dernières années pour les applications grand public. Les enregistrements EEG ont tout particulièrement retenu l'attention des chercheurs dans le domaine de l'informatique affective, affective computing, dans le but de réaliser des travaux sur l'analyse du comportement humain, et ce plus particulièrement sur la reconnaissance automatique des émotions. Comparé à d'autres modalités utilisées dans de précédents travaux sur la reconnaissance des émotions, telles que la parole, les expressions du visage, les mouvements et d'autres signaux physiologiques, l'EEG a l'avantage de pouvoir capturer des informations liées à l'état émotionnel interne qui n'est pas forcément traduit par des manifestations extérieures observables.

La reconnaissance des émotions est habituellement envisagée sous l'angle d'un problème de classification où le choix de caractéristiques appropriées est essentiel dans le but de s'assurer une précision de reconnaissance satisfaisante. Une de nos problématiques repose sur le fait que, dans le cadre des caractéristiques EEG, un consensus n'a toujours pas été réalisé sur un ensemble standard de caractéristiques qui permettrait de garantir une distinction performante des émotions d'un sujet humain. Nous explorons une grande variété de caractéristiques temporelles, spectrales et spatiales pouvant être potentiellement utiles dans le cadre de la reconnaissance des émotions et nous les comparons à d'autres, exposées dans de précédents travaux, en utilisant un protocole expérimental rigoureux. Nous évaluons plus particulièrement l'efficacité de plusieurs caractéristiques spectrales, n'ayant pas été précédemment proposées pour le problème de la classification. Nos résultats montrent que les nouvelles caractéristiques spectrales que nous proposons sont compétitives comparées à celles précédemment utilisées. Elles nous amènent de plus vers une configuration mono-canal performante de la reconnaissance des émotions, ce qui implique un fort potentiel pour les applications destinées au grand public.

Au sein des corpus existants et accessibles destinés à la reconnaissance des émotions en informatique affective, l'aspect de la dynamique de l'émotion n'est pas pris en considération. Ces corpus présentent également un manque de variabilité dans les données et ne possèdent les enregistrements que d'un nombre limité de participants. Ces raisons nous ont menées à proposer un corpus multi-modal destiné à l'analyse de l'état émotionnel qui s'attache à répondre au mieux à certaines faiblesses des corpus existants. Nous employons différentes stratégies d'élicitation de l'émotion, par le biais de l'utilisation de stimuli visuels et audio-visuels, et nous proposons également une approche novatrice dans la stratégie d'annotation de l'émotion ressentie, en intégrant en plus de la retranscription de l'émotion ressentie de manière globale, la retranscription de ses variations.

Ce nouveau corpus permettra dans un premier temps de renforcer la validation de l'approche proposée dans nos travaux. D'autre part, en accord avec les résultats encourageants obtenus, il sera possible d'envisager à plus long terme la caractérisation de la dynamique de l'état émotionnel de l'utilisateur qui ouvre la voie à de nouveaux modèles qui pourraient prédire, par exemple, l'augmentation de l'anxiété de l'utilisateur en fonction de la situation dans laquelle il est placé.

MOTS-CLEFS: EEG, émotion, classification, corpus

ABSTRACT: While the usage of electroencephalographic (EEG) recording has been for long confined in the medical field, the recent years have seen a growing interest in EEG-based brain-computer interfaces (BCI) for general public applications. In particular EEG-recording has attracted the attention of researchers in the field of affective computing as part of the effort to perform human-behaviour analysis tasks, especially automatic emotion recognition. Compared to other modalities which have been considered in previous work on emotion recognition, such as speech, facial expressions, gestures or other physiological signals, EEG has the advantage of capturing information related to internal emotional states not necessarily resulting in any observable external manifestations.

Emotion recognition is usually approached as a classification problem where the choice of appropriate features is critical to ensure satisfactory recognition accuracy. As far as EEG-features are concerned, a consensus has not yet been reached as to a standard set of attributes that could guarantee a successful characterization of a human-subject's emotions. We explore a wide range of temporal, spectral and spatial features potentially useful for emotion recognition, comparing them to previous proposals through a rigorous experimental evaluation. In particular we assess the effectiveness of various spectral features that were not previously envisaged for the problem of classifying emotions. Our results show that the new spectral shape features that we propose are very competitive compared to previously used ones. They are additionally amenable to successful emotion recognition in single-channel setups, which holds a great potential for general public applications.

Within the existing and accessible datasets for emotion recognition in affective computing, the aspect of the dynamics of emotion is not being considered. Those datasets also show a lack of variability in the data and contain the records of a limited number of participants. These reasons have led us to propose a new multimodal dataset for the analysis of the emotional state that attaches to respond better to certain weaknesses of the existing datasets. We use different strategies for the emotion elicitation, through the use of both visual and audio-visual stimuli. We also provide an innovative approach to the strategy on the annotation of the emotion experienced by the user by integrating more than the transcript of the emotion globally felt, but also the transcript of its variations.

This new dataset will, in a first stage, enhance the validation of the approach we propose in our work. Secondly, in agreement with the encouraging results, it will be possible to consider a longer-term characterization of the dynamics of the emotional state of the user, which opens the way for new models that could predict, for example, the increase in the user's anxiety according to the situation in which he is placed.

KEY-WORDS: EEG, emotion, machine learning, dataset



