



HAL
open science

Reconnaissance d'états émotionnels à partir des interactions avec un smartphone : Conception des méthodes et outils pour le domaine de la relation client

Nicolas Simonazzi

► To cite this version:

Nicolas Simonazzi. Reconnaissance d'états émotionnels à partir des interactions avec un smartphone : Conception des méthodes et outils pour le domaine de la relation client. Automatique. Université de Bordeaux, 2021. Français. NNT : 2021BORD0272 . tel-03474164

HAL Id: tel-03474164

<https://theses.hal.science/tel-03474164>

Submitted on 10 Dec 2021

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.

THÈSE PRÉSENTÉE
POUR OBTENIR LE GRADE DE
DOCTEUR DE
L'UNIVERSITÉ DE BORDEAUX

École Doctorale Sciences Physiques et de l'Ingénieur
Automatique, Productique, Signal et Image, Ingénierie Cognitive
Spécialité Ingénierie Cognitive

Par Nicolas Simonazzi

**Reconnaissance d'états émotionnels à partir des
interactions avec un smartphone**

Conception des méthodes et outils pour le domaine de la relation client

Sous la direction de : Jean-Marc Salotti

Soutenue le 19/11/21

Membres du jury :

M. CLAVERIE, Bernard
M. CAUSSE, Mickaël
Mme. GALY, Edith
Mme TCHERKASSOF, Anna
M. SALOTTI, Jean-Marc
Mme. DUBOIS, Caroline
M. LE GOFF, Philippe

IMS, ENSC Institut Polytechnique de Bordeaux
ISAE-SUPAERO, Université de Toulouse
LAPCOS, Université Côte d'Azur
LIP/PC2S, Université Grenoble Alpes
IMS, ENSC Institut Polytechnique de Bordeaux
Orange Labs Service
Orange Labs Service

Président
Rapporteur
Rapporteur
Examineur
Directeur de thèse
Invité
Invité

Reconnaissance d'états émotionnels à partir des interactions avec un smartphone. Conception des méthodes et outils pour le domaine de la relation client.

Résumé :

L'idée de cette thèse est née dans l'entité Orange Labs Service de l'entreprise de télécommunication Orange. L'objectif de ce projet de recherche est d'améliorer la qualité de la relation client de l'entreprise par une meilleure compréhension des émotions des clients.

Les émotions sont essentielles à comprendre en situation de communication. En effet, elles sont indispensables pour s'adapter et identifier les intentions d'un interlocuteur. De plus, les smartphones représentent un nouveau moyen de communication privilégié pour dialoguer avec des conseillers clients ou des agents conversationnels (chatbot), notamment par l'intermédiaire de messageries instantanées textuelles. Cependant, ces moyens de communication coupent les interlocuteurs de nombreuses modalités pour exprimer leurs états émotionnels, tels que les expressions non verbales.

L'objectif de cette thèse est donc d'explorer la possibilité de reconnaître des émotions de manière automatique et non invasive. Dans un premier temps, nous avons étudié les théories existantes dans la science des émotions. Nous avons fondé notre cadre théorique, à partir de la théorie de l'évaluation cognitive et du modèle des processus-composant, afin d'identifier les composantes associées à l'expression d'états émotionnels. Dans un second temps, nous avons réalisé une revue de littérature des méthodes utilisées dans le domaine de l'informatique affective, pour concevoir des systèmes de reconnaissance automatiques d'émotions.

À l'issue de ces recherches, nous avons fait le choix de mesurer la composante de l'expression motrice des états émotionnels à partir des capteurs inertiels (accéléromètre et gyroscope) et des interactions avec le clavier du smartphone. L'apport de cette thèse concerne, notamment, la conception et le développement d'outils pour réaliser des expérimentations sur smartphone et pour enregistrer les données nécessaires. Une application d'expérimentation a ainsi été développée, sous un format de messagerie instantanée, pour induire des émotions chez les utilisateurs et enregistrer les données nécessaires.

Plusieurs méthodologies d'induction d'émotions ont été testées en laboratoire, afin de recueillir des données comportementales à partir des capteurs de smartphones. Les données ont été labellisées à partir des auto-évaluations des participants. À la suite du traitement des signaux temporels, nous avons généré un corpus de données utilisables

pour de l'apprentissage automatique. Ce corpus a permis d'entraîner et d'évaluer des modèles de classification supervisée par apprentissage automatique. Des modèles de classification inter-personnels et intra-personnels ont été créés pour classifier des niveaux de valence déclarés par les sujets des expérimentations. Les résultats montrent une forte variabilité de performance de classification en fonction des participants.

Mots clés : Reconnaissance d'émotions, Informatique affective, Interaction homme-machine, Apprentissage automatique, Relation client, Smartphone

Emotion recognition through interaction with smartphones. Design of methods and tools for the customer relationship domain.

Abstract :

The idea of this thesis came from the Orange Labs Service entity in the telecommunication company Orange. The objective of the research project is to improve the quality of the company's support services through a better understanding of customers' emotions.

Emotions are essential to be considered in a communication situation. Indeed, they are necessary to adapt to and understand the intentions of a conversational partner. Moreover, smartphones represent a new medium for communicating with customer advisors or chatbot, especially through instant messaging. However, these means of communication deprive the interlocutors of many ways to express their emotional states, such as non-verbal expressions.

The objective of this thesis is therefore to explore the feasibility of recognizing emotions in an automatic and non-invasive way. First, we explored the existing theories in the science of emotions. We based our theoretical framework on the theory of cognitive appraisal and the process-component model in order to identify the components associated with the expression of emotional states. Then, we conducted a literature review of the methods used in the field of affective computing to design automatic emotion recognition systems.

As a result of this research, we have decided to measure the motor expression component of emotional states using inertial sensors (accelerometer and gyroscope) and interactions with the smartphone keyboard. The contribution of this thesis concerns, among other things, the design and development of tools to perform experiments on a smartphone and to record the necessary data. An experimentation application was developed, under an instant messaging format, to induce emotions and to record the necessary data.

Several emotion induction methodologies were tested in the laboratory to collect interaction data from smartphone sensors. The data were labeled according to the participants' self-reports. Following the processing of the temporal signals, we generated a corpus of data suitable for machine learning. This corpus was used to train and evaluate supervised classification models by machine learning. Inter-personal and intra-personal classification models were created to classify self-reported valence levels. The results show a high degree of variability in classification performance across participants.

Keywords : Emotions Recognition, Affective computing, Human-System Interface, Machine Learning, Customer Relationship, Smartphone

Equipe Cognitive Ingénierie Humaine

UMR 5218 – Laboratoire de l'Intégration du Matériau au Système Université de Bordeaux. 351, cours de la Libération – F 33 405 TALENCE.

&

“Il n’y a pas de réussite facile ni d’échecs définitifs.”
Marcel Proust

Remerciements

Cette thèse ne porte pas seulement sur les émotions, elle en a aussi été la source de nombreuses... Mais parmi toutes les difficultés et les réussites, je voudrais surtout souligner l'importance de l'aventure humaine et des très nombreuses personnes qui ont collaboré de près ou de loin à ce projet.

Dans un premier temps, je voudrais exprimer ma reconnaissance envers mes encadrants, pour la confiance qu'ils m'ont accordée, leur grand soutien et les connaissances qu'ils m'ont apportés.

Je voudrais d'abord remercier mon directeur de thèse, Jean-Marc Salotti, pour sa disponibilité, ses enseignements et ses nombreux conseils. Ainsi que pour m'avoir donné l'opportunité de contribuer aux cours d'intelligence artificielle.

Je tiens à remercier chaleureusement Caroline Dubois, qui a consacré beaucoup d'énergie pour amorcer ce projet, ainsi que pour m'encadrer et me diriger. C'est notamment grâce à son implication et son humanité que ce projet a été aussi enrichissant.

Je remercie Philippe Le Goff d'avoir accepté de m'encadrer en plein milieu de parcours. Ce n'était pas une tâche facile, pourtant ses contributions pour la thèse et le programme DI4H sont très nombreuses.

Je remercie Dominique Seminel d'avoir proposé ce sujet et pour avoir été à mes côtés au début de ce projet. Sa vision a un impact important sur cette thèse.

Je tiens à remercier Édith Galy et Mickaël Causse d'avoir accepté de consacrer du temps sur mon sujet en tant que rapporteurs et de faire partie du jury de thèse. De plus, je remercie aussi Anna Tcherkassof et Bernard Claverie de me faire l'honneur d'être membre du jury de thèse, en tant qu'examineurs.

J'ai une pensée particulière pour mes collègues avec qui j'ai partagé plus de trois années au sein du groupe Orange et de l'École Nationale Supérieure de Cognitive :

À toute l'équipe Digital Intelligence 4 Human : Nolwenn Le Gonidec, Caroline Fenzy-Pierre, Christian Sirchia, Sylvie Plantin, Patricia Holysz-Fouquet, Pierre Levis et Valérie Chapuis, merci pour votre aide, votre dynamisme et toute votre bienveillance. Cette équipe possède une alchimie unique.

À Maud Levavasseur et Amina Makhlof, merci pour votre soutien, vos retours sur mon sujet et de m'avoir intégré sein de l'équipe DJACK.

À Laëtitia Gros, je te remercie pour ton aide pour mes expérimentations et pour

tous les échanges passionnants que nous avons eue autour des émotions.

Je remercie Guillaume Tardiveau pour l'intérêt qu'il a porté à ces recherches, pour ses retours pertinents et pour toutes les opportunités.

Je remercie les doctorants de l'ENSC : Baptiste Prébot, Théodore Letouze et Hélène Unrein pour les diverses aides et échanges.

Je voudrais aussi remercier les membres de l'équipe AUCTUS avec qui j'ai pu interagir et qui m'ont donné des retours sur mon sujet.

Enfin, je tiens à exprimer ma gratitude envers les étudiants et stagiaires de l'ENSC que j'ai eu le plaisir d'encadrer et qui ont apporté des contributions significatives à cette thèse : Sylvain Hazard, Tom Vallée, Katell Maro, Emma Goumillout, Martin Rollin, Maudeline Marlier, Julien Baillet et Marie Morelle. Félicitations pour votre travail !

Bien sûr, je n'oublie pas tous mes amis et ma famille :

À Charles qui n'hésite pas à traverser la France pour qu'on puisse passer du temps ensemble. Merci.

À Florian et Alexandre : colocataires, cobayes et correcteurs, mais surtout compères. Merci les amis !

À Chafik, tu es désormais mon conseiller attitré en traitement du signal. Merci pour ton aide.

À Jade, Océane, Luc et Nicolas. Merci pour vos relectures.

À tous mes amis : Aline, Alexis, Aubin, Florie-Anne, Jean, Jessica, Johann, Kilian, Marilène, Thibaud, Sébastien, Sylvain... Rendez-vous au prochain apéro !

Merci à Christophe et à Béatrice pour vos invitations et vos encouragements.

Je remercie mes parents, merci de m'avoir toujours poussé à apprendre, à être curieux et à me dépasser intellectuellement. C'est grâce à vous que je me suis senti capable de me lancer dans un tel projet.

Et pour terminer, je voudrais te remercier Margaux, toujours en première ligne pour les moments difficiles comme les moments de joie. Tu m'as toujours soutenu et encouragé avec une bienveillance sans faille et tu m'as donnée l'énergie pour aller de l'avant et surmonter les défis. Je te dois beaucoup...

Glossaire

Acronyme	Nom entier
API	Interface de Programmation d'Application
CE	Conductance Électrodermale
CRM	Customer Relationship Management
ECG	Électrocardiogramme
EEG	Électroencéphalogramme
GEW	Geneva Emotion Wheel
GEWMI	Geneva Emotion Wheel Mobile Interface
IHM	Interface Homme-Machine
IIC	Intervalle Inter-battement Cardiaque
LOO	<i>Leave One Out</i> Validation croisée un contre tous
PS	Pulsation Sanguine
PPM	Photopléthysmographe
RGP	Réponse Galvanique de la Peau
RGPD	Règlement Général sur la Protection des Données
SAM	Self-Assesment Manikin
SMOTE	Synthetic Minority Oversampling TechniquE
SVM	Machine à Vecteur de Support
TAL	Traitement Automatique du Langage
TIC	Technologie de l'Information et de la Communication
VFC	Variabilité de la Fréquence Cardiaque



Table des matières

Résumé	5
Glossaire	9
I Introduction	22
I.1 Contexte des travaux de recherche	22
I.2 Problématique générale	23
I.3 Organisation du manuscrit	25
II Science des émotions : Définitions, théories et modèles	27
II.1 Introduction	28
II.2 Historique des théories et débats	28
II.2.1 Origine et définitions du terme émotion	28
II.2.2 Les émotions avant le 19 ^e siècle	29
II.2.3 Approche darwiniste et néodarwiniste	31
II.2.4 Débat James-Lange versus Cannon-Bard	32
II.2.5 Approche cognitiviste	34
II.2.6 Approche socio-constructiviste	36
II.3 Théorie de l'évaluation cognitive	37
II.3.1 Origine de la théorie	37
II.3.2 Définitions et composantes	40
II.3.3 Évaluation cognitive	42
II.3.4 Réponses psychophysologiques	47
II.3.5 Tendance à l'action	49
II.3.6 Expression motrice	50
II.3.7 Sentiment subjectif	52

II.3.8	Discussion	55
II.4	Classification et représentation des émotions	55
II.4.1	Taxonomie des états affectifs	55
II.4.2	Classification des émotions	57
II.4.3	Modèles discrets et catégoriels	58
II.4.4	Modèles dimensionnels	63
II.4.5	Discussion	66
II.5	Conclusion et positionnement	67
III	Étude des émotions et nouvelles technologies	69
III.1	Introduction	70
III.2	Informatique affective	71
III.2.1	Définition et enjeux	71
III.2.2	Application au contexte du service client	73
III.2.3	Expression des émotions dans le cadre des messageries instantanées	74
III.3	Reconnaissance automatique des émotions	76
III.3.1	Introduction	76
III.3.2	Modalité visuelle	77
III.3.3	Modalité audio	80
III.3.4	Modalité textuelle	82
III.3.5	Modalité physiologique	84
III.3.6	Approche multimodale	87
III.3.7	Difficultés en reconnaissance d'émotions	88
III.4	Reconnaissance automatique d'émotions et interfaces homme-machine .	90
III.4.1	Introduction	90
III.4.2	Ordinateur	91
III.4.2.1	Dynamique de frappe au clavier	92
III.4.2.2	Usage de la souris	93
III.4.3	Smartphone	93
III.4.3.1	Données des écrans tactiles et claviers virtuels	96
III.4.3.2	Données des capteurs inertiels	97
III.4.4	Discussion	100
III.5	Informatique affective et contexte expérimental	102

III.5.1	Protocoles expérimentaux	102
III.5.2	Induction d'émotions	103
III.5.2.1	Images	104
III.5.2.2	Vidéos	104
III.5.2.3	Manipulation de la difficulté	105
III.5.2.4	Comparaison des méthodes	106
III.5.3	Mesures déclaratives	107
III.5.4	Discussion	108
III.6	Conclusion	109
IV	Matériel et méthode	111
IV.1	Introduction et positionnements méthodologiques	113
IV.2	Conception des outils d'expérimentation	116
IV.2.1	Présentation des outils	116
IV.2.2	Fonctionnement général de l'application d'expérimentation	117
IV.2.3	Interface de l'application d'expérimentation	118
IV.2.3.1	Questionnaires	119
IV.2.3.2	Tâches	119
IV.2.3.3	Autoévaluation	120
IV.2.4	Adaptation de la <i>Geneva Emotion Wheel</i> en format mobile	120
IV.3	Pré-expérimentation	122
IV.3.1	Présentation et objectif	122
IV.3.2	Protocole expérimental	122
IV.3.2.1	Induction d'émotions et mesure subjective	122
IV.3.2.2	Plan expérimental	125
IV.3.3	Hypothèse générale	126
IV.3.4	Passation	127
IV.4	Expérimentation principale	127
IV.4.1	Présentation et objectif	127
IV.4.2	Protocole expérimental	128
IV.4.2.1	Induction d'émotions et mesure subjective	128
IV.4.2.2	Enregistrement des capteurs du smartphone	130
IV.4.2.3	Enregistrement des mesures physiologiques	131

IV.4.2.4	Plan expérimental	132
IV.4.3	Hypothèses générales	133
IV.4.4	Passation	134
IV.5	Expérimentation complémentaire	136
IV.5.1	Présentation et objectif	136
IV.5.2	Protocole expérimental	136
IV.5.2.1	Induction d'émotions et mesure subjective	136
IV.5.2.2	Plan expérimental	138
IV.5.3	Passation	138
IV.5.4	Hypothèse générale	139
IV.6	Méthode de reconnaissance automatique d'émotions	140
IV.6.1	Introduction	140
IV.6.2	Démarche générale	141
IV.6.3	Extraction des signaux bruts	143
IV.6.4	Prétraitement des signaux	144
IV.6.4.1	Enregistrements des capteurs inertiels	144
IV.6.4.2	Mesures physiologiques	146
IV.6.5	Génération de descripteurs	146
IV.6.5.1	Échantillonnage	146
IV.6.5.2	Descripteurs des signaux inertiels	147
IV.6.5.3	Descripteurs des évènements clavier	150
IV.6.5.4	Descripteurs des capteurs physiologiques	151
IV.6.5.5	Fusion de données multimodale	152
IV.6.6	Mise en œuvre de l'apprentissage automatique	153
IV.6.6.1	Prétraitement des descripteurs	157
IV.6.6.2	Sélection des descripteurs	158
IV.6.6.3	Apprentissage des modèles de classification	159
IV.6.6.4	Évaluation des modèles de classification	160
IV.6.7	Discussion	164
IV.7	Conclusion	166
V	Résultats	168
V.1	Introduction	169

V.2	Analyse du sentiment subjectif	170
V.2.1	Méthodologie d'analyse	170
V.2.2	Pré-expérimentation	170
V.2.2.1	Analyse de la valence des auto-évaluations en fonction des extraits vidéo	171
V.2.2.2	Analyse de l'activation des auto-évaluations en fonction des extraits vidéo	173
V.2.2.3	Relation entre les réponses au test et auto-évaluations .	174
V.2.2.4	Discussion	175
V.2.3	Expérimentation principale	177
V.2.3.1	Nombre d'émotions déclarées	177
V.2.3.2	Intensité des émotions déclarées	179
V.2.3.3	Valence des émotions déclarées	181
V.2.3.4	Analyse des verbatims	186
V.2.3.5	Discussion	188
V.2.4	Expérimentation complémentaire	191
V.2.4.1	Valence des émotions déclarées	191
V.2.4.2	Détails des libellés	193
V.2.4.3	Discussion	194
V.2.5	Discussion	196
V.3	Résultats du système de reconnaissance automatique	197
V.3.1	Labellisation des descripteurs	197
V.3.2	Classification de données d'interaction avec un smartphone . . .	198
V.3.3	Classification avec les données de la montre Empatica E4	201
V.3.4	Classification avec fusion des modalités	204
V.3.5	Discussion	206
V.4	Conclusion	209
VI	Conclusion	211
VI.1	Synthèse et apports des recherches	211
VI.2	Réflexions et perspectives	214
Bibliographie		217
A.	Captures d'écran de l'application GEWMI	234

B.	Formulaire de consentement	234
C.	Questionnaires	237
C.I	Caractéristiques personnelles	237
C.II	Habitudes et sentiment de compétence pour l'utilisation de smart- phone	237
C.III	Évaluation du système d'auto-évaluation des émotions	238
C.IV	Évaluation de l'impulsivité émotionnelle	238
D.	Détails des vidéos de l'expérimentation principale	239
E.	Réponses aux questionnaires de l'expérimentation principale	240
E.I	Caractéristiques personnelles	240
E.II	Habitudes et sentiment de compétence pour l'utilisation de smart- phone	241
E.III	Évaluation du système d'auto-évaluation des émotions	242
E.IV	Évaluation de l'impulsivité émotionnelle	244
F.	Représentation des signaux bruts pour une tâche de l'expérimentation principale	248
F.I	Capteurs du smartphone	248
F.II	Capteurs de la montre Empatica E4	251
G.	Détails des calculs sur les séries temporelles	255

Table des figures

II.1	Séquence temporelle de la théorie de James (1884)	33
II.2	Séquence temporelle de la théorie de Cannon (1927)	34
II.3	Séquence temporelle de la théorie de Schachter-Singer (Schachter & Singer, 1962)	35
II.4	Séquence temporelle de la théorie de l'évaluation cognitive d'Arnold (1960)	38
II.5	Modèle des composantes de Scherer et de leurs interactions avec d'autres systèmes fonctionnels (D. M. Grandjean & Scherer, 2009 ; Scherer, 2005)	43
II.6	Représentation des trois modules du sentiment subjectif (Dan-Glauser, 2008 ; Scherer, 2004)	54
II.7	Photographies des expressions faciales caractéristiques des six émotions basiques d'Ekman et Friesen (1971)	59
II.8	Modèle de représentation d'émotions de Plutchik (1984)	60
II.9	Modèle de représentation d'émotions de Russell (1980, 2003)	64
II.10	Modèle PAD proposé par (mehrabian1996pleasure), schéma inspiré de Bakker <i>et al.</i> (2014)	65
III.1	Description des différentes approches d'apprentissage automatique	79
III.2	Caméra kinect de l'entreprise Microsoft (Kinect, 2020), licence creative commons CC BY 3.0	80
III.3	Appareils de mesures physiologiques	85
III.4	Lunettes d'occulométrie Tobii Glasses 2 développées par l'entreprise Tobii (Tobii, 2021), licence creative commons CC BY 3.0	85
III.5	Montre E4 développée par l'entreprise Empatica (Empatica, 2021)	86
III.6	Smartphone et principaux capteurs associés. (Sparkfun, 2021) licence creative commons CC BY 2.0 / ©Raimond Spekking / CC BY-SA 4.0	95

III.7	Représentation spatiale des axes d'accélération enregistrés par un accéléromètre et des axes de rotation enregistrés par un gyromètre	98
III.8	Représentation de l'échelle SAM (<i>Self Assessment Manikin</i>) (Bradley & Lang, 1994)	107
III.9	Représentation de la Geneva Emotion Wheel (Sacharin <i>et al.</i> , 2012)	108
IV.1	Schéma fonctionnel	118
IV.2	Captures d'écran des interactions avec l'agent conversationnel par questionnaire	119
IV.3	Captures d'écran des médias et activités pouvant être intégrés à l'application	120
IV.4	Écrans de sélection du label d'émotion et de l'intensité	121
IV.5	Représentation des conditions expérimentales sur le circomplexe de Russell (Russell, 1980)	123
IV.6	Exemple de feedback de la part du chatbot	124
IV.7	Intégration de l'instrument SAM dans l'application d'expérimentation (Bradley & Lang, 1994)	125
IV.8	Déroulement global de l'expérimentation	126
IV.9	Contre-balancements	132
IV.10	Déroulement global de l'expérimentation	133
IV.11	Capture de l'enregistrement vidéo de l'expérimentation	135
IV.12	Déroulement du test de logique	137
IV.13	Déroulement global de l'expérimentation	139
IV.14	Démarche générale pour le traitement des signaux et l'apprentissage automatique	142
IV.15	Représentation d'une série temporelle de signal d'accéléromètre pour une tâche	143
IV.16	Représentation d'une série temporelle filtrée de signal d'accéléromètre pour une tâche	145
IV.17	Représentation des processus pour l'échantillonnage et l'extraction des descripteurs	147
IV.18	Représentation des processus pour l'échantillonnage et l'extraction des descripteurs	148

IV.19	Représentation des processus pour l'échantillonnage et l'extraction des descripteurs	149
IV.20	Démarche générale pour l'apprentissage des modèles de classification . . .	156
IV.21	Matrices de confusion sur la classification de phase (<i>watching</i> = visionnage). La matrice de gauche est pour le modèle de référence et le modèle de droite un modèle de régression logistique	162
V.1	Distribution des auto-évaluations de valence (1 = valence négative, 5 = valence positive	170
V.2	Distribution des auto-évaluations d'activation (1 = activation faible, 5 = activation forte	171
V.3	Moyenne de valence déclarée par condition	172
V.4	Moyenne d'activation déclarée par condition	173
V.5	Nombre moyen de réponses correctes en fonction de l'auto-évaluation de la valence	175
V.6	Nombre moyen d'émotions déclarées	178
V.7	Moyenne d'intensité des émotions déclarées	180
V.8	Nombre moyen d'émotions par valence et par condition	181
V.9	Distribution des libellés sous la condition positive	184
V.10	Distribution des libellés sous la condition négative	185
V.11	Distribution des libellés en condition neutre	186
V.12	Moyenne de polarité des émotions déclarées	187
V.13	Nombre moyen d'émotions déclarés par valence et par condition	192
V.14	Distribution des libellés en condition de succès	193
V.15	Distribution des libellés en condition d'échec	194
V.16	Matrices de confusion pour un sujet sélectionné sur les données d'interaction avec un smartphone	200
V.17	Matrices de confusion pour un sujet sur les données d'Empatica E4	203
V.18	Matrices de confusion pour un sujet à partir des données fusionnées	205

Liste des tableaux

II.1	Liste des composantes du modèle des processus composants (D. M. Grandjean & Scherer, 2009 ; Scherer, 2005)	41
II.2	Classification des phénomènes affectifs (Scherer <i>et al.</i> , 2000 ; Scherer, 2005)	56
II.3	Catégorisation des émotions du modèle OCC (Ortony <i>et al.</i> , 1990), (adaptation de (Mahboub, 2011)	62
III.1	Différenciation des dispositifs numériques les plus courants (Lugtig & Toepoel, 2016)	91
III.2	Criticité des données recueillies sur smartphone (Politou <i>et al.</i> , 2017) . .	101
III.3	Type de protocole pour la collecte de données	103
IV.1	Moyenne et écart-type des annotations de valence et d'activation des vidéos sélectionnées (Baveye <i>et al.</i> , 2013)	124
IV.2	Moyenne et écart-type des annotations de valence et d'activation des vidéos sélectionnées (Samson <i>et al.</i> , 2016)	129
IV.3	Enchaînement d'actions pour chaque extrait vidéo	130
IV.4	Données enregistrées	131
IV.5	Capteurs enregistrés par la montre Empatica E4	131
IV.6	Descripteurs issus des signaux des capteurs inertiels	150
IV.7	Descripteurs issus des mesures physiologiques	152
IV.8	Matrice de confusion	161
IV.9	Métriques d'évaluation	161
IV.10	Moyenne d'exactitude pour une classification de phase de description vs visionnage	164
V.1	Test post-hoc de Mann-Whitney sur la dimension de la valence	172

V.2	Tests de Mann-Whitney sur la dimension d'activation	174
V.3	Tests de Mann-Whitney sur le nombre d'émotions déclarées	179
V.4	Tests de Mann-Whitney sur les moyennes d'intensité	180
V.5	Tests de Mann-Whitney sur la valence positive	182
V.6	Tests de Mann-Whitney sur la valence négative	183
V.7	Tests de Mann-Whitney sur la polarité des descriptions	188
V.8	Modèles inter-personnels sur les données d'interaction : moyenne d'exac- titude pour la classification de la valence (négative vs non négative) . . .	199
V.9	Modèles intra-personnels sur les données d'interaction : moyenne d'exac- titude pour la classification de la valence (négative vs non négative) . . .	200
V.10	Liste des dix descripteurs les plus corrélés à la variable dépendante	201
V.11	Modèles intra-personnels sur les données physiologiques : moyenne d'exactitude pour la classification de la valence (négative vs non négative)	202
V.12	Liste des dix descripteurs à partir des données fusionnées les plus corrélés à la variable dépendante	204
V.13	Modèles intra-personnels sur les données fusionnées : moyenne d'exacti- tude pour la classification de la valence (négative vs non-négative)	205
V.14	Liste des dix descripteurs de la montre Empatica E4 les plus corrélés à la variable dépendante	206

Chapitre I

Introduction

Sommaire

I.1	Contexte des travaux de recherche	22
I.2	Problématique générale	23
I.3	Organisation du manuscrit	25

I.1 Contexte des travaux de recherche

La transformation numérique a profondément bouleversé nos moyens de communication au cours des dernières années, notamment dans le domaine de la relation client, dont les pratiques et les usages sont en pleine métamorphose. En effet, les évolutions des Technologies de l'Information et de la Communication (TIC) ont démultiplié les canaux de communication (emails, messageries instantanées, réseaux sociaux...), s'ajoutant aux canaux habituels (boutiques, téléphones, courriers...). Cette transformation s'accompagne de l'apparition de nouveaux supports de plus en plus performants et mobiles, tels que les téléphones intelligents (*smartphone*), les tablettes ou autres objets connectés (montres connectées...). Cette évolution a permis non seulement de réduire significativement les coûts, mais aussi de générer et collecter une grande quantité de données. Il en résulte une meilleure connaissance des clients, ainsi que la possibilité de proposer une expérience personnalisée pour le consommateur. L'intelligence artificielle joue un rôle central dans la personnalisation et l'optimisation du support et de l'expérience client dans le monde de l'entreprise. Elle offre, par exemple, une forme d'automatisation des communications, avec l'utilisation d'agents conversationnels textuels (*chatbot*) ou vocaux (*voicebot*). Ainsi, le domaine du service client s'est largement diversifié et par conséquent complexifié.

Ces transformations des usages ont des conséquences concrètes sur la manière d'interagir entre humains, conditionnant les communications à certaines modalités

(seulement textuelle ou vocale par exemple). Les situations de communications entre conseillers et clients, qui sont au cœur de la relation client, sont particulièrement impactées par ces transformations. Nous pouvons nous poser la question des effets de l'usage de nouveaux canaux de communication sur les échanges humains, et plus particulièrement sur l'aspect émotionnel et cognitif.

Le groupe Orange, en tant qu'opérateur de télécommunication, est un acteur majeur de cette mutation numérique. C'est pourquoi l'entité de recherche et développement, Orange Labs Service, étudie de près ce phénomène, que cela soit pour des applications concrètes ou pour anticiper les changements à venir. Le programme de recherche *Digital Intelligence 4 Human*, dans lequel s'intègre cette thèse, s'intéresse notamment à mesurer les effets de la numérisation du service client autant pour les conseillers que pour les clients. La place de l'humain au sein des nouvelles technologies est une problématique déterminante de ce programme de recherche, avec l'objectif d'instaurer la meilleure interaction possible. Cette démarche de recherche multidisciplinaire implique, entre autres, la mise en œuvre d'études, de méthodologies et d'outils pour développer les connaissances et les compétences du groupe sur ce sujet.

La discipline de la cognitique partage ces réflexions, avec la perspective d'améliorer les interactions homme-machine, en faisant intervenir les sciences humaines et les nouvelles technologies. Ainsi, l'équipe CIH, Cognitique et Ingénierie Humaine, du laboratoire IMS Bordeaux, s'intéresse particulièrement à l'évaluation des utilisateurs et des usages individuels, notamment en situation de communication. Les apports de ce laboratoire sont nombreux, avec des connaissances théoriques en sciences cognitives, ainsi que des compétences en traitement de données et en intelligence artificielle.

Cette thèse a été menée dans le cadre du partenariat entre ces deux entités, avec l'objectif de contribuer à la prise en compte du facteur humain dans le contexte du service client.

I.2 Problématique générale

Les émotions jouent un rôle clé dans les communications inter-personnelles. Selon, Dacher Keltner et Haidt (1999), elles permettent d'exprimer des intentions et des besoins à ses interlocuteurs. Une bonne compréhension des émotions de son interlocuteur est essentielle pour réussir à se coordonner et s'adapter à celui-ci. Par conséquent, les émotions sont indispensables dans le domaine de la relation client, pour faciliter la prise en charge des clients par les conseillers. Elles rentrent aussi en compte dans la satisfaction globale des clients. En effet, cette dernière est à la fois conditionnée par l'atteinte de l'objectif désiré (résolution du problème, réponses aux questions, achat d'un produit...), mais aussi par l'expérience d'émotions positives au cours de l'échange. De plus, considérer les émotions des conseillers clients constitue

un élément déterminant pour estimer le bien-être ou la difficulté ressentie suite aux interactions. Accorder une plus grande place aux émotions dans ce cadre est donc essentiel. Cela représente un enjeu global pour les entreprises qui désirent apporter une valeur ajoutée à leurs services.

Les émotions peuvent s'exprimer par de nombreux moyens. Ainsi, diverses informations peuvent nous aider à identifier les émotions d'une personne, comme les expressions verbales, non verbales ou le contexte. Cependant, ces informations ne sont pas véhiculées de la même manière selon le canal de communication utilisé. Ainsi, des échanges en face à face facilitent largement la compréhension des émotions, avec l'accès d'expressions non verbales visuelles (postures, expressions faciales) ou vocales (intonation). À l'inverse, des communications uniquement textuelles ou vocales limitent largement la quantité et la qualité des informations disponibles pour inférer les émotions de son interlocuteur. Des interactions médiées par un canal textuel, telles que des emails ou des messageries instantanées, sont donc caractérisées par une plus grande difficulté pour inférer l'état émotionnel d'un interlocuteur. Des moyens supplémentaires peuvent être mis en œuvre pour compenser cette perte de données. Cependant, il nous semble nécessaire de concevoir des solutions permettant de pallier ce manque d'information. Par exemple, l'utilisation de données complémentaires pourrait faciliter la reconnaissance automatique d'émotions, exprimées lors de l'usage de TIC.

Comme nous l'avons évoqué, de nouveaux supports de communication ont émergé au cours des dernières années. Cependant, la progression des smartphones est sans doute la plus remarquable, avec une explosion des ventes au cours des dix dernières années. Au niveau mondial, plus de smartphones sont vendus que d'ordinateur en 2021 et cette tendance est encore plus forte dans les pays en voie de développement (Statcounter, 2021). Ce type de support concrétise l'informatique ubiquitaire, définie par des usages connectés, mobiles et accessibles de partout. Il s'agit désormais d'un objet incontournable de la vie quotidienne. Outre des microphones ou des caméras, les smartphones embarquent tout un ensemble de capteurs qui pourrait permettre de représenter l'activité des utilisateurs (accéléromètre, écrans tactiles, puce GPS...). Les smartphones sont particulièrement utilisés pour communiquer, notamment par l'intermédiaire de messageries instantanées. De nombreuses applications sont dédiées aux échanges entre utilisateurs, mais aussi dans le cadre de service client. C'est pourquoi grâce à leur popularité croissante et leur capacité d'enregistrer des données diversifiées, les smartphones nous semblent particulièrement appropriés pour étudier les émotions des utilisateurs en situation de communication.

Nous pouvons donc formuler une problématique générale à partir des éléments énoncés : **comment reconnaître automatiquement des états émotionnels exprimés par un utilisateur, lors d'une situation de communication sur smartphone?** Notre objectif final consiste à proposer une approche pour concevoir un système de

reconnaissance automatique d'émotions, en prenant le cas d'usage d'une messagerie instantanée.

I.3 Organisation du manuscrit

De nombreuses questions découlent de cette problématique générale et permettent d'assembler les éléments nécessaires pour y répondre.

Premièrement, nous consacrons le chapitre **II** aux aspects théoriques des émotions. En effet, les émotions constituent à elles seules un vaste sujet d'étude : bien que familières, elles sont complexes à définir. Il est donc nécessaire d'étudier la science des émotions et de réaliser une revue des théories dominantes, afin de sélectionner une définition pertinente pour notre contexte. Ainsi, nous cherchons dans un premier temps à répondre à la question suivante : qu'est-ce qu'une émotion ? Ensuite, pour déterminer comment se manifestent les émotions, il nous faut sélectionner des modèles théoriques, nous permettant de caractériser les phénomènes associés aux émotions. Enfin, nous nous posons la question suivante : comment différencier des états émotionnels ? Pour cela, nous explorerons les modèles de représentation et catégorisation des émotions de la littérature scientifique. Nous exposerons finalement les positionnements théoriques de cette thèse.

Deuxièmement, le chapitre **III** se focalise sur la relation entre les états émotionnels et les TIC, ainsi que les émotions dans le contexte expérimental. Nous présenterons ainsi la discipline de l'informatique affective, qui considère la prise en compte des états affectifs et émotionnels des humains comme déterminante, afin de faciliter les interactions homme-machine. C'est pourquoi, dans un premier temps, nous chercherons à répondre à la question suivante : en quoi l'informatique affective peut-elle s'appliquer au contexte du service client ? Nous présenterons ensuite la revue des principales modalités utilisées pour détecter ou reconnaître automatiquement des états émotionnels. Nous prendrons comme référence les publications issues de la discipline de l'informatique affective. Plus spécifiquement, nous nous intéresserons aux modalités issues des interactions homme-machine avec des TIC et plus particulièrement avec des smartphones. L'ensemble de ces éléments nous permettra de comprendre quels sont les types de données utilisables pour concevoir un système de reconnaissance automatique d'états émotionnels. Enfin, afin de diriger notre démarche pour constituer un corpus de données approprié, nous expliciterons les approches existantes permettant d'enregistrer des données expérimentales labellisées selon des émotions.

Troisièmement, le chapitre **IV** est dédié à la présentation du matériel et de la méthode. L'objectif de ce chapitre est de présenter notre démarche pour étudier les émotions dans un contexte expérimental et pour recueillir les données nécessaires

pour répondre à notre problématique. Dans un premier temps, nous décrivons les outils développés pour permettre la mise en place d'expérimentations dans un contexte d'utilisation de smartphone. Trois protocoles expérimentaux seront présentés, avec leurs objectifs et hypothèses. En effet, nous expliciterons plusieurs approches d'induction d'émotions, ainsi que les mesures proposées pour évaluer leur efficacité sur les sujets. Nous décrivons également les enregistrements réalisés à partir des capteurs d'un smartphone pour constituer une base de données labellisées, accompagnés de mesures supplémentaires à l'aide de capteurs physiologiques. Le recueil de l'ensemble de ces données a pour objectif de servir de base pour la mise en œuvre d'un système de reconnaissance automatique d'émotions. Une partie explicative sera consacrée au développement d'un système de reconnaissance automatique d'émotions multimodal. Les explications porteront sur les traitements des enregistrements bruts jusqu'à l'entraînement de modèles de classification par apprentissage automatique.

Quatrièmement, l'analyse des résultats sera abordée dans le chapitre V. Afin de vérifier l'induction d'émotions des trois expérimentations, nous mettrons en lumière les résultats obtenus à partir des différents outils de mesure utilisés. Enfin, nous présenterons les résultats obtenus à partir des données de capteurs de smartphone et des capteurs physiologiques. Nous décrivons l'évaluation de la robustesse des différents modèles de classification réalisés.

Finalement, nous ferons la synthèse des apports de ce travail de recherche pour répondre à la problématique générale dans le chapitre VI. De même, nous présenterons les perspectives possibles pour approfondir ces travaux.

Chapitre II

Science des émotions : Définitions, théories et modèles

Sommaire

II.1	Introduction	28
II.2	Historique des théories et débats	28
II.2.1	Origine et définitions du terme émotion	28
II.2.2	Les émotions avant le 19 ^e siècle	29
II.2.3	Approche darwiniste et néodarwiniste	31
II.2.4	Débat James-Lange versus Cannon-Bard	32
II.2.5	Approche cognitiviste	34
II.2.6	Approche socio-constructiviste	36
II.3	Théorie de l'évaluation cognitive	37
II.3.1	Origine de la théorie	37
II.3.2	Définitions et composantes	40
II.3.3	Évaluation cognitive	42
II.3.4	Réponses psychophysiologiques	47
II.3.5	Tendance à l'action	49
II.3.6	Expression motrice	50
II.3.7	Sentiment subjectif	52
II.3.8	Discussion	55
II.4	Classification et représentation des émotions	55
II.4.1	Taxonomie des états affectifs	55
II.4.2	Classification des émotions	57
II.4.3	Modèles discrets et catégoriels	58
II.4.4	Modèles dimensionnels	63
II.4.5	Discussion	66
II.5	Conclusion et positionnement	67

II.1 Introduction

Afin de répondre à notre problématique de recherche, il nous est nécessaire d'établir un cadre théorique sur la notion d'émotions. Ainsi, l'objectif de ce chapitre est de répondre aux questions suivantes :

- Qu'est-ce qu'une émotion et comment la définir ?
- Comment modéliser les phénomènes qui sont inclus dans cette notion ?
- Comment différencier et classifier les différents types d'émotion ?

Bien que le terme d'émotion soit couramment utilisé dans le quotidien et que nous en fassions tous l'expérience au cours de nos vies, il reste très compliqué de donner une définition claire sur cette notion pour le moins abstraite. Ainsi, nous ne pouvons généralement qu'observer les « symptômes » des émotions sur nous et sur les autres, sans pour autant en comprendre l'essence. Les chercheurs du domaine de la science des émotions admettent eux-mêmes qu'il n'existe pas une définition unanime et absolue pour décrire ce concept et les mécanismes qu'il englobe. La science des émotions n'est pas elle-même homogène. En effet, elle inclut de nombreuses disciplines qui ont chacune tenté de comprendre ce phénomène à partir de leurs outils et de leurs connaissances. Ainsi, les émotions ont été abordées du point de vue du comportement par la psychologie cognitive ou bien par l'intermédiaire des réactions psychophysologiques par la biologie. Il en résulte donc de nombreuses approches et points de vue pour tenter d'expliquer ce que sont les émotions.

Dans le cadre de cet état de l'art, nous allons, dans un premier temps, réaliser une revue historique des théories sur les émotions, afin de suivre l'évolution des idées et des débats sur le sujet. Dans un deuxième temps, nous expliciterons plus en détail la théorie de l'évaluation cognitive et des composantes qu'elle intègre. Enfin, nous présenterons les différentes approches pour différencier, classifier et représenter les émotions.

II.2 Historique des théories et débats

II.2.1 Origine et définitions du terme émotion

Si l'on s'en tient au dictionnaire, nous pouvons citer la définition proposée sur le site du Centre National de Ressources Textuelles et Lexicales :

« Conduite réactive, réflexe, involontaire vécue simultanément au niveau du corps d'une manière plus ou moins violente et affectivement sur le mode du plaisir ou de la douleur. » (CNTRL, 2012)

L'étymologie du mot émotion provient des termes latins *motio* signifiant « mouvement, le trouble ou le frisson de fièvre » et *e* signifiant « qui vient de ». Selon l'historien

Alain Corbin, le terme émotion est donc initialement associé à des mouvements physiques et n'apparaît dans la langue française qu'à partir du 15^e siècle (Corbin *et al.*, 2016a). Il peut être utilisé aussi bien pour désigner des perturbations du corps que des agitations de foule ou de la météo.

L'association du terme émotion au mental se fera graduellement au fil de l'histoire. D'après Thomas Dixon (2012a), c'est au début du 19^e siècle que le terme émotion acquiert un statut à part entière en tant que concept propre des sciences de l'esprit et remplace les notions de passion ou affection. Le philosophe écossais Thomas Brown serait à l'origine de cette distinction en différenciant clairement les émotions des pensées ou des sensations. Ainsi, il définit les émotions comme un sentiment vif (*vivid feelings*) causé par la considération mentale d'un objet (Dixon, 2012a, 2012b). Toujours au cours du 19^e siècle, un autre théoricien, Charles Bell affirme qu'une émotion est un mouvement de l'esprit et peut devenir visible à travers des signes extérieurs au travers du visage et du corps. Ces pensées ont fortement influencé les scientifiques des décennies suivantes qui ont ainsi formulé de nombreuses définitions et théories afin d'appréhender ce phénomène.

De nos jours, nous pouvons compter un nombre impressionnant de définitions de l'émotion. Il y a déjà 40 ans, en 1981, Anne et Paul Kleinginna ont listé pas moins de 91 définitions dans la littérature scientifique et ont tenté de les catégoriser (Kleinginna & Kleinginna, 1981). Depuis, même si leur travail n'a pas été reproduit récemment, nous pouvons imaginer que le nombre de définitions proposées a augmenté significativement, ajoutant encore plus de profondeur et de diversité à ce concept.

Il faut bien discerner les émotions des autres termes du même registre, dans le langage courant, comme les affects, les humeurs ou les sentiments. Les affects, par exemple, concernent des dispositions basiques positives ou négatives : « Disposition affective élémentaire » (CNTRL, 2012). Tandis que l'humeur fait référence à un état affectif plus long sur la durée, plus faible en intensité et qui n'a pas nécessairement de cause définie : « Disposition de caractère, état de réceptivité dans lequel se trouve une personne à un moment donné. » (CNTRL, 2012). En effet on peut par exemple se réveiller de bonne ou de mauvaise humeur. Enfin, le sentiment comporte un aspect subjectif et conscient : « État affectif complexe, assez stable et durable, composé d'éléments intellectuels, émotifs ou moraux, [...] » (CNTRL, 2012). La distinction entre les différents termes affectifs sera abordée de manière plus approfondie dans la section II.4.

II.2.2 Les émotions avant le 19^e siècle

Selon l'historien Corbin, il n'existe pas d'équivalent direct à la notion d'émotion dans l'antiquité grecque. Les Grecs font référence aux passions (*pathè*) qui désignent

des phénomènes internes affectant l'âme, aux côtés des facultés et de la façon d'être. Ces passions, que ce soit le plaisir ou la souffrance, sont subies par l'individu. Le terme *pathos* signifiant « ce que l'on éprouve » sera d'ailleurs à l'origine du mot pathologie dans la langue française (CNTRL, 2012). Sans proposer de classification fixe, Aristote intègre dans les passions les manifestations suivantes : le désir, la colère, la peur, le courage, l'envie, la joie, l'amitié, la haine, le désir ardent, la jalousie et la pitié. Selon lui, les passions mettent l'âme en mouvement. On peut rapprocher cette notion de mouvement à l'étymologie du mot émotion présentée précédemment (Corbin *et al.*, 2016a).

Bien que le mot émotion provienne du latin, on ne peut pas directement faire référence à ce terme pour les Romains non plus. À l'image d'Aristote, Cicéron établit des listes de passions opposées, telles que la volonté stable face au désir ou la précaution face à la crainte. De plus, les Romains ont un vocabulaire étendu pour se référer aux modifications d'états internes : *passio*, *affectio*, *perturbatio*... Autant de termes qui font référence aux mouvements subis par l'âme (Corbin *et al.*, 2016a).

Il serait complexe de synthétiser la vision des émotions sur la période du Moyen Âge au vu de l'étendue de cette ère. Le terme émotion ne sera utilisé que plus tardivement. En revanche, au 13^e siècle on pourra trouver les expressions comme *esmouvoir le coeur* ou *estre emu*. *Esmouvoir* reste lié à la notion de mouvement physique dans la continuité du latin *moveo* signifiant mouvoir. Ainsi, l'expression *esmouvoir le coeur* fait référence à des mouvements affectant l'organe du cœur, amenant une notion de trouble intérieur (Corbin, Courtine & Vigarello, 2016a).

Le 16^e siècle voit apparaître un usage plus courant du terme émotion, ainsi qu'une distinction entre les « émotions du corps » et les « émotions de l'esprit » ou « émotions de l'âme ». Descartes rapproche les émotions de l'âme aux pensées et les distingue des sensations du corps liées aux sens comme les odeurs, le toucher ou la vision. À cette époque Descartes ne distingue pas explicitement les émotions des passions, mais amorce une vision des émotions plus intériorisée et moins centrée sur le corps (Corbin, Courtine & Vigarello, 2016a).

La période des Lumières voit s'approfondir cette réflexion sur la place des émotions entre le corps et l'esprit. Dans l'*Encyclopédie* des Lumières, on commence à distinguer une sensibilité physique d'une sensibilité morale, cette dernière étant unique à l'Homme. L'idéologie de l'empirisme suggère une continuité de la sensibilité physique vers les pensées et les émotions. De son côté, Jean-Jacques Rousseau propose de différencier l'être *sensitif*, lié au corps et à l'organisme, de l'être *sensible*, lié au spirituel. On s'interroge alors sur le statut des émotions et de sa dépendance par rapport au corps (Corbin *et al.*, 2016b).

Les émotions prennent ensuite une signification plus proche de notre vision actuelle

au 19^e siècle avec des penseurs comme Thomas Brown et Charles Bell que nous avons présentés précédemment. Ainsi, les émotions deviennent un objet d'étude à part entière de la part du monde scientifique.

II.2.3 Approche darwiniste et néodarwiniste

À la fin du 19^e siècle, Charles Darwin (1877) a été un des premiers chercheurs à marquer la recherche sur les émotions avec la publication de *L'expression des émotions chez l'homme et l'animal*. Cet écrit s'inscrit dans la continuité des travaux du biologiste sur sa théorie de l'évolution de l'univers du vivant. Ainsi, Charles Darwin applique sa vision évolutionniste sur les expressions des émotions, en les considérant comme des comportements sélectionnés au sein d'une espèce animale pour s'adapter à son environnement. En observant les expressions faciales ou corporelles, Darwin constitue une classification des émotions et associe des rôles utilitaires aux expressions de ces émotions. Ces expressions seraient issues de comportements primaires hérités de nos ancêtres et présents chez d'autres espèces animales à des niveaux plus rudimentaires, chez les grands singes par exemple. En conséquence, Darwin attribue aux émotions une dimension adaptative et une dimension universelle pour chaque espèce. De plus, il met en avant le rôle communicatif des expressions faciales comme un moyen d'avertir ses congénères d'une situation. L'expression de peur sur un visage suggère par exemple un danger imminent aux autres membres d'un groupe (Christophe, 1998 ; Nugier, 2009). Les travaux de Charles Darwin auront une influence significative sur la recherche sur les émotions et seront à l'origine de l'approche néodarwiniste au cours du 20^e siècle.

À partir des années 60, des chercheurs tels que Paul Ekman, Sylvain Tomkins ou Carroll Izard reprendront cette approche en associant une émotion à un pattern comportemental et physiologique spécifique. Selon ces chercheurs, il existe chez l'Homme un ensemble d'émotions primaires ou émotions de base, discrètes et universelles. Ces émotions seraient d'origine génétique et répondraient à des besoins concrets pour la survie des individus d'une espèce (Nugier, 2009). Tomkins propose un ensemble de neuf émotions primaires à partir d'expressions faciales spécifiques : l'intérêt, la joie, la surprise, l'anxiété, la peur, le mépris, le dégoût, la honte et la colère (Christophe, 1998 ; Tomkins, 1984). Dans le modèle de Tomkins, chacune des émotions est activée en fonction de l'intensité de la stimulation neuronale et en fonction de la durée de cette activation.

Paul Ekman propose d'étudier l'universalité de six émotions fondamentales : la joie, la colère, la tristesse, le dégoût, la surprise et la peur (Ekman & Friesen, 1971). Pour cela il propose à des individus de cultures diverses, n'ayant pas été affectés par les médias occidentaux, de choisir quelle expression faciale doit être associée à une histoire typique d'une émotion parmi les six émotions discrètes (par exemple pour la tristesse : apprendre qu'un de ses enfants est mort). Les conclusions de ses études

soutiendraient que les patterns d'expressions faciales seraient associés aux émotions discrètes proposées indépendamment de l'origine culturelle des participants.

Robert Plutchik (1984, 2001) considère les émotions à partir de « séquences comportementales prototypiques ayant chacune une fonction adaptative distincte » (Christophe, 1998). En effet, il identifie un ensemble de huit émotions primaires, chacune caractérisée par un ensemble de comportements expressifs et évaluatifs, afin de répondre à une fonction adaptative spécifique. Plutchik se différencie des autres chercheurs en intégrant dans son modèle des dimensions pour affiner les nuances des émotions. Il propose de représenter les émotions en fonction de leur intensité (le chagrin est d'intensité plus forte que la tristesse), par polarité (la tristesse s'oppose à la joie) et par similitude (la peur est similaire à la surprise). Par l'intermédiaire de la dimension d'intensité, l'auteur soumet l'idée d'émotions secondaires dérivées des émotions primaires. Ce modèle de représentation sera explicité plus en détail dans la sous-section II.4.3.

En résumé, l'approche évolutionniste des émotions, amorcée par Charles Darwin, s'appuie sur l'utilité fonctionnelle des émotions ainsi que sur leur caractère inné et universel. Selon les auteurs, le nombre d'émotions basiques considérées dans les modèles est variable, cette diversité s'explique notamment par l'utilisation de procédures méthodologiques différentes (prise en compte des expressions faciales, corporelles, fonctionnelles...). Cette approche des émotions ne fait pas l'unanimité. En effet, beaucoup de chercheurs n'adhèrent pas à la thèse selon laquelle les émotions sont innées. De plus, les chercheurs de ce courant de pensée se sont vu reprocher leurs méthodes. Notamment, de ne prendre en compte qu'un certain nombre d'indices externes, comme les expressions faciales, pour créer leur classification d'émotions basiques. La notion d'émotion basique elle-même est décriée par certains auteurs qui la considèrent comme trop simpliste pour représenter les phénomènes émotionnels (Nugier, 2009).

II.2.4 Débat James-Lange versus Cannon-Bard

Alors que de Darwin s'est principalement concentré sur le rôle et l'expression des émotions, en 1884 William James, psychologue et philosophe américain, expose une première théorie scientifique sur le fonctionnement des émotions en publiant *Qu'est-ce qu'une émotion?* (James, 1884). Dans cet article, James avance l'idée qu'un évènement pertinent pour l'organisme entraîne des modifications corporelles spécifiques, puis que la perception de ces changements est à l'origine de l'émotion. Une émotion peut donc être caractérisée par la séquence représentée sur la figure II.1 (Sander, Scherer *et al.*, 2014).

Cette séquence se fonde sur le principe d'une rétroaction corporelle, l'émotion est donc l'expérience subjective des modifications perçues dans notre organisme suite à

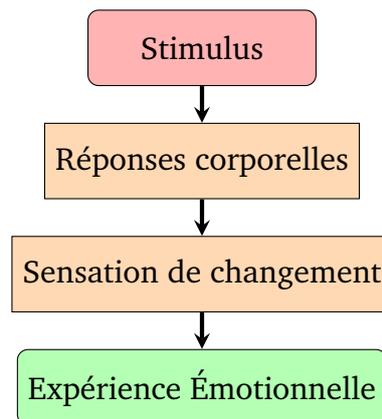


FIGURE II.1 – Séquence temporelle de la théorie de James (1884)

un évènement (James, 1884). James dépeint sa théorie avec un exemple célèbre dans le domaine des émotions : un promeneur croise un ours dans une forêt. La vue de cet ours provoque chez la personne des tremblements aux genoux, puis le déclenchement d'une course pour s'échapper. La perception des changements physiques dans le corps permet au promeneur d'identifier qu'il a peur.

En 1885, à peine une année plus tard, Carl Lange présente une théorie similaire dans un traité sur les émotions (Lange, 1885). En cherchant à différencier les phénomènes physiques spécifiques de plusieurs émotions, il identifie deux systèmes qui en seraient à l'origine : le système sensori-moteur et le système végétatif (ou système nerveux autonome) (Christophe, 1998). En plus de s'accorder sur la séquence et sur le fait que les émotions découlent de la perception subjective des changements corporels, les deux scientifiques s'entendent sur le fait qu'une émotion spécifique possède une réponse corporelle distincte. De fait, il existe de nombreux points communs entre les théories James-Lange et darwiniste, en effet les deux considèrent les émotions comme étant un phénomène biologique et inné. La théorie James-Lange donnera ce qu'on appelle de nos jours la position dite *périphaliste*, elle sera révolutionnaire pour l'époque et posera une base de réflexion sur le fonctionnement des mécanismes émotionnels.

Au début du 20^e siècle, Walter Cannon et son collaborateur Philipp Bard réagissent à la théorie *périphaliste*. En 1927, Cannon publie un argumentaire contre la théorie James-Lange (Cannon, 1927). Premièrement, à partir d'expériences sur des animaux, Cannon et Bard affirment que la réaction émotionnelle n'est pas altérée lorsque le système végétatif est déconnecté du cerveau. Deuxièmement, ils observent des réponses physiologiques similaires entre des états émotionnels et non émotionnels dans le système végétatif (fièvre, digestion...), infirmant le fait que les émotions ont des réponses physiologiques spécifiques (Sander, Scherer *et al.*, 2014). Enfin, à partir d'expériences de stimulation du système végétatif avec des injections d'adrénaline (Marañón, 1924), ils affirment que la stimulation de ce système ne déclenche pas

d'émotions particulières. Ainsi ils considèrent que le système périphérique n'est pas nécessaire pour ressentir des émotions et n'en est donc pas la source. Ils expliquent ce point de vue en exposant le fait que les réponses du système végétatif sont trop lentes et trop peu sensibles pour être à l'origine des émotions (Sander, Scherer *et al.*, 2014).

En réaction Cannon propose une autre théorie, dite *centraliste*. En effet, selon des observations neurophysiologiques, l'émotion serait issue du système nerveux central et plus particulièrement du thalamus. Ainsi, les modifications du système périphérique ne seraient que des conséquences de l'activation du thalamus et non spécifiques à des états émotionnels ou non émotionnels. Cannon considère le thalamus comme le centre de l'expérience affective (Christophe, 1998). Il suggère donc la séquence temporelle représentée sur la figure II.2.

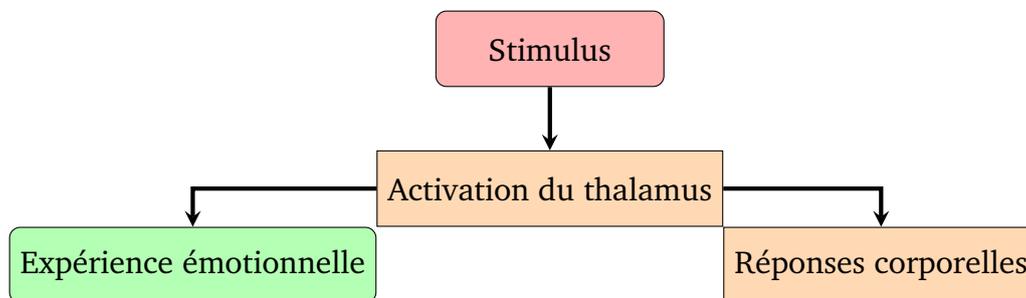


FIGURE II.2 – Séquence temporelle de la théorie de Cannon (1927)

L'argumentaire de Cannon a été critiqué du fait qu'il reposait essentiellement sur le système végétatif ou viscéral et ne prenait pas en compte le système sensori-moteur pourtant inclus dans la théorie James-Lange. Néanmoins, cette opposition entre *périphaliste* et *centraliste* a mis au cœur du débat des problématiques essentielles. Les questions sur la localisation des émotions dans le corps et sur la séquence entre les réactions physiologiques du corps et l'expérience subjective seront essentielles pour établir les théories à venir (Nugier, 2009).

II.2.5 Approche cognitive

À la suite du débat entre *périphalistes* et *centralistes*, l'idée d'une relation entre émotion et cognition prend une place centrale dans les discussions. Véronique Christophe propose la définition générale suivante pour expliquer ce qu'est la cognition :

« [...] L'ensemble des connaissances que le sujet a du monde et de lui-même, et l'ensemble des processus qui permettent d'acquérir ces connaissances. »
(Christophe, 1998)

En 1962, Stanley Schachter est parmi le premier à proposer une théorie de l'émotion avec une composante cognitive (Moors, 2009 ; Schachter & Singer, 1962). Il désapprouve en partie la théorie de James-Lange, car il ne considère pas que des réponses

physiologiques spécifiques soient associées à des émotions ni qu'elles en soient l'origine. Cependant, il est d'accord avec le principe qu'une activation physiologique précède l'expérience émotionnelle, de manière non spécifique (Moors, 2009). Schachter propose donc une théorie *bifactorielle*, en ajoutant à l'activation physiologique une composante cognitive. Pour qu'il existe l'expérience d'une émotion, il doit donc y avoir une activation corporelle suffisante associée à une composante cognitive pour attribuer à la situation vécue un sentiment subjectif. Par exemple, croiser un chien dangereux ou une personne chère causerait, initialement, une activation physique similaire jusqu'à ce que l'identification de la situation amène l'expérience émotionnelle de la peur ou de la joie. La séquence temporelle est représentée sur la figure II.3.

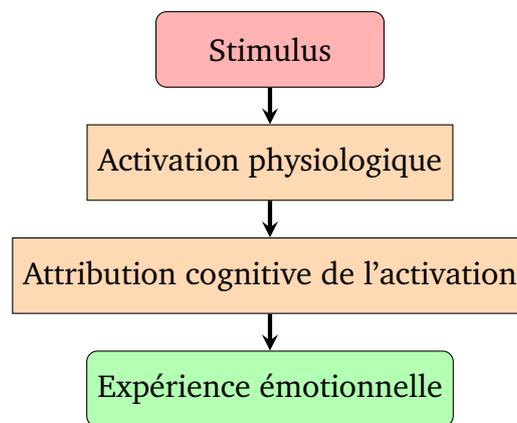


FIGURE II.3 – Séquence temporelle de la théorie de Schachter-Singer (Schachter & Singer, 1962)

Pour approfondir cette théorie, Schachter propose d'ajouter la notion de « besoin évaluatif ». En effet, soit l'individu peut immédiatement identifier son état émotionnel à partir des marqueurs physiologiques s'il possède une cognition explicative en fonction de son vécu, soit l'individu doit alors évaluer la situation afin de comprendre son état. Si l'individu ne trouve pas une cause externe évidente pour expliquer son activation physiologique, alors il réalisera qu'il ressent une émotion et cherchera à l'évaluer en comparant les réactions émotionnelles d'autrui (Christophe, 1998). Ces processus cognitifs ne se font pas de manière consciente, mais par des traitements implicites (Sander, Scherer *et al.*, 2014). Schachter a tenté de confirmer sa théorie à l'aide d'expériences, en injectant de l'épinéphrine, un stimulant, à des participants et en induisant des émotions. Un groupe de participants était correctement informé des effets de l'épinéphrine sur le corps, tandis que d'autres groupes n'étaient pas ou mal informés (Schachter & Singer, 1962). Les groupes de sujets mal ou peu informés auraient ressenti des émotions plus fortes selon les chercheurs, prouvant l'hypothèse qu'une activation physiologique forte et non connue entraîne le participant à s'attribuer un état émotionnel par l'intermédiaire d'un processus cognitif.

La théorie *bifactorielle* a été amplement utilisée à partir des années soixante jusqu'aux années quatre-vingt-dix. Cependant la théorie et les expériences réalisées ont été largement critiquées par la suite. Une des principales failles dans la théorie était l'absence d'explications sur la cause des activations physiologiques. Une autre critique vise l'absence de la prise en compte d'un lien causal entre l'activation physiologique et l'attribution cognitive (Leventhal, 1980). Les expériences ont été aussi remises en question à cause de la méthodologie et de la difficulté à répliquer les expériences (Schachter & Singer, 1962). Malgré ces critiques, les travaux de Schachter et Singer ont permis de mettre en avant l'importance de la cognition et de l'expérience personnelle dans l'attribution des expériences émotionnelles vécues et ont donc marqué l'histoire de la recherche sur les émotions. L'apport de cette théorie permettra de faire évoluer l'approche cognitiviste sous sa forme plus actuelle avec les théories de l'évaluation cognitive que nous expliciterons dans la section II.3.

II.2.6 Approche socio-constructiviste

Les théories socio-constructivistes proposent une approche différente des théories darwiniste et jamesienne axées sur un déterminisme biologique. James Averill (1985) réfute l'idée que les émotions soient universelles et innées. Au contraire, il considère que les réactions émotionnelles des individus sont des constructions sociales et culturelles. L'expression d'une émotion serait le produit de normes et d'attentes sociales acquises avec l'expérience, la sociabilisation et le statut social. Il en résulte des schémas cognitifs amenant à traiter certains événements avec des comportements spécifiques (Christophe, 1998). Ces schémas seraient automatiques et rempliraient un rôle social. Ainsi, les émotions provoquées par une même situation dépendront fortement de l'interprétation, elle-même conditionnée par la culture et l'expérience (Nugier, 2009).

Le concept d'émotion de base est particulièrement décrié par les partisans du socio-constructivisme. Ils avancent la critique suivant laquelle les émotions primaires, proposées par certains chercheurs, sont souvent issues d'une certaine langue et ne s'appliquent donc pas nécessairement à des concepts issus d'autres cultures, elles ne peuvent donc pas être considérées comme universelles. Les listes d'émotions de base constituées seraient donc biaisées par la culture d'origine des chercheurs. James Russell (1991) et Anna Wierzbicka (1986) mettent en évidence, à partir d'études ethnographiques, la difficulté de traduire les émotions basiques en des termes simples entre les langues et d'appliquer exactement les mêmes significations aux traductions lorsqu'elles existent. Wierzbicka prend l'exemple du terme dégoût, souvent inclus dans les modèles d'émotions basiques, qui ne possède pas de traduction directe dans la langue polonaise. En effet, les immigrés polonais aux États-Unis ont donc emprunté le terme anglais *disgust* dans leur langue (*dysgust*) pour désigner le même concept. Pourtant la langue polonaise contient des termes signifiant la répulsion ou des concepts analogues au dégoût, mais aucun, selon les locuteurs polonais, qui n'englobe le même

sens que le terme anglais. Ainsi, même sur des cultures pourtant similaires, car issues de langues indo-européennes, nous pouvons noter des différences essentielles sur des émotions perçues comme universelles par l'approche évolutionniste.

L'apport de l'approche socio-constructiviste met en lumière le rôle social des émotions ainsi que les différences culturelles qui les conditionnent. À l'inverse de l'approche néonéodarwiniste elle se propose d'étudier les différences entre les cultures plutôt que de chercher des similitudes. Ainsi, les émotions ne sont plus seulement étudiées à un niveau individuel, mais aussi à un niveau sociétal, apportant un nouveau regard sur cet objet d'étude.

II.3 Théorie de l'évaluation cognitive

II.3.1 Origine de la théorie

Dans la continuité de l'approche cognitiviste, présentée dans la sous-section [II.2.5](#), la deuxième moitié du 20^e siècle est marquée par l'apparition de théories donnant une place centrale à la cognition dans les processus menant à des états émotionnels. Cependant, plusieurs problématiques se posent quant à la place de la cognition dans le processus émotionnel et sur les informations traitées. Selon Schachter et Singer, un traitement cognitif s'effectue après une activation physiologique (Schachter & Singer, [1962](#)). Cependant, cela soulève la problématique de ce qui cause en premier lieu cette activation physiologique suite à un stimulus interne ou externe.

En 1960, Magda Arnold ([1960](#)) propose qu'une évaluation cognitive (ou *appraisal* en anglais) intervienne donc au début de la séquence temporelle directement après la survenue d'un stimulus. Cette évaluation cognitive peut se traduire comme une « appréciation d'un évènement particulier » (Mahboub, [2011](#)). Cette évaluation serait à l'origine d'une première analyse de la situation, produisant une « tendance à l'action », à savoir une préparation vers une action liée à la situation (Christophe, [1998](#)). Cette préparation à l'action doit être distinguée de l'action en elle-même, qui est plus tardive et considérée comme une conséquence de l'état émotionnel et non comme une composante. Selon Arnold, cette tendance à l'action est suivie par le déclenchement de modifications physiologiques et comportementales, causées par des impulsions nerveuses en provenance du système nerveux central vers le système périphérique. Dans une boucle de rétroaction, les transformations physiologiques provoquées sont perçues par le cortex. Cette rétroaction met en œuvre une nouvelle évaluation cognitive de la situation, permettant de qualifier l'expérience émotionnelle, à partir des modifications physiologiques perçues, ainsi que la prise en compte de la situation. La figure [II.4](#) résume la séquence temporelle de la théorie d'Arnold.

Ainsi, Arnold et Schachter s'accordent sur la présence d'un traitement cognitif

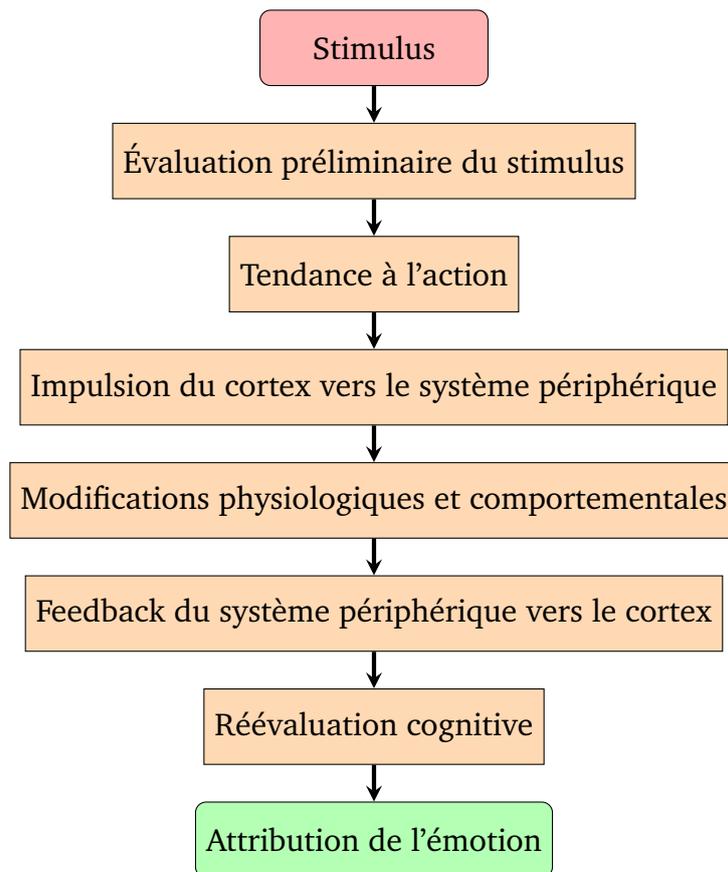


FIGURE II.4 – Séquence temporelle de la théorie de l'évaluation cognitive d'Arnold (1960)

nécessaire à l'expérience émotionnelle et indispensable pour qualifier l'émotion, mais n'identifient pas ce processus à la même étape (Moors, 2009). Dans la théorie de l'évaluation cognitive, la qualification finale de l'émotion n'intervient qu'à la fin du processus. D'une certaine manière, cette théorie de l'évaluation cognitive concilie les rôles du système nerveux central et périphérique du débat James-Lange vs Cannon-Bard (voir la sous-section II.2.4). Les deux systèmes seraient alors impliqués dans le processus d'expérience émotionnelle par le prisme de plusieurs évaluations cognitives. Arnold est considérée comme une précurseuse de ce que l'on nomme à présent : les théories de l'évaluation cognitive ou théorie de *l'appraisal*.

Les théories issues du courant de *l'appraisal* mettent en évidence de nouvelles composantes dans le processus émotionnel. Ainsi, la notion de tendance à l'action, présentée précédemment, est au cœur de la théorie de Nico Frijda (Frijda *et al.*, 1986). Elle apparaît comme une composante motivationnelle qui interagit en parallèle de la composante cognitive pour analyser la situation et déclencher une action appropriée pour l'intérêt de l'individu. Pour Frijda cette préparation à l'action est à l'origine de

l'émotion et de l'initialisation de patterns physiologiques et comportementaux en fonction des ressources de l'organisme (Aue Seil, 2009).

Richard Lazarus (1968) affirme lui aussi l'importance d'une composante motivationnelle en plus de l'évaluation cognitive pour le déclenchement de l'expérience émotionnelle. Il considère ainsi que la modification de la relation entre l'individu et son environnement est le sujet de l'interprétation. L'aspect motivationnel s'articule par la prise en compte des besoins et objectifs de l'individu d'une part et des moyens, ressources et contraintes qui s'appliquent à l'individu pour atteindre ses objectifs d'autre part. Lazarus identifie plusieurs types d'évaluations cognitives (Christophe, 1998 ; Sander & Scherer, 2009) :

- Des évaluations primaires : elles portent sur la pertinence du stimulus pour l'individu, par rapport à ses propres objectifs, ses besoins et l'environnement. Une situation peut alors être évaluée comme non pertinente et ne pas susciter de réaction ou à l'inverse être jugée comme pertinente. Si une situation est pertinente, l'individu va alors évaluer à quel point l'évènement est bénéfique (congruent avec les objectifs et besoins) ou néfaste (non congruent avec les objectifs et besoins).
- Des évaluations secondaires : elles sont centrées sur l'appréciation des ressources de l'individu pour faire face aux conséquences de cette situation. Les évaluations secondaires dépendent des stratégies mises en place pour répondre à la situation.
- Des réévaluations : elles consistent à mettre à jour les évaluations primaires et secondaires en fonction de l'évolution de la relation entre l'individu et son environnement.

La combinaison des types d'évaluations cognitives permet de différencier les émotions entre elles. Elles mettent notamment en lumière les différences inter et intra-individuelles et expliquent la diversité des réactions par rapport à un même évènement. Par exemple, un chasseur expérimenté ne réagira pas de la même manière en croisant un ours dans une forêt par rapport à un promeneur non préparé. Le premier pourra percevoir un gain devant cet évènement, s'il estime qu'il a les ressources nécessaires pour gérer cette situation, au contraire le deuxième percevra un danger et pourra se sentir démuné.

En conclusion, les théories des évaluations cognitives se distinguent par plusieurs paramètres. Premièrement, la notion de *appraisal* consistant à une prise d'informations de la situation portant sur la relation entre l'individu, l'environnement et les évènements déclencheurs. Cette évaluation prend en compte une large gamme de critères pour répondre aux besoins de l'individu, ce qui implique donc une dimension motivationnelle. Deuxièmement, la mise en avant de composantes qui interagissent à partir de cette évaluation cognitive. Selon les chercheurs, la liste des composantes impliquées dans les émotions n'est pas identique et ne s'articule pas de la même

manière. Ces composantes permettent de former des modèles *multicomponentiels* intégrant des sous-systèmes distincts et des relations entre ces sous-systèmes afin de représenter la complexité du processus émotionnel.

II.3.2 Définitions et composantes

Les théories de l'évaluation cognitive offrent une vision moderne des émotions et permettent d'étudier sous plusieurs aspects la complexité de ce phénomène. Dans le cadre de cette thèse, nous avons choisi de fonder notre base théorique sur le modèle *multicomponentiel* de Scherer, issu du courant des théories de l'évaluation cognitive. Avant d'expliquer plus en détail ce modèle, il nous semble essentiel de prendre du recul sur le concept d'émotion et sur sa complexité. En effet, en introduction du livre *Traité psychologique des émotions*, David Sander et Klaus Scherer établissent le constat suivant (Sander & Scherer, 2009) :

« L'émotion est un concept hypothétique qui ne peut pas être observé directement, mais dont on infère l'existence à partir d'un nombre d'indices. »
(Sander & Scherer, 2009)

Cette affirmation nous oblige à envisager une approche pragmatique pour étudier les émotions dans notre contexte. À l'image de nombreuses notions issues des sciences humaines et particulièrement en psychologie, les émotions constituent en elles-mêmes un concept abstrait. En effet, il n'est pas possible d'observer directement une émotion, d'autant plus que le concept même d'émotion reste compliqué à définir. Comme nous l'avons vu dans la partie précédente, les caractéristiques des émotions, les processus qu'elles impliquent et leurs articulations ont été et sont encore vivement débattus dans la communauté scientifique. Au vu de la complexité du problème, il nous semble alors essentiel de nous appuyer sur une définition pragmatique d'une part et sur un modèle constitué d'indices concrets, observables et mesurables d'autre part.

Pour répondre à la complexité des émotions, Scherer propose donc de définir les émotions à partir d'un certain nombre de sous-systèmes distincts, fonctionnels et interconnectés. Il propose alors la définition suivante :

« Un épisode de changements corrélés et synchronisés de l'état de tous ou de la plupart des cinq sous-systèmes de l'organisme en réponse à l'évaluation d'un stimulus ou d'un événement interne ou externe évalué comme pertinent pour les préoccupations majeures de l'organisme. » (D. M. Grandjean & Scherer, 2009 ; Scherer, 2005).

Selon le modèle des *processus composants* de Scherer ces sous-systèmes sont au nombre de cinq. La plupart ont été déjà évoqués, au moins partiellement, dans les théories présentées précédemment. L'ensemble des cinq composantes, leurs fonctions

TABLE II.1 – Liste des composantes du modèle des processus composants (D. M. Grandjean & Scherer, 2009 ; Scherer, 2005)

Nom simplifié de la composante	Fonctions	Composantes	Sous-systèmes organiques (substrat majeur)
Évaluation cognitive	Événements et stimulus	Composante cognitive	Traitement de l'information (SNC)
Réponses psychophysiologiques	Système de régulation	Composante périphérique efférente	Support (SNC, SNE, SNA)
Tendance à l'action	Préparation et direction de l'action	Composante motivationnelle	Exécutif (SNC)
Expression motrice	Communication des réactions et des intentions comportementales	Composante expressive motrice	Action (SNS)
Sentiment subjectif	Contrôle et interactions états internes-environnement	Composante du sentiment subjectif	Moniteur (SNC)

SNC : système nerveux central ; SNE : système neuro-endocrinien ; SNA : système nerveux autonome ; SNS : système nerveux somatique.

et leur localisation dans un système organique, sont détaillés dans le tableau II.1.

Les réponses psychophysiologiques, les expressions corporelles et comportementales et les expressions motrices sont les composantes qui ont été le plus longtemps associées à des aspects des émotions. Il s'agit en effet des modalités les plus facilement observables lorsqu'une émotion se produit. La tendance à l'action, qu'il faut bien différencier de l'action en elle-même, ne sera incluse qu'à partir des travaux de Frijda *et al.* (1986) et d'Arnold (1960). La dimension cognitive est incarnée par l'*appraisal* qui constitue un processus nécessaire à l'émotion selon cette théorie (Scherer, 2005).

Il est important de noter la dimension temporelle dans ce modèle. Ainsi, Sander et Scherer (2009) préfèrent parler « d'épisode émotionnel » plutôt que « d'état émotionnel » car une émotion est un processus dynamique d'une durée limitée. De plus, chacun de ces sous-systèmes peut être considéré comme indépendant des autres dans une situation normale, n'impliquant pas d'émotions particulières. Il y a une émotion seulement quand plusieurs de ces sous-systèmes se modifient en même temps de manière synchronisée. Dans l'exemple du promeneur qui croise un ours en forêt : avant qu'il ne croise l'ours et qu'il ne vive un épisode émotionnel, l'ensemble de ces sous-systèmes évoluent de manière indépendante. Sa respiration est adaptée à sa marche, ses muscles faciaux sont détendus et ses pensées dirigées vers la planification de son itinéraire ou de sa destination. Au moment où l'ours est perçu comme un danger,

sa respiration s'accélère, ses muscles faciaux se figent et ses pensées sont complètement dirigées vers l'objet de sa peur, dirigeant toutes ses ressources disponibles vers un même objectif de survie face au danger.

Le modèle des processus composants présente donc une flexibilité qui permet de rendre compte de la complexité des réactions émotionnelles et des mécanismes qu'elles provoquent face à des événements pertinents. Il permet notamment de mettre en évidence les différences individuelles et culturelles. Le fait d'impliquer plusieurs composantes distinctes avec un cadre défini offre la possibilité de mesurer des indices permettant de rendre compte d'une partie de l'expérience émotionnelle. Afin de mieux comprendre le modèle des processus composants, nous allons détailler les cinq composantes qui les composent dans les prochaines sous-parties.

II.3.3 Évaluation cognitive

L'évaluation cognitive est considérée comme la composante la plus essentielle pour le déclenchement et la différenciation d'un épisode émotionnel. Afin de mieux comprendre ce concept, Agnes Moors propose la définition suivante de l'*appraisal*.

« Appraisal is a process that detects and assesses the significance of the environment for well-being. » (Moors *et al.*, 2013)

Par cette définition et par *well-being* ou bien être, Moors intègre dans l'évaluation cognitive les notions de satisfaction ou d'obstruction des événements par rapport aux besoins, valeurs, objectifs de l'individu. Cette évaluation, comme présentée précédemment, est transactionnelle, elle implique la prise en compte de la relation entre l'évènement et l'individu selon Lazarus (1968). Il faut noter que l'évaluation n'est pas nécessairement un processus haut niveau conscient, mais peut aussi être un processus non contrôlable et limité. Arnold et Roseman la décrivent comme une sorte de perception involontaire dépendante de nos motivations (Arnold, 1960 ; Roseman & Smith, 2001).

Cependant, de nombreuses problématiques émergent par rapport aux aspects qui sont pris en compte dans le cadre de cette évaluation. En effet, une émotion n'est pas un processus anodin pour l'organisme, elle implique la mobilisation de ressources afin de répondre à un évènement aux conséquences importantes. Quels sont donc les critères pris en compte par cette évaluation cognitive ?

Scherer a établi une liste de *critères d'évaluation de stimulus* (CES) afin de lister et catégoriser les modalités à l'origine du déclenchement d'émotions. D. M. Grandjean et Scherer (2009) distinguent ces CES au sein de quatre objectifs évaluatifs :

- La pertinence de l'évènement (Détection de la pertinence).

- Les implications et les conséquences de cet évènement sur le long terme (Évaluation de l'implication).
- La capacité de s'adapter aux conséquences de cet évènement (Potentiel de maîtrise).
- La signification de cet évènement en rapport avec l'individu et les normes sociales (Évaluation de la signification normative).

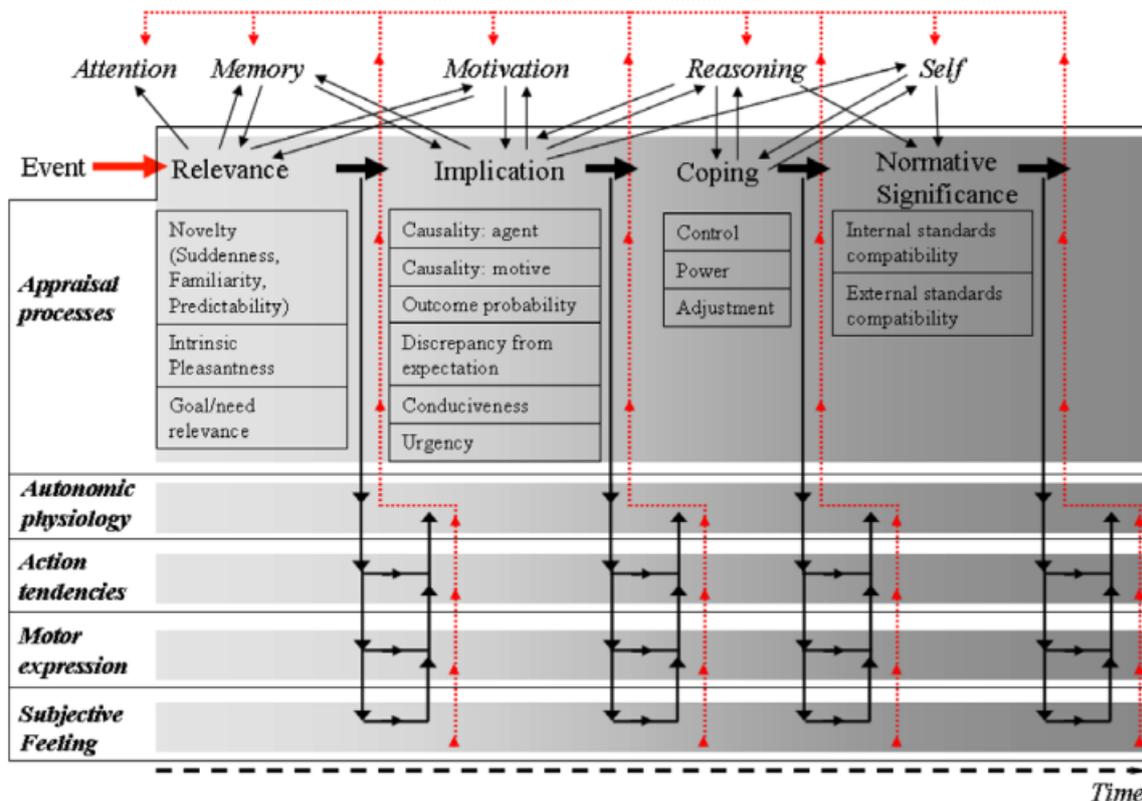


FIGURE II.5 – Modèle des composants de Scherer et de leurs interactions avec d'autres systèmes fonctionnels (D. M. Grandjean & Scherer, 2009 ; Scherer, 2005)

La figure II.5 représente ces quatre objectifs et leur intégration temporelle dans la séquence émotionnelle. Chacun des objectifs évaluatifs est accompagné d'un certain nombre de CES et est relié à des systèmes fonctionnels comme l'attention, la mémoire, la motivation, le raisonnement et la conscience de soi. De plus, chacune de ces étapes d'interprétation de la situation implique les autres sous-systèmes d'une séquence émotionnelle. Cette figure représente le rôle central de la composante cognitive dans la théorie de l'*appraisal* et le modèle des processus composants de Scherer (2005).

Détection de la pertinence

Notre organisme est habitué à percevoir et à traiter des stimuli internes et externes de manière constante et automatique. L'objectif de cette catégorie de CES est donc de déterminer les critères des événements suffisamment importants pour déclencher un épisode émotionnel. La détection de la pertinence impliquerait en elle-même trois CES pour répondre à ce besoin (D. Grandjean *et al.*, 2008) :

- Le critère de nouveauté : il se caractérise des modalités sensorielles (visuelles, auditives, olfactives...) pour évaluer la nouveauté d'un stimulus afin d'allouer des ressources émotionnelles nécessaires. Trois aspects seront pris en compte : la soudaineté, le degré de familiarité et la prédictibilité du stimulus.
- Le critère d'agrément intrinsèque : il s'agit d'une évaluation rapide et automatique d'un stimulus. D'après D. Grandjean *et al.* (2008), ce processus est automatique et sert à évaluer le niveau d'attractivité ou d'aversion d'un stimulus. Par exemple, on sera généralement naturellement plus attiré par un gâteau qu'un animal dangereux.
- Le critère de pertinence : ce processus est plus tardif. Premièrement, il permet de hiérarchiser le niveau d'importance d'un stimulus. Deuxièmement, d'évaluer plus précisément la congruence par rapport aux objectifs de l'individu. Un événement pouvant être dangereux pour l'organisme sera jugé comme plus pertinent.

Globalement, les critères de nouveauté et d'agrément intrinsèque seraient les plus précoces dans la séquence de détection de la pertinence. Ils seraient aussi inconscients et automatiques, tandis que le critère de pertinence des buts et besoins serait plus tardif, conscient et contrôlé (D. Grandjean *et al.*, 2008).

Évaluation de l'implication

Suite à une évaluation préalable et primaire avec l'évaluation de la pertinence d'un événement, l'évaluation cognitive s'engage vers une estimation plus approfondie de cette situation afin d'en extraire les causes, et les impacts plus précis en rapport avec les motivations individuelles. Ainsi, cinq CES ont été référencés :

- Le critère de l'attribution causale : l'objectif étant d'identifier l'agent à l'origine du phénomène observé. L'origine peut être attribuée à l'individu lui-même, à un autre individu ou à aucune cause spécifique. L'émotion ressentie sera différente si l'individu se sent directement responsable d'un événement ou s'il l'attribue à une source externe.
- Le critère de probabilité des conséquences : il s'agit de l'estimation des conséquences futures d'un événement. Selon Ira Roseman, des événements aux caractéristiques différentes peuvent déclencher une expérience émotionnelle similaire

si les conséquences sont perçues comme similaires par l'organisme (le décès d'un proche ou une séparation pouvant tous les deux être attribués à une perte par exemple) (Roseman & Smith, 2001).

- Le critère de différence avec les attentes : les conséquences peuvent être en opposition avec les prédictions. Un événement complètement inattendu pourra alors causer de la surprise ou de la déception.
- Le critère de facilitation/obstruction aux buts-besoins : un événement sera analysé en fonction des buts de l'individu. Un événement comme rater un examen sera perçu comme un obstacle majeur pour atteindre l'objectif et sera donc considéré comme perturbant.
- Le critère d'urgence : ce critère apporte une dimension temporelle à l'analyse de la situation et va directement impacter la tendance à l'action afin de répondre au mieux à la situation. L'individu peut alors se sentir sous contrôle ou débordée par la temporalité.

La diversité de ces critères permet de rendre compte des différences inter-individuelles dans les attributions émotionnelles pour un même événement, de la même manière qu'elle peut expliquer la survenue de patterns similaires chez une même personne pour des événements différents. Les causes, les attentes et les motivations sont des paramètres essentiels qui vont impacter le type d'émotion déclenché.

Potentiel de maîtrise

Comme l'a suggéré Paul Lazarus (1968), l'évaluation cognitive ne comprend pas seulement l'analyse de la situation, mais surtout la prise en compte de la relation entre l'individu et son environnement. Cela intègre donc l'évaluation des ressources internes perçues par l'individu pour répondre à la situation. Selon Didier Grandjean, cette évaluation secondaire, qui suit l'évaluation de la situation en fonction des objectifs et des besoins, se découpe en trois CES :

- Le critère de contrôle : ce critère porte sur l'évaluation de la possibilité d'influencer une situation. Par exemple, un étudiant peut tenter de préparer au mieux un examen en révisant. En revanche, face à un événement météorologique comme une tempête, il n'y a pas de possibilités d'agir pour changer cette situation.
- Le critère de puissance des conséquences : il s'agit de l'estimation des ressources à disposition pour réussir le scénario d'action envisagé. Si l'étudiant qui doit préparer un examen a peu de temps libre à disposition et peu de moyens d'accéder aux connaissances, alors sa puissance pour agir sur cet événement est limitée.
- Le critère d'ajustement : il s'agit d'une estimation sur le plus long terme des impacts de la situation sur l'individu, une fois une stratégie mise en place et les

ressources épuisées. Un étudiant peut, par exemple, estimer finalement que le diplôme qu'il cherche à obtenir n'est pas si important pour sa poursuite d'étude.

L'ensemble de ces critères a pour objectif de choisir la meilleure option pour résoudre la situation avec un résultat correspondant aux motivations (D. M. Grandjean & Scherer, 2009). Cela implique donc d'anticiper les scénarios possibles et leurs conséquences dans la durée.

Évaluation de la signification normative

Cette dernière catégorie de critères d'évaluation permet d'intégrer un aspect de construction sociale dans l'évaluation cognitive. Cet aspect est considéré comme postérieur dans la séquence temporelle par rapport aux autres évaluations. Deux CES sont impliqués dans cette évaluation : le critère des standards internes et le critère des standards externes.

Ces standards comprennent donc les valeurs morales et idéaux personnels, ainsi que les valeurs morales et idéaux des groupes sociaux dans lesquels l'individu évolue. Un étudiant peu scrupuleux pourra donc considérer que tricher pour obtenir l'examen n'est pas un problème par rapport à ses propres convictions (la réussite étant évaluée comme plus importante que l'intégrité). En revanche, l'étudiant sait bien que si ses collègues ou les professeurs apprennent la fraude, cela provoquera des sentiments aversifs à son égard et remettra en question sa réputation et sa réussite.

Comme l'a soulevé Lazarus (1968), l'évaluation de la situation n'est pas linéaire ni unique, car elle évolue avec le temps. Plusieurs « cycles de réévaluations » peuvent avoir lieu à la suite d'un événement déclencheur. L'évaluation cognitive est un processus dynamique et l'ensemble des CES proposés dans le modèle des processus composants est un moyen pour conceptualiser la complexité du phénomène et des informations prises en compte. Ainsi, les traitements cognitifs impliqués peuvent à la fois être de bas niveau, automatiques et inconscients, dans le cas des critères d'évaluation les plus basiques et rapides des stimuli, jusqu'à engager des mécanismes de haut niveau, pour anticiper les conséquences en fonction de la motivation et des normes sociales. L'aspect cognitif des émotions constitue donc, en lui-même, un objet d'étude à part entière dans la littérature scientifique sur les émotions. Les séquences temporelles des évaluations, leur classification et leur définition peuvent faire l'objet de débats et d'interprétations variables.

II.3.4 Réponses psychophysologiques

L'aspect physiologique constitue un des aspects les plus visibles et mesurables des épisodes émotionnels. La théorie de James-Lange, précurseur dans ce domaine, mettait déjà l'accent sur le rôle des réactions physiologiques suite à un stimulus comme le déclencheur de l'expérience émotionnelle. Dans cette sous-partie, nous présenterons les différentes catégories de réponses psychophysologiques associées aux épisodes émotionnels.

Selon Anna Pecchinenda (2001), il existerait de nombreuses observations dans la littérature scientifique permettant de mettre en évidence un lien entre des expériences émotionnelles et des modifications physiologiques. Bien qu'il soit envisageable d'associer un ensemble de modifications physiologiques à une émotion prototypique, ces modifications resteraient conditionnées à l'appréciation cognitive de la situation par l'individu, selon la théorie de l'*appraisal*. De même, les réponses psychophysologiques peuvent être associées à la composante de la tendance à l'action. En effet, les motivations de l'individu et la stratégie mise en place pour répondre à la situation peuvent impacter les réactions de l'organisme. Ainsi, la mesure des réponses physiologiques pourrait constituer un marqueur de l'évaluation cognitive lors d'un épisode émotionnel.

Nous pouvons associer les réponses psychophysologiques au système nerveux qui peut être divisé en deux parties : le système nerveux central (SNC), composé du cerveau et de la moelle épinière, et le système nerveux périphérique (SNP), composé de toutes les autres structures nerveuses. Le SNP remplit deux principaux objectifs : capter et communiquer les informations sensibles vers le SNC et transmettre les informations du SNC vers le corps. Le SNP est lui-même divisé en un système nerveux somatique (SNS) et un système nerveux autonome ou viscéral (SNA). Le SNS intègre principalement des fonctions soumises au contrôle conscient et volontaire tandis que le SNA comprend les fonctions automatiques de régulation de l'organisme. (Breedlove *et al.*, 2012).

Le SNA prend en charge, de manière non exhaustive : la digestion, la respiration, le rythme cardiaque, la transpiration ou la dilatation pupillaire... Tatjana Aue (2009) liste trois catégories de modifications physiologiques opérées au sein du SNA lors d'un épisode émotionnel :

- L'activité cardiovasculaire, qui se concrétise par des modifications du rythme cardiaque en fonction de l'intensité de l'émotion et afin de s'adapter au contexte. On peut mesurer la fréquence cardiaque à partir d'un électrocardiogramme qui, à l'aide d'électrodes posées sur la surface de la peau, enregistre les impulsions électriques des contractions du cœur. Le volume de pulsations sanguines peut aussi être mesuré par l'intermédiaire d'un photopléthysmographe afin de calculer la fréquence cardiaque, cette méthode étant moins invasive (Haag *et al.*, 2004). De nombreuses expériences ont montré la relation entre la réalisation de tâches perçues comme difficiles et la fréquence cardiaque (Pecchinenda, 2001).

- L'activité électrodermale est déterminée par l'activité des glandes sudoripares qui provoquent de la transpiration. Elle est mesurée par la conductance sur la surface de la peau. La relation entre l'activité électrodermale et l'évaluation cognitive a été testée dans une célèbre expérience de Lazarus. À partir d'un extrait vidéo montrant une opération chirurgicale de circoncision en Nouvelle-Guinée et en proposant soit une version sans commentaires, soit une version avec une bande-son explicative, allégeant la gravité de la situation. Les expérimentateurs ont enregistré des activations de la conductance électrodermale plus faibles pour ceux qui avaient reçu des explications au cours de la vidéo et qui ont donc évalué différemment ce média par rapport à ceux qui ont visionné seulement la vidéo (Lazarus & Alfert, 1964 ; Pecchinenda, 2001).
- L'activité motrice peut être mesurée par électromyographie, procédé consistant à capter les champs électriques produits lors de l'excitation d'un muscle. La tension des muscles pouvant significativement changer dans le cas d'une émotion intense. Les muscles du visage sont particulièrement sollicités lors des états émotionnels et donc utilisés dans les études de psychologie (Aue, 2009).

Notons qu'il existe dans la littérature de nombreuses autres mesures réalisables et associables avec les émotions comme la dilatation pupillaire ou la température de la peau. De plus, avec l'avancée des neurosciences, des outils de mesure peuvent porter directement sur le SNC pour capter l'activité cérébrale. En effet, les outils de neuro-imagerie fonctionnelle, comme les IRM (Imagerie par résonance magnétique), ont permis de réaliser des études pour identifier les structures corticales associées à certaines émotions (Murphy *et al.*, 2003). Des émotions considérées comme basiques, telles que la joie ou la peur ont fait l'objet d'études particulières (Lotstra, 2002). Joseph LeDoux (2000) aurait identifié deux circuits neuronaux liés à la peur : un circuit court passant du thalamus à l'amygdale et un circuit long transitant par le cortex avant d'atteindre l'amygdale. Le premier circuit permettrait un traitement automatique de l'information pour assurer la survie de l'organisme avec des réactions rapides, le circuit long permettrait de réévaluer la situation avec plus de recul. De la même manière, l'accumbens, un sous-ensemble du cerveau, a été identifié comme un ensemble de substrats neurologiques liés au système de la récompense. Ainsi, des situations sources de plaisir activeraient les neurones du tronc cérébral qui déclencheraient une libération de dopamine vers l'accumbens. En plus de ces analyses perfectionnées, des outils moins invasifs peuvent être employés comme les électro-encéphalogrammes (Conneau, 2016), nécessitant seulement des capteurs à poser sur le cuir chevelu.

Les réponses psychophysiologiques fournissent donc un vaste choix de moyens pour réaliser des mesures expérimentales. Elles permettent de mesurer l'activation de l'organisme au cours d'un épisode émotionnel. Cependant, elles peuvent impliquer des outils de mesure perfectionnés et contraignants pour le sujet. De plus, relier les mesures physiologiques à l'expérience émotionnelle, dans des conditions écologiques, constitue une

tâche complexe, au vu des nombreux signaux pouvant perturber les signaux enregistrés. Il est donc nécessaire de combiner les mesures physiologiques avec d'autres modalités pour vérifier la pertinence des mesures.

II.3.5 Tendance à l'action

La tendance à l'action est englobée dans la composante motivationnelle. Pourtant, il est important de distinguer deux types de motivations : la motivation antécédente à l'expérience émotionnelle et la motivation conséquente de l'expérience émotionnelle, désignée comme la tendance à l'action. La première est déterminante pour identifier les stimuli pertinents en fonction des buts et besoins. C'est pourquoi elle est incluse dans la composante de l'évaluation cognitive. Selon Aue Seil (2009), la motivation conséquente, ou tendance à l'action, prend en compte la motivation antécédente à l'action pour diriger l'organisme vers une stratégie appropriée.

Bernard Weiner (2001) propose une dimension d'apprentissage inhérente aux émotions afin d'améliorer la direction des actions à l'avenir. Ainsi, il considère que chaque individu va évaluer, à la suite d'un épisode émotionnel, les causes de sa réussite ou de son échec pour répondre à la situation. L'attribution des causes s'intégrera dans l'expérience personnelle d'un individu et influencera ses stratégies futures. Cet apprentissage se concrétise par trois variables pour évaluer les causes des conséquences d'un événement, déclencheur d'épisode émotionnel :

- La dimension interne ou externe. À savoir est-ce que les raisons de l'échec ou de la réussite de la situation sont le fait de l'individu ou de circonstances externes (autres personnes, environnement...)
- La stabilité ou l'instabilité. Est-ce que les moyens de répondre à la situation sont constants dans la durée? Par exemple, liés à une aptitude de l'individu ou alors complètement sujets au hasard.
- La contrôlabilité de la cause. Est-ce que les ressources investies par l'individu auront une influence sur l'issue de la situation?

La causalité externe ou interne de la réussite ou de l'échec impactera les émotions ressenties par une personne. Une réussite attribuée à des causes internes, donc à des capacités ou à un effort personnel, sera plus valorisante pour l'individu, qui pourra à l'avenir évaluer un événement similaire comme moins menaçant. La stabilité démontre des attentes d'un individu face à un événement, un étudiant doué en math s'attendra par exemple à réussir ses examens, en revanche s'il est évalué dans un domaine inconnu, il ne peut pas anticiper autant la possibilité de réussir ou non. La contrôlabilité est directement liée au potentiel d'action. Un individu pourra ressentir de la frustration suite à un échec où il ne pouvait rien faire ou de la culpabilité s'il avait pu agir, mais ne l'a pas fait. Les motivations conséquentes aux épisodes et qui conduisent à la tendance

à l'action sont donc influencées par les expériences vécues et leurs attributions causales.

Cette notion de tendance à l'action a été particulièrement abordée par Nico Frijda *et al.* (1986). En effet, il résume les tendances à l'action comme une disposition, à modifier d'une part sa relation avec l'environnement dans le cas d'une situation non désirable ou à maintenir d'autre part sa relation avec l'environnement en cas de situation désirable. Cette composante est peut-être fortement associée aux réponses psychophysiologiques et aux expressions motrices pour préparer les actions conséquentes de l'émotion. Plus spécifiquement, cette composante aurait trois rôles :

- Faciliter certains comportements moteurs. Il peut alors s'agir d'une tendance à l'action différente pour une même émotion, mais elle doit avoir la même finalité (par exemple une personne agressée peut vouloir se battre ou s'enfuir, la finalité étant la survie). Cela peut comprendre la régulation de l'attention ou l'inhibition de certaines expressions motrices.
- Déclencher des réactions physiologiques pour accompagner les comportements moteurs.
- Communiquer aux autres son émotion à partir d'expressions faciales, de gestes ou de postures.

De même, les tendances à l'action peuvent alors être associées à des termes émotionnels, à savoir des sentiments subjectifs. Selon Frijda (2003), l'action de s'éloigner peut être associée à la peur, celle du rapprochement à l'affection, l'hyperactivité à de la joie et l'hypoactivité à de la tristesse... Ces comparaisons entre émotions et actions se rapprochent de travaux de Plutchik (1984) sur les liens entre fonctions évolutives, actions et sentiments subjectifs.

En conclusion, les motivations représentées par les tendances à l'action jouent un rôle prépondérant, non seulement pour comprendre les stratégies mises en place dans le cadre des processus émotionnels, mais aussi pour comprendre les évolutions inter et intra-individuelles qui se forment avec l'expérience par les attributions causales. Tatjana Aue Seil (2009) considère que cette composante est l'une des moins étudiées dans la littérature, alors que la motivation constitue l'un des mécanismes cognitifs essentiels pour catégoriser et comprendre les émotions.

II.3.6 Expression motrice

Dans la littérature scientifique, les expressions faciales sont largement majoritaires lorsqu'il s'agit d'étudier les comportements causés par les émotions. Les raisons de cet engouement pour les expressions faciales sont à la fois historiques et pragmatiques. L'histoire de l'étude des émotions est, en effet, particulièrement marquée par les apports de scientifiques comme Charles Darwin et de l'approche évolutionniste qui a donné une

place importante aux expressions faciales. Elles revêtent à la fois un rôle fonctionnel et communicatif, directement lié à la tendance à l'action ou au sentiment subjectif dans une forme non verbale. Elles sont à la base de nombreuses catégorisations d'émotions (Darwin, 1877 ; Ekman & Friesen, 1971). Enfin, elles sont un des éléments les plus facilement observables d'un œil extérieur.

De manière générale, les expressions faciales ont un rôle prépondérant dans les interactions sociales selon Suzanne Kaiser *et al.* (2009). En effet, au-delà de l'expression des émotions, elles indiquent aussi le niveau d'attention d'une personne, son engagement cognitif dans une tâche et permettent d'engager des relations sociales (le sourire de politesse par exemple). De nombreux muscles sont engagés dans les expressions du visage : le muscle frontal qui plisse le front et lève les sourcils, les zygomatiques qui soulèvent les lèvres pour n'en citer que quelques-uns. Une part significative du cortex moteur serait dédiée au contrôle de ces muscles du visage (Breedlove *et al.*, 2012). Certaines espèces de grands singes seraient aussi en mesure de produire des expressions du visage pour communiquer. Cependant, les expressions des grands singes n'ont pas nécessairement la même signification que chez les humains. Néanmoins, cela démontre un rôle écologique des émotions et un gain évolutif pour les espèces concernées.

Pour les défenseurs des théories de l'évaluation cognitive, l'existence d'expressions faciales prototypiques universelles n'est pas à remettre en question. En revanche, ils ne considèrent ces expressions que comme des patterns déclenchés à la suite d'évaluations cognitives, au sein de nombreux autres processus. Elles ne permettent donc pas, à elles seules, de catégoriser les émotions, mais constituent un moyen d'inférer une émotion (Kaiser *et al.*, 2009). Les expressions faciales auraient un lien fort avec la composante des tendances à l'action, car elles donneraient des informations essentielles sur les actions envisagées par un individu pour modifier sa relation avec l'environnement (Frijda & Tcherkassof, 1997).

Les caractéristiques des expressions vocales ont aussi été étudiées comme l'un des effets des émotions sur le comportement. En effet, la tension des cordes vocales est influencée par l'intensité des épisodes émotionnels selon Didier D. M. Grandjean et Baenziger (2009). La prosodie est alors affectée et nous pouvons percevoir lorsqu'une personne s'exprimant à l'oral est en train de ressentir un stress ou est émue par exemple. Ces dernières années, la recherche sur le traitement automatique de la voix pour en extraire les composantes émotionnelles s'est particulièrement développée. L'usage de plus en plus important de technologies de communication accentuant d'autant plus ces recherches.

Cependant, les expressions motrices intègrent des manifestations plus étendues que les seules expressions faciales et vocales. Nous pouvons noter que les gestes, les postures et les mouvements globaux du corps sont aussi impactés par les épisodes émotionnels. Ces aspects comportementaux sont moins étudiés dans la littérature, pourtant dans

le cadre de ce projet de recherche, ces modalités nous intéressent particulièrement, car elles constituent des données que nous pouvons enregistrer par l'intermédiaire des capteurs de smartphone. Selon Nele Dael *et al.* (2012), les travaux de recherche portant sur les gestes sont relativement récents dans l'histoire de la recherche sur les émotions, mais commencent à émerger. À l'image de la relation entre les expressions faciales et les tendances à l'action, présentés précédemment, les mouvements du corps exprimeraient aussi des tendances à l'action. Plus spécifiquement, Anna Tcherkassof et Frijda (2014) proposent que les émotions s'incarnent dans le corps comme des attitudes corporelles envers un évènement ou un objet. Une des études les plus anciennes, menée par Marco De Meijer (1989) a montré que les mouvements effectués par des acteurs pouvaient être attribués par un public à certaines préparations à l'action comme l'acceptation, le rejet ou l'approche... Autre exemple, Frank Pollick *et al.* (2001) ont démontré un lien entre les accélérations des mouvements du bras et le niveau d'activation provoqué par des épisodes émotionnels. Bien que peu exploités, les mouvements du corps sont donc représentatifs de la composante comportementale. L'utilisation de ce type d'informations est d'autant plus pertinente lorsqu'elle est combinée, dans des études multimodales, avec les expressions faciales ou vocales.

En conclusion, les expressions motrices s'avèrent être une composante à privilégier pour inférer un épisode émotionnel. La diversité des modalités constitue à la fois une richesse, car elle permet d'étudier les comportements causés sur de nombreux aspects, mais elle constitue aussi une difficulté à cause de la complexité de fusionner toutes ces informations. L'état actuel de la recherche montre que les expressions faciales ont été une porte d'entrée pour étudier et classifier les émotions, car plus faciles à analyser. Pourtant, elles ne sont qu'un aspect de cette composante et peuvent être complétées par d'autres informations permettant de faire le lien avec les tendances à l'action et donc de tenter d'inférer des émotions.

II.3.7 Sentiment subjectif

Antonio Damasio (2003, 2006) considère que les sentiments sont la partie privée et les émotions sont la partie publique d'un même processus. Ainsi, les émotions précèdent les sentiments, qui sont une perception des modifications causées par les émotions. Il définit les sentiments de la façon suivante :

« Un sentiment est la perception d'un certain état du corps ainsi que celle d'un certain mode de pensée et des pensées ayant certains thèmes. » (Damasio, 2003)

Ainsi, le sentiment proviendrait de la perception d'un état interne corporel spécifique (par exemple l'accélération de la fréquence cardiaque et la contraction des intestins) associée à des pensées portées sur un objet interne ou externe responsable. La représentation de Damasio est donc similaire au modèle des processus composants de Scherer qui l'intègre comme une composante à part entière de l'émotion. Pour Scherer,

les termes « émotion » et « sentiment » (*feeling* en anglais) ont régulièrement été utilisés pour désigner les mêmes concepts. Ainsi, il considère que la séquence temporelle proposée par James, basée sur la perception des modifications corporelles, devrait en réalité être nommée sentiment et non pas émotion (James, 1884 ; Scherer, 2005).

Dans le cadre du modèle des processus composants, Elise Dan-Glauser (2008) définit le sentiment subjectif de la manière suivante :

« Nous considérons le sentiment subjectif comme l'émotion telle qu'elle est vécue par l'individu ; il s'agit de l'aspect conscient du processus émotionnel. »
(Dan-Glauser, 2008)

La composante du sentiment subjectif a donc un rôle particulier dans le modèle, il a pour rôle de contrôler les modifications des autres composantes. Les émotions étant un processus dynamique et non linéaire, l'individu doit puiser dans des représentations corporelles et cognitives pour reconstituer un ressenti d'ensemble. Le sentiment subjectif est donc « le reflet conscient des changements dans toutes les autres composantes » (Dan-Glauser, 2008). Le sentiment subjectif n'est accessible, pour une personne extérieure, qu'à partir de la verbalisation d'une personne faisant l'expérience d'une émotion. Elle peut alors mettre un mot synthétisant l'ensemble des processus provoqués par l'épisode émotionnel. Cependant, la labélisation d'un processus émotionnel est une tâche complexe, elle ne rend pas nécessairement compte de l'ensemble des modifications corporelles, psychiques et cognitives. En effet, il n'est pas toujours aisé de décrire ce que nous vivons, chaque épisode étant unique, nous ne pouvons que tenter de faire des rapprochements entre des représentations similaires, sur la base de notre vécu.

Comme cela était suggéré par les théoriciens de l'approche socio-constructiviste (voir sous-section II.2.6), notre langage, et par extension notre culture, conditionne la verbalisation des expériences émotionnelles (Russell, 1991 ; Wierzbicka, 1986). Des processus différents pourraient donc être rapportés sur un même label, car il n'existe pas suffisamment de labels pour les différencier. Fondamentalement, chaque situation étant unique, il pourrait exister une infinité de labels pour décrire les expériences que nous vivons. Le langage verbal est donc à la fois un moyen d'expression de nos expériences et une limitation de nos représentations.

Pour mettre en évidence les limites de l'expression verbale pour représenter le sentiment subjectif, Scherer modélise cette composante en trois modules, visible sur la figure II.6 :

- Le module A représente le suivi des changements dans l'ensemble des quatre autres composantes. Il intègre les informations pour créer une représentation cohérente et centralisée des processus en cours. Ce module se différencie des autres par sa nature inconsciente.
- Le module B constitue la perception consciente. Ce module intègre une partie du

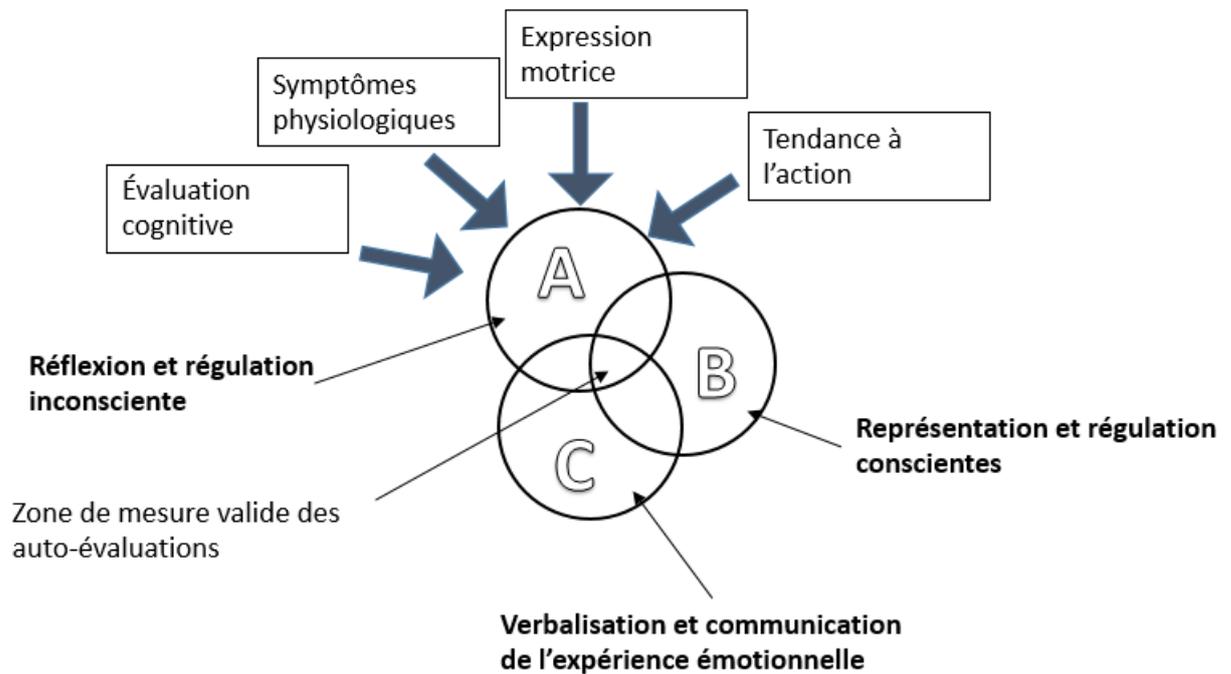


FIGURE II.6 – Représentation des trois modules du sentiment subjectif (Dan-Glauser, 2008 ; Scherer, 2004)

traitement inconscient provenant du module A, sans pour autant en saisir l'entièreté. De plus, il est complété par des représentations externes, comme des normes sociales ou culturelles.

- Tandis que les modules A et B constituent la partie « privée » des émotions, le module C en est la partie « publique ». En effet, il correspond à la verbalisation de l'expérience émotionnelle. Elle reflète les informations inconscientes du module A traduites de manière consciente par le module B, cela serait alors la sous-partie la plus valide de l'auto-évaluation. De plus, elle intègre aussi des représentations externes du module B ne provenant pas du traitement inconscient du module A. Enfin, elle ajoute un surplus d'information, causé par le choix de labels englobant des représentations plus larges que ce qui est vécu, les labels disponibles dans la langue parlée ne pouvant refléter l'expérience avec exactitude.

Ainsi, l'expression du sentiment subjectif est altérée de trois manières différentes : la perte des informations inconscientes sur les modifications des autres composantes (module A), l'ajout de représentations conscientes externes différentes (module B) et le manque de précision des termes utilisés dans la langue (module C). En conclusion,

le sentiment subjectif est une composante complexe. En effet, il s'agit à la fois de l'une des composantes les plus accessibles et le plus à même de nous aider à comprendre l'expérience émotionnelle vécue par un tiers, tout en ne représentant qu'une faible partie de l'épisode émotionnel. Cette problématique d'altération de l'information montre l'importance de concevoir des outils d'auto-évaluation à même de limiter l'altération de l'information et d'aider les participants à exprimer le plus facilement possible leur expérience.

II.3.8 Discussion

La théorie de l'évaluation cognitive peut être considérée comme une approche moderne pour expliquer l'ensemble des processus émotionnels. Elle offre une flexibilité permettant d'inclure des réflexions sur les différences intra-individuelles, inter-individuelles ou même culturelles. Ainsi, chacune de ces composantes peut former un objet d'étude à part entière. Cependant, une des limites du modèle repose sur la difficulté de délimiter exactement la nature et les relations entre certaines composantes. En effet, la description des différentes composantes proposées met en évidence quelques zones floues. Nous pouvons par exemple noter de nombreuses similitudes entre l'aspect motivationnel des tendances à l'action et l'évaluation cognitive. Ainsi, la théorie de l'évaluation cognitive et l'ensemble de ses composantes sont source de discussions et de débats entre les théoriciens des émotions. Les émotions étant à la fois un concept fondamentalement abstrait et un ensemble de processus dynamiques, il est complexe d'en saisir la globalité. En revanche, nous pouvons étudier un certain nombre de caractéristiques d'un épisode émotionnel à partir des manifestations mesurables, catégorisées dans le modèle des processus composants. Afin de mesurer les émotions, il est cependant nécessaire de pouvoir les catégoriser et les représenter, afin de les différencier. Nous présenterons donc dans la prochaine partie les modèles de représentation couramment utilisés dans la littérature.

II.4 Classification et représentation des émotions

II.4.1 Taxonomie des états affectifs

Comme cela a été évoqué dans la sous-section [II.2.1](#), le registre des émotions est accompagné d'un vocabulaire varié pour désigner des états affectifs. Avant de décrire les différentes représentations des émotions, il nous semble nécessaire de faire la distinction avec d'autres phénomènes affectifs présents dans la littérature scientifique ou dans le langage courant. Klaus Scherer propose de discerner les phénomènes affectifs à partir de la prise en compte de plusieurs variables (voir le tableau [II.2](#)) (Scherer *et al.*, [2000](#); Scherer, [2005](#)) :

- La présence d'un objet déclencheur, le phénomène affectif pouvant en effet être causé soit par un stimulus interne ou externe, soit ne pas être relié à un événement particulier. Cela suppose aussi le besoin d'une évaluation cognitive pour apprécier le stimulus.
- L'intensité des réponses à l'expérience émotionnelle. Par exemple, pour les réponses physiologiques, le niveau d'augmentation de la fréquence cardiaque. Un état affectif intense impliquant une plus forte utilisation des ressources de l'organisme.
- La durée de l'épisode affectif.
- Le niveau de synchronisation des différentes composantes lors de l'épisode affectif.
- La rapidité de changement de l'état, cette variable porte sur l'évolution de l'évaluation cognitive dans le temps. Certains phénomènes peuvent être rapidement réévalués, car ils sont associés à un événement très changeant.
- L'impact sur le comportement. Un fort impact peut, par exemple, stopper le comportement en cours et générer une nouvelle stratégie.

TABLE II.2 – Classification des phénomènes affectifs (Scherer *et al.*, 2000 ; Scherer, 2005)

Variables	Objet déclencheur	Durée	Intensité	Synchronisation	Impact comportemental
Préférences	Sur des objets ou des événements	Intermédiaire	Faible	Très faible	Intermédiaire
Attitude	Sur des objets ou des personnes	Élevée	Intermédiaire	Très faible	Faible
Humeur	Non spécifique	Élevée	Intermédiaire	Faible	Faible
Disposition affective	Non spécifique	Très Élevé	Faible	Très faible	Faible
Émotion	Sur des objets, des événements ou des personnes	Faible	Élevée	Élevée	Élevé

L'émotion constitue donc un des phénomènes affectifs les plus extrêmes en comparaison avec les autres états affectifs. En effet, elle est caractérisée par la prise en compte d'un stimulus spécifique, une forte intensité et une durée faible (de l'ordre de quelques minutes). De plus, elle provoque une réponse synchronisée des différentes composantes. C'est un phénomène prioritaire pour l'organisme, car il implique un changement

de comportement rapide pour s'adapter à la situation. À l'inverse, les humeurs sont des états diffus, faibles en intensité, de longue durée (de l'ordre de plusieurs heures) et ne sont pas nécessairement portées sur un stimulus particulier. Elles n'impliquent pas une modification comportementale immédiate. Scherer compare aussi d'autres phénomènes affectifs comme les préférences, les attitudes et les dispositions affectives.

Voici une explication plus détaillée de ces phénomènes affectifs (Scherer, 2005) :

- Les préférences sont des phénomènes affectifs stables sur la durée et caractérisés par l'attraction ou l'aversion envers un stimulus ou un objet. L'intensité et la synchronisation de l'expérience sont faibles. En revanche, elles peuvent amener un comportement et des tendances à l'action de rapprochement ou d'évitement.
- Les attitudes sont généralement déclenchées à la vue ou à la pensée d'un objet ou d'une personne spécifiques. Elles comporteraient une composante cognitive, affective et motivationnelle, mais une intensité relativement faible et une durée prolongée. Ainsi, selon Scherer le fait de ressentir de l'amitié ou de l'amour pour une personne pourrait rentrer dans cette catégorie de phénomène affectif, bien que cela implique sans doute des épisodes affectifs plus intenses à des instants précis.
- Les dispositions affectives sont des traits de personnalité stables, elles ont donc une consistance sur le long terme. Ces dispositions influencent les humeurs qu'un individu peut expérimenter plus fréquemment et les tendances à réagir de certaines manières lors d'épisodes émotionnels. On peut proposer les termes suivants pour décrire des dispositions affectives : nerveux, jaloux, irritable... Les dispositions affectives peuvent inclure des troubles cliniques comme une dépression sur le long terme.

En conclusion, en parallèle des émotions, il existe une multitude de phénomènes psychologiques affectifs impactant notre quotidien. Ils sont nettement différenciables des émotions bien que certains puissent influencer le déclenchement des émotions et les réactions associées. Chacun de ces phénomènes constitue un objet de recherche à part et implique donc d'utiliser des outils et des méthodologies adaptés pour les étudier.

II.4.2 Classification des émotions

En parallèle des approches théoriques, nous pouvons référencer un certain nombre de modèles de représentation et de classification des états émotionnels. Si la théorie de l'*appraisal* met en évidence la complexité des phénomènes émotionnels, il est possible de simplifier notre représentation des différents états observables afin de construire une méthodologie. Dans ce sens, des modèles de classification permettent d'appréhender les différents types d'émotions de manière concrète et opérationnelle à partir d'un certain nombre de critères. Ces critères peuvent correspondre à des composantes du

modèle des processus composants. Des chercheurs ont alors construit des classifications en fonction des expressions motrices, des critères de l'évaluation cognitive ou par une approche linguistique fondée sur les sentiments subjectifs.

Dans le cadre de cet état de l'art, nous différencierons les modèles de classification en deux types :

- Les modèles discrets ou catégoriels, qui se définissent par un nombre fini d'émotions dans leur représentation. Les émotions peuvent être différenciées par une seule caractéristique ou par un ensemble de critères complexes. Il faut noter que l'adoption d'un modèle de représentation catégoriel n'implique pas nécessairement le postulat d'un nombre fini d'émotions basiques dans la réalité, ce qui peut être le cas dans le cadre des théories évolutionnistes. En effet, pour les scientifiques défendant la théorie de l'évaluation cognitive, qui considèrent le nombre d'expériences émotionnelles potentiellement infini, proposer des modèles catégoriels permet d'illustrer les variables de leurs modèles à partir d'expressions verbales.
- Deuxièmement, les modèles dimensionnels ou continus se définissent par un certain nombre d'axes ou de dimensions permettant de représenter les émotions dans un espace continu. Les émotions ne rentrent donc pas dans des catégories fermées mais sont un ensemble de valeurs pour chacune de ces dimensions.

Les modèles de classification d'émotions sont des outils puissants pour différencier des groupes d'expériences émotionnelles. Ils offrent l'avantage de faciliter la compréhension entre chercheurs, de labéliser du matériel d'induction d'émotions à partir de catégories ou dimensions définies et de faciliter l'annotation et la comparaison des émotions ressenties par un sujet. Le choix d'un modèle de représentation est dépendant de la problématique et du type de données exploitées.

II.4.3 Modèles discrets et catégoriels

Dans le langage courant, nous catégorisons naturellement les émotions dans un ensemble fini, par besoin de compréhension, afin de verbaliser nos expériences pour autrui et nous-même. Ainsi, selon les langues, nous avons à notre disposition un panel de labels émotionnels, offrant une certaine granularité. Comme évoqué dans la sous-section [II.3.7](#) sur le sentiment subjectif, nous associons derrière chaque terme un certain nombre de représentations et tentons de trouver le plus adapté pour décrire notre expérience. Naturellement, des chercheurs se sont fondés sur le langage courant pour concevoir des catégorisations, en sélectionnant un certain nombre d'émotions et en leur associant des caractéristiques propres. Les critères sous-jacents à ces classifications sont très diversifiés selon le domaine de recherche des scientifiques qui les

II.4. CLASSIFICATION ET REPRÉSENTATION DES ÉMOTIONS

ont construites. Dans cette sous-partie, nous décrirons quelques modèles représentant des nombres finis d'émotions. Selon la complexité des modèles, ils peuvent intégrer seulement un nombre discret d'émotions ou intégrer des catégories pour classer les émotions et proposer plus de granularité.

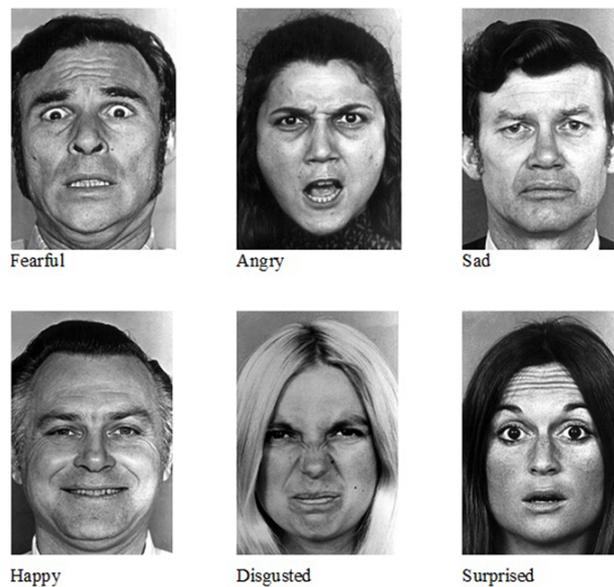


FIGURE II.7 – Photographies des expressions faciales caractéristiques des six émotions basiques d'Ekman et Friesen (1971)

Les modèles discrets les plus emblématiques proviennent de l'approche évolutionniste ou néodarwiniste que nous avons déjà évoquée dans la partie historique du chapitre (voir sous-section II.2.3). Les caractéristiques de ces approches évolutionnistes sont de considérer que les émotions sont innées, qu'elles ont un rôle fonctionnel pour la survie et qu'un certain nombre d'émotions basiques sont héréditaires chez les membres d'une espèce. Cependant, les scientifiques adoptant ces approches n'incluent pas les mêmes critères pour créer leurs modèles. Comme nous avons déjà évoqué précédemment, les travaux d'Ekman et Friesen (1971) portent sur l'étude des expressions faciales et sur l'hypothèse de leur universalité dans toutes les cultures humaines. La figure II.7 représente les expressions faciales basiques. Sa simplicité, par la proposition d'un nombre limité de six émotions, sans compter une expression faciale neutre, a favorisé la large diffusion dans la littérature scientifique et auprès du grand public. En effet, le faible nombre d'émotions et leur représentation très visuelle, à l'aide des expressions faciales caractéristiques, rendent ce modèle facilement accessible. Il est particulièrement privilégié dans la conception d'outils de détection automatique d'émotions à partir de vidéos ou de photos de visages, de plus en plus répandu grâce aux méthodes d'apprentissage automatique profond (*deep learning*). La simplicité de ce modèle constitue aussi la principale critique à son encontre, les émotions proposées

peuvent être considérées comme trop basiques et peu nombreuses pour représenter avec pertinence la complexité des expériences émotionnelles.

L'approche psycho-évolutionniste de Plutchik (1984), également déjà évoquée, propose de classifier les émotions à partir de leur utilité fonctionnelle pour la survie d'un individu. Initialement, Plutchik a proposé un modèle composé de huit émotions basiques. Il oppose, par exemple, joie et tristesse, comme étant d'un point de vue fonctionnel un gain ou une perte. La distance avec le centre du modèle indique le niveau d'intensité, les émotions centrales étant les plus intenses. La rage serait, par exemple, une émotion plus prononcée que la colère. De plus, des émotions mixtes ou secondaires sont formées en combinant les émotions voisines, ainsi la combinaison de joie et d'anticipation formerait de l'espoir.

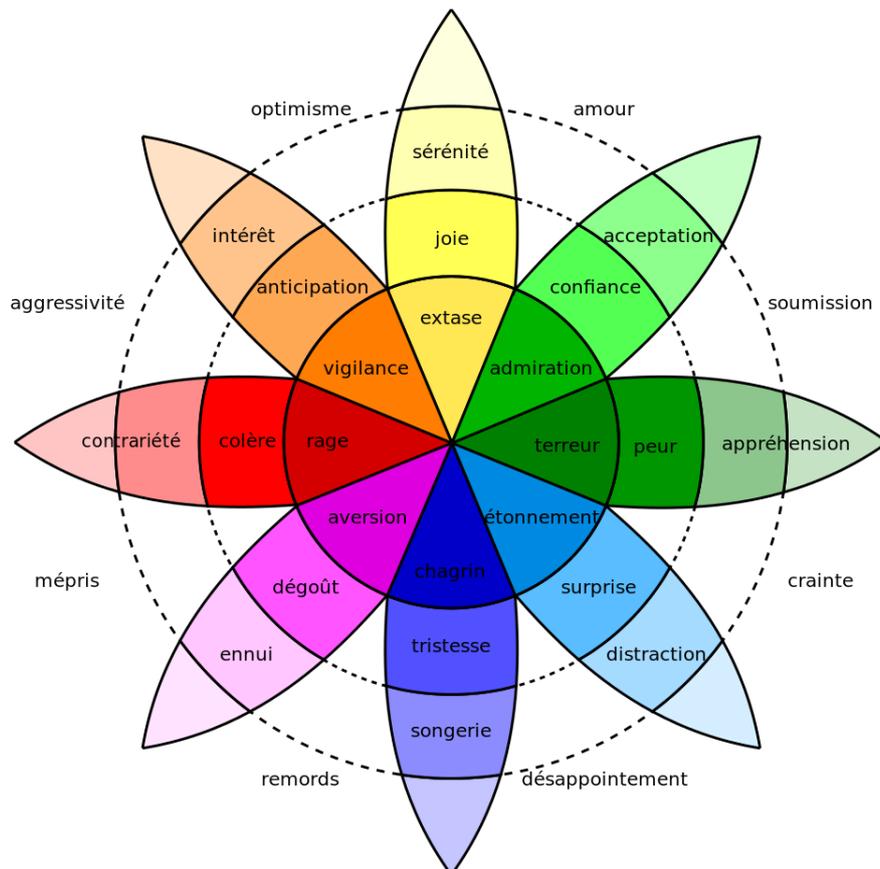


FIGURE II.8 – Modèle de représentation d'émotions de Plutchik (1984)

Force est de constater que les modèles d'émotions catégorielles ne se sont pas

nécessairement cantonnés aux approches évolutionnistes. Pour autant, les classifications discrètes sont-elles compatibles avec la théorie de l'évaluation cognitive ? Les cinq composantes de la théorie ne peuvent être entièrement représentées par un modèle d'émotions basiques. Selon Phoebe Ellsworth (2013), la plupart des chercheurs travaillant sur cette théorie admettent qu'il existe un nombre potentiellement infini d'états émotionnels, au vu de la complexité du phénomène et de la multitude des variables qui pourraient être prises en compte. Cependant, même si les émotions sont des phénomènes dynamiques, la langue les a catégorisées dans un ensemble de labels fini. À l'image des couleurs qui sont, d'un point de vue physique, un spectre continu et pour lesquelles nous avons établi un certain nombre de termes pour les représenter. Le concept d'émotions basiques pose donc problème aux théoriciens de l'*appraisal*, mais il offre tout de même une porte d'entrée pour approcher les phénomènes émotionnels et illustrer les différents critères appréciés par la cognition ou la motivation.

Des scientifiques des théories de l'*appraisal* ont donc catégorisé des critères pouvant être pris en compte dans l'évaluation cognitive pour créer des modèles descriptifs. Le modèle OCC proposé par Ortony *et al.* (1990) est l'un des plus emblématiques. En effet, ce modèle liste les variables et les modalités affectant le déclenchement d'une émotion type. Il propose de hiérarchiser 22 émotions composites dans trois branches, chacune de ces branches étant associée à une perception plaisante ou déplaisante de la situation par l'individu :

- Les conséquences des actions : Les émotions provoquées dépendent alors des agents qui vont subir les conséquences, les autres ou soi-même. Le modèle OCC introduit un aspect empathique avec des émotions portées vers les autres, mais affectant l'individu. On peut donc ressentir de la pitié pour un ami qui subit les conséquences négatives d'un événement ou au contraire jubiler face aux déboires d'un adversaire. De plus, une dimension d'anticipation est introduite, des émotions peuvent être vécues en anticipation d'un événement à venir ou en conséquence, selon que l'évènement se produise ou non. On ressentira de l'espoir face à un événement futur attractif puis de la satisfaction si celui-ci se produit tel qu'il est attendu. De plus, un événement peut être inattendu et provoquer d'autres types d'émotions.
- Les actions des agents : Les émotions dépendent de nos propres actions ou de celles des autres, et de leurs conséquences directes sur nous-mêmes. On pourra alors ressentir de la honte, si on pense ne pas avoir agi correctement, mais sans conséquence directe sur soi-même. De même, on pourra ressentir de la gratitude envers un autre agent ou envers soi-même par rapport à des actions allant dans nos buts et nos intérêts.
- La perception d'un objet : Un objet concret (une personne, un objet inanimé...) ou abstrait (une pensée, un souvenir...) peut provoquer des comportements d'aversion ou d'attraction, illustrées par exemple par l'amour ou la haine.

Pour chaque branche, les émotions peuvent être plaisantes ou déplaisantes pour l'individu. Il y a donc une composante motivationnelle dans le modèle OCC, car les événements ou les actions des autres sont perçus en anticipant les impacts sur nos objectifs personnels. De plus, le modèle OCC apporte une composante sociale et empathique en prenant en compte les effets sur les autres agents comme sources d'émotions. Ce modèle est particulièrement populaire dans les approches computationnelles des émotions pour modéliser les émotions d'agents artificiels (Steunebrink *et al.*, 2009). Le modèle OCC est synthétisé sur le tableau II.3.

TABLE II.3 – Catégorisation des émotions du modèle OCC (Ortony *et al.*, 1990), (adaptation de (Mahboub, 2011))

			Réaction personnelle	
			Plaisante	Déplaisante
Conséquences d'événements	Pour les autres	Désirable	Heureux pour	Ressentiment
		Non désirable	Jubilation	Pitié
	Pour soi-même	Anticipation d'un événement	Espoir	Peur
		Évènement anticipé se confirmant	Satisfaction	Peur confirmée
		Évènement anticipé ne se confirmant pas	Soulagement	Déception
Évènement non anticipé	Joie	Détresse		
Action des agents	Soi-même		Fierté	Honte
		Conséquence pour soi-même et faible intensité	Gratification	Remords
		Conséquence pour soi-même et forte intensité	Gratitude	Colère
	Autre agent		Admiration	Reproche
		Conséquence pour soi-même et faible intensité	Gratification	Remords
		Conséquence pour soi-même et forte intensité	Gratitude	Colère
Aspect des objets			Amour	Haine

En conclusion les modèles de représentation catégoriels, généralement associés aux approches évolutionnistes, présentent des avantages pour illustrer et comparer des critères des théories de l'évaluation cognitive. Ces modèles permettent de manipuler des termes du langage courant, ce qui facilite la compréhension. De plus, il est envisageable d'associer des émotions à des composantes de la théorie de l'évaluation cognitive. Ainsi, le modèle d'Ekman est particulièrement associé à l'expression motrice, tandis que le modèle OCC est associé aux composantes motivationnelles et cognitives. Il n'en demeure pas moins une forte complexité pour établir des modèles prenant en compte chaque composante. De plus, comme nous l'avons présenté au début de cette partie, il faut considérer que l'utilisation d'un label n'est pas suffisant pour décrire une expérience émotionnelle. Il peut être alors important de considérer d'autres types de représentations pour compléter les modèles catégoriels.

II.4.4 Modèles dimensionnels

Une autre approche de représentation des émotions, consiste à les représenter dans des espaces multidimensionnels et continus. Les émotions prennent alors des valeurs continues dans ces différentes dimensions, permettant d'illustrer des nuances plus fines d'expériences émotionnelles. À l'image des modèles catégoriels, il n'est pas possible de rendre compte de tous les aspects de chaque composante à partir d'un nombre limité de dimensions. Cette approche était déjà adoptée au 19^e siècle par le psychologue Allemand Wilhelm Wundt (1874) qui proposait alors de caractériser les sentiments en opposant : l'excitation et la dépression, la tension et la relaxation, enfin l'agrément et le désagrément.

Selon Sander et Scherer (2009) cette approche dimensionnelle est le plus souvent appariée à la composante du sentiment subjectif. En effet, les chercheurs adoptant cette démarche multidimensionnelle se sont fondés sur les descriptions verbales et sur l'étude de leurs similarités afin d'établir des dimensions bien définies. Deux dimensions ressortent particulièrement des études réalisées (Bakker *et al.*, 2014) :

- La dimension de la « Valence », qui oppose le plaisir et le déplaisir. Il s'agit d'une évaluation du sentiment positif ou négatif global causé par une émotion. Elle se rapproche de la dimension de l'agrément ou désagrément proposée par Wundt.
- La dimension de « l'Activation » ou éveil, souvent nommé *arousal* en anglais, qui représente le niveau de mobilisation d'énergie que provoque l'émotion dans l'organisme. Une activation élevée correspond donc à un état d'excitation extrême et à l'inverse une activation faible à un état en sommeil. Albert Mehrabian (1996) la définit comme un mélange de vigilance et d'activité mentale.

Ces deux dimensions ont été particulièrement popularisées par le circumplex, proposé par le chercheur James Russell (1980). Ainsi, il combine ces deux dimensions

dans un espace circulaire. À partir d'études, où il propose à des étudiants de classer des expressions émotionnelles par rapport à la valence et l'activation, il positionne des libellés émotionnels qui seraient particulièrement saillants pour illustrer chacun des espaces entre les dimensions. Le circumplex est représenté en figure II.9.

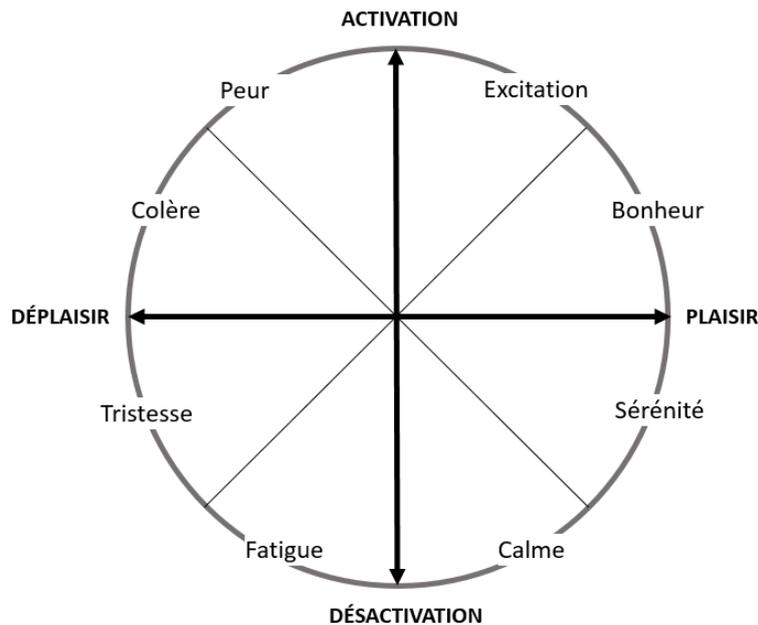


FIGURE II.9 – Modèle de représentation d'émotions de Russell (1980, 2003)

Néanmoins, il peut être compliqué de distinguer la peur de la colère, qui sont chacune des émotions avec un éveil a priori élevé et une valence négative. La dimension de la « Dominance » ou puissance a donc été suggérée par certains chercheurs. Elle se définit comme le sentiment de contrôle. Une dominance élevée indique donc que l'individu a le sentiment de pouvoir influencer la situation en cours, tandis qu'une dominance faible correspond au sentiment de subir la situation. Ainsi, Mehrabian (1996) a proposé le modèle PAD « Plaisir-Activation-Dominance » (traduit de l'anglais *Pleasure-Arousal-Dominance*). De la même manière que Russell, Mehrabian a associé à chacun des sous-espaces du modèle des termes représentatifs et opposés. L'hostilité par exemple, serait négative sur l'échelle du plaisir, élevée en termes d'activation et de dominance, à l'opposé de la docilité qui est élevée en plaisir avec une faible activation et une faible dominance.

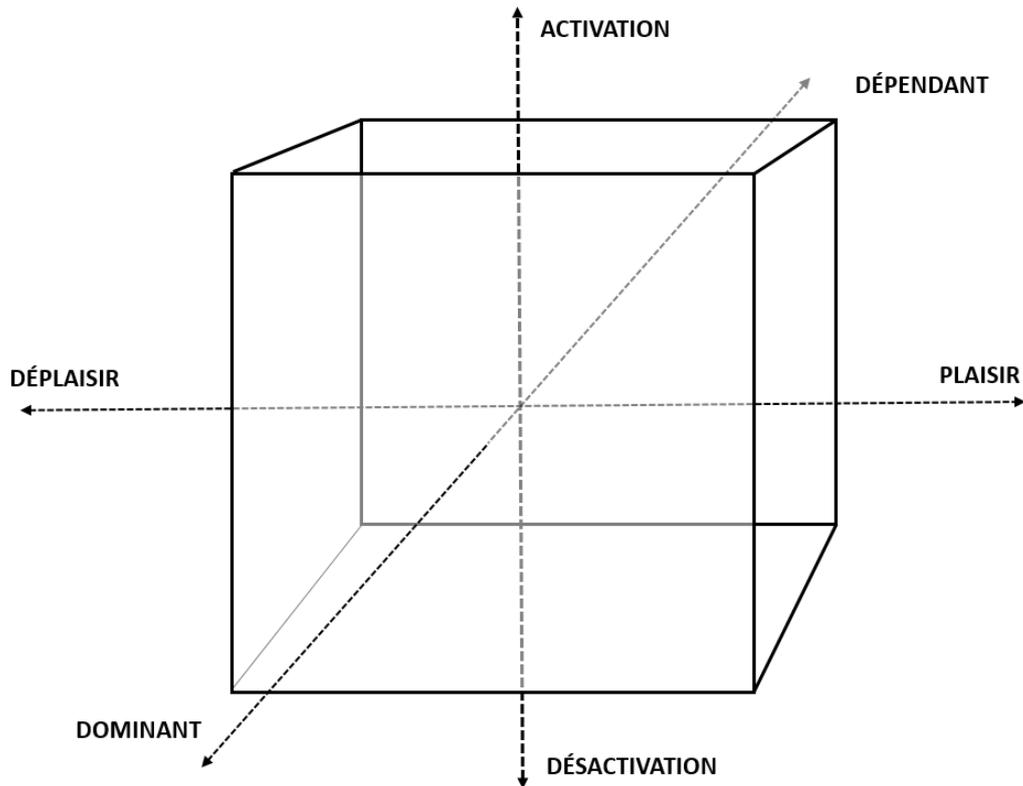


FIGURE II.10 – Modèle PAD proposé par (mehraban1996pleasure), schéma inspiré de Bakker *et al.* (2014)

À l'échelle neurologique, Hugo Lövheim (2012) propose d'associer trois neurotransmetteurs à des dimensions émotionnelles. Ainsi, il associe la dopamine à la récompense et au renforcement et donc à la dimension du plaisir ou déplaisir. La noradrénaline correspondrait aux réponses comportementales de fuite ou de combat et pourrait être appariée à l'activation, en intégrant la vigilance et l'attention. Enfin, la sérotonine pourrait être mise en relation avec la dimension de contrôle et de confiance. Il propose donc une représentation tridimensionnelle et intègre des termes émotionnels provenant des travaux de Tomkins (Tomkins, 1984).

Johnny Fontaine et Klaus Scherer suggèrent de faire le lien entre des composantes du modèle des processus composants à des dimensions du modèle PAD (Fontaine *et al.*, 2007). Ils utilisent la méthode GRID, élaborée par Scherer (2005), un questionnaire comportant 144 descripteurs et des termes émotionnels. Ils mettent en évidence quatre axes et des corrélations marquées entre certains de ces axes et les descripteurs. Les trois premiers axes correspondent aux trois dimensions que nous avons présentées précédemment, auxquels s'ajoute la dimension de l'imprédictibilité, principalement représentée par la surprise. Pour chaque descripteur, des corrélations sont calculées.

Le fait de considérer la situation comme « non congruente avec les buts et objectifs » sera fortement négativement corrélé avec la valence, faisant donc correspondre un critère de l'évaluation cognitive avec une dimension. De même, le fait de transpirer, caractéristique de la composante psychophysiological, est fortement corrélé avec la dimension d'activation.

Les modèles dimensionnels offrent donc la possibilité de représenter une infinité d'expériences émotionnelles sur des échelles continues. Paradoxalement, même si ces modèles permettent de s'extraire de labels limités, leur interprétation implique souvent d'utiliser au final des expressions du langage. Une des difficultés qu'ils soulèvent est d'identifier et définir les axes qui les composent pour offrir des représentations compréhensibles et pratiques. Cependant, ils offrent un moyen très efficace pour synthétiser des expériences émotionnelles. Ils sont particulièrement utilisés dans la recherche pour créer des outils d'annotations, afin d'aider un sujet à déclarer son sentiment subjectif ou pour classer du matériel d'induction.

II.4.5 Discussion

Il est envisageable, d'une part de lister des critères permettant de distinguer les émotions des autres phénomènes affectifs tels que les humeurs ou les dispositions affectives et, d'autre part, de proposer des modèles de représentation des émotions pour en extraire des points de ressemblance ou d'opposition. Bien que représenter et classer des émotions implique de réduire la réalité de l'épisode émotionnel, complexe et dynamique, dans un ensemble plus limité, cela permet de les étudier. Par conséquent, il nous semble indispensable d'utiliser ces approches pour établir une méthodologie, permettant de différencier des mesures dans le cadre d'expérimentations.

En ce qui concerne les situations de communication, les deux types de représentations présentent des avantages certains. Globalement, les modèles dimensionnels nous semblent offrir une plus grande flexibilité par rapport aux modèles catégoriels. En effet, au vu de la complexité à définir des épisodes émotionnels, pouvoir les représenter selon des dimensions facilite leur classification. Dans le cadre d'un service client, la dimension de la valence est essentielle pour inférer, de manière synthétique, la satisfaction des clients. Les autres dimensions peuvent représenter des informations complémentaires pour mieux s'adapter à leurs réactions. Un client ressentant une forte activation nécessitera, sans doute, une réponse différente de celle à proposer à un client peu activé. En outre, elles peuvent faciliter la compréhension des facteurs à l'origine d'un épisode émotionnel, notamment par l'intermédiaire de modèle de classification comme le modèle OCC (Ortony *et al.*, 1990). C'est pourquoi les représentations catégorielles semblent pertinentes, car elles sont plus proches du langage courant. L'utilisation combinée des deux types de représentations est donc souhaitable.

II.5 Conclusion et positionnement

L'ensemble de ce chapitre a permis d'aborder la science des émotions d'un point de vue théorique. Dans un premier temps, nous pouvons considérer que la science des émotions n'est pas une discipline homogène et consensuelle. En effet, les différentes approches historiques ont été marquées par des débats, pour certains encore d'actualité, qui ont divisé les chercheurs. Les concepts abstraits et propres aux sciences humaines, comme le sont la cognition ou l'émotion, sont particulièrement sources de discussion lorsqu'il s'agit de trouver une définition satisfaisante. L'émotion en elle-même, concept relativement récent, a peiné à trouver sa place dans l'étude de la psychologie humaine et animale. Sa place dans l'organisme, par rapport au système nerveux ou périphérique, a été l'objet de théories et de propositions tout au long du 19^e siècle. De même, la question de l'inné ou de l'acquis entre les approches évolutionnistes et socio-constructivistes continue de diviser les chercheurs. Enfin, bien que sa relation avec la cognition ait été source de confusion, elle a participé à l'émergence de nouvelles propositions théoriques. L'ensemble des débats et des oppositions a permis de mettre en évidence les limites des définitions et de construire des théories plus à même de répondre à la complexité des phénomènes émotionnels. Ainsi, l'histoire de l'étude des émotions, loin d'être achevée, est riche d'une littérature scientifique foisonnante. Elle met en évidence la profondeur des questionnements sur cette facette incontournable de la nature humaine.

Devant la difficulté à définir et encadrer ce que sont les émotions, il nous a semblé essentiel d'établir un cadre théorique solide. La théorie de l'évaluation cognitive, couplée avec le modèle des processus composants, propose de considérer les processus émotionnels comme des épisodes dynamiques et le produit d'un ensemble de composantes. Dans le contexte du service client, ces éléments théoriques nous semblent appropriés. Premièrement, comme il s'agit de situations évoluant avec le temps et les événements, il est pertinent de considérer les émotions comme des épisodes dynamiques plutôt que comme des états sur un instant. Deuxièmement, la place centrale de la cognition au sein du processus émotionnel et plus particulièrement des critères d'évaluation des stimuli, permet d'expliquer les réactions des clients ou des conseillers en fonction de leurs buts à atteindre dans la communication. Enfin, les différentes composantes nous permettent d'identifier les modalités à recueillir pour développer des outils de reconnaissance automatique. Dans le cadre de cette thèse, nous porterons particulièrement notre attention sur les composantes de l'expression motrice, du sentiment subjectif et des réponses psychophysologiques. En effet, ces composantes constituent l'ensemble des manifestations les plus visibles et les plus mesurables d'un épisode émotionnel. C'est pourquoi la théorie de l'évaluation et le modèle des processus composants constituent un cadre théorique adapté pour nos recherches. Nous proposons d'en reprendre les définitions et les concepts pour élaborer nos protocoles expérimentaux et interpréter les résultats.

Enfin, il est nécessaire d'établir des modèles de représentation des émotions pour

accorder ces phénomènes avec des schémas de pensée plus intelligible. La variété des épisodes émotionnels étant infinie, nous devons donc les réduire, soit dans un ensemble de labels, soit par l'intermédiaire d'espaces multidimensionnels. Cependant, le choix d'un modèle de représentation implique la prise en compte du contexte dans lequel vont se produire les épisodes émotionnels. Les deux types de représentations sont complémentaires pour le contexte du service client. Les modèles dimensionnels offrent une flexibilité élevée pour concevoir des conditions expérimentales et différencier des caractéristiques globales. Tandis que les modèles catégoriels permettent des analyses avec une granularité supplémentaire pour comprendre les causes d'une émotion.

Dans le chapitre [III](#), nous aborderons l'étude des émotions d'un point de vue expérimental, plus particulièrement dans le domaine des interfaces homme-machine.

Chapitre III

Étude des émotions et nouvelles technologies

Sommaire

III.1	Introduction	70
III.2	Informatique affective	71
III.2.1	Définition et enjeux	71
III.2.2	Application au contexte du service client	73
III.2.3	Expression des émotions dans le cadre des messageries instantanées	74
III.3	Reconnaissance automatique des émotions	76
III.3.1	Introduction	76
III.3.2	Modalité visuelle	77
III.3.3	Modalité audio	80
III.3.4	Modalité textuelle	82
III.3.5	Modalité physiologique	84
III.3.6	Approche multimodale	87
III.3.7	Difficultés en reconnaissance d'émotions	88
III.4	Reconnaissance automatique d'émotions et interfaces homme-machine	90
III.4.1	Introduction	90
III.4.2	Ordinateur	91
III.4.3	Smartphone	93
III.4.4	Discussion	100
III.5	Informatique affective et contexte expérimental	102
III.5.1	Protocoles expérimentaux	102
III.5.2	Induction d'émotions	103
III.5.3	Mesures déclaratives	107
III.5.4	Discussion	108

III.1 Introduction

Au cours du chapitre II, nous avons exploré les principales théories sur l'étude des émotions. Plus spécifiquement, nous avons opté pour les théories de l'évaluation cognitive comme base théorique de notre projet de recherche. Cela nous a donc permis de proposer une définition de l'expérience émotionnelle et de mettre en évidence l'ensemble de ses composantes et de leurs modalités permettant de la mesurer. De plus, nous avons listé différentes approches de représentations des états émotionnels et trouvé dans la littérature les éléments fondant les observations futures.

Afin de répondre à nos problématiques de recherche, ancrées sur les usages des technologies de l'information et de la communication (TIC), nous accordons donc un chapitre d'état de l'art plus spécifique à l'étude des émotions dans ce contexte particulier. Pour commencer, l'émergence des TIC est à l'origine de nouvelles formes de pratiques. Cet usage des outils numériques, de plus en plus inscrit dans nos quotidiens, constitue en lui-même une source et un terrain d'expression d'émotions. En effet, les technologies ont pour rôle de nous aider à réaliser des actions. En ce sens elles augmentent et étendent nos possibilités. Parallèlement, elles nous contraignent aussi dans leur logique d'usage et leurs modalités. Ainsi, les interactions créées par les nouvelles technologies sont source d'émotions spécifiques. La communication par l'intermédiaire des smartphones ne fait pas exception. Tout d'abord, parce qu'il s'agit d'une situation sociale et donc à même de nous faire ressentir ou exprimer des émotions. Ensuite, parce qu'elle nous impose ses limites techniques et réduit nos capacités à communiquer naturellement. Cela se concrétise particulièrement sur le terrain de recherche de la relation client. En effet, l'ensemble des canaux disponibles pour communiquer avec les conseillers clients d'une entreprise se sont multipliés (emails, réseaux sociaux, messagerie instantanée). Chacune de ces modalités possède des avantages pratiques pour les clients et les entreprises (accessibilité, coûts ...). En revanche, elles conditionnent les interactions et par conséquent les moyens d'expression des émotions. Ainsi, il est important de mieux comprendre l'expression des émotions, afin de concevoir des procédés pour pallier ces limites dans la transmission via les TIC.

En plus d'être un contexte propice à la production d'émotions, les nouvelles technologies rassemblent des capteurs et des outils propices au recueil des informations sur les utilisateurs. En prenant en compte ces nombreuses données, nous pouvons imaginer qu'elles puissent servir à traduire les comportements des utilisateurs. Elle pourraient donc permettre d'inférer des caractéristiques des épisodes émotionnels vécus. Pour autant, les données recueillies sont complexes à analyser et le recours à des techniques d'intelligence artificielle nous semble pertinent pour progresser. Par exemple, nous

pouvons citer les méthodes d'apprentissage automatique (*machine learning* en anglais) pour mettre en évidence des patterns spécifiques ou des descripteurs particulièrement pertinents pour inférer une expérience émotionnelle. De plus, la mise en place de modèles d'apprentissage nécessite la création de corpus de données labellisées avec des expériences émotionnelles.

L'objectif de ce chapitre est donc de répondre aux questions suivantes :

- Comment étudier les émotions dans le contexte des TIC ?
- Quels sont les systèmes de reconnaissance automatique d'émotions adaptés ?
- Comment réaliser des bases de données pour concevoir des systèmes de reconnaissance automatique d'émotions ?

Nous introduirons la discipline de l'informatique affective, à l'origine de nombreux travaux sur l'étude des relations entre nos émotions et les technologies. Nous présenterons ensuite les méthodes existantes pour la reconnaissance d'émotions et plus spécifiquement dans le cadre des interactions homme-machine. Enfin, nous présenterons de manière plus détaillée les méthodes pour étudier les émotions dans le cadre d'expérimentations, avec l'objectif de créer des corpus de données exploitables.

III.2 Informatique affective

III.2.1 Définition et enjeux

L'*affective computing*, que nous nommerons informatique affective en français, est un champ de recherche relativement récent. Le terme prend sa source à partir d'une publication de Rosalind Picard, une chercheuse du MIT (Picard *et al.*, 1995). Elle popularisa cette expression pour désigner l'étude des émotions humaines dans le cadre des technologies de l'information et de la communication. Dans cet article, Picard fait le constat que les émotions ne sont majoritairement pas prises en compte dans la conception des outils informatiques. Par extension, elle considère que la science néglige le concept d'émotion en l'opposant à la pensée rationnelle. En s'appuyant sur des travaux en neurosciences, elle présente les émotions comme étant en fait essentielles pour la prise de décision, au même titre que des processus cognitifs rationnels. Elle propose donc de développer des systèmes informatiques « intelligents », en mesure de comprendre les émotions des utilisateurs pour s'adapter à leurs besoins. En effet, de la même manière que nous nous adaptons aux autres en reconnaissant leurs émotions, les machines devraient prendre en compte cette information pour s'adapter à nos usages. Ainsi, selon Picard la prise en compte des états affectifs dans nos interactions avec les machines aurait de nombreuses applications, par exemple dans l'éducation, le divertissement, la santé et les interactions humaines...

L'informatique affective sera donc le nom donné à l'ensemble des travaux impliquant, d'une part de concevoir des outils de détection d'états affectifs pour adapter les outils informatiques, et d'autre part la conception de machines en mesure d'exprimer des émotions. Soujanya Poria *et al.* (2017) définit l'informatique affective de la manière suivante :

« Affective computing is an emerging field of research that aims to enable intelligent systems to recognize, feel, infer and interpret human emotions. It is an interdisciplinary field which spans from computer science to psychology, and from social science to cognitive science. »

L'informatique affective est donc, par nature, un domaine multidisciplinaire. Comme le présente Picard (2010), elle implique la collaboration d'expertises qui peuvent paraître éloignées, comme la psychologie et l'ingénierie. Elle distingue trois types d'applications (Picard, 2015) :

- La mesure des états émotionnels des utilisateurs, l'objectif étant d'améliorer les technologies en prenant en compte la frustration ou la satisfaction. Les ingénieurs pourraient alors se servir de ces mesures pour mieux concevoir les outils afin d'améliorer l'expérience des utilisateurs.
- Le développement de systèmes en mesure d'exprimer des émotions reconnaissables pour des humains. Cela permettrait d'engager les utilisateurs en induisant des émotions positives, par exemple dans le cadre d'interactions avec des robots ou des chatbots.
- De plus, la conception des technologies peut être améliorée, en s'inspirant du fonctionnement des émotions d'un point de vue biologique. Les systèmes informatiques pourraient alors aussi ressentir des émotions avec l'optique des prendre des décisions adaptées au contexte.

Selon Rosalind Picard, il est possible de prendre en compte plusieurs modalités pour détecter les émotions d'un utilisateur : les expressions du visage, les gestes, la voix ou encore les modifications physiologiques du corps. Étant donné qu'il n'est pas possible d'être dans la tête des utilisateurs, il est essentiel de diversifier les types de données récoltées. Les seules expressions et modifications physiologiques ne sont pas suffisantes. Il serait nécessaire, selon Picard, que les systèmes informatiques accèdent au contexte. Cela comprend l'information ambiante, comme la température de la pièce et les informations cognitives, comme les dernières actions que vient de réaliser l'utilisateur (Picard *et al.*, 1995). D'un point de vue théorique, les chercheurs de l'informatique affective ont donc tendance à s'appuyer sur les théories de l'évaluation cognitive (R. Calvo *et al.*, 2015). En effet, beaucoup estiment que la cognition est une composante essentielle à prendre en compte dans le cadre d'une interaction homme-machine et par extension des émotions ressenties dans ce contexte. Cependant, à l'image de la recherche sur les émotions, il n'y a pas de consensus sur ce qu'est une émotion. La

complexité théorique des émotions constitue donc une problématique de ce domaine d'étude.

En 2005, Jianhua Tao et Tan (2005) notaient plusieurs problématiques à l'informatique affective : le manque de corpus de données pertinentes, la complexité d'intégrer des données provenant de différentes modalités et la difficulté de s'adapter à des scénarios de la vie courante. Ils envisagent l'importance, qui se confirme de nos jours, de l'informatique ubiquitaire qui permettra de recueillir des données émotionnelles. Ces dernières années, l'idée d'une informatique ubiquitaire, qui se définit comme l'informatique embarquée dans des objets du quotidien, s'est particulièrement concrétisée au travers des smartphones. Cependant, malgré l'essor de la quantité d'informations recueillies, il reste encore la difficulté d'obtenir des corpus labellisés d'une part, et d'autre part de traiter des données provenant de différents canaux pour en extraire une émotion.

Selon une étude bibliométrique de Maria Helena Pestana *et al.* (2018), l'informatique affective aurait connu un essor important. En effet, en 2000 le nombre annuel de publications dans le domaine est de 20 et augmente jusqu'à plus de 200 publications par an à partir de 2013. 54,7% des publications proviendraient de chercheurs issus des domaines de l'ingénierie ou de l'informatique. Enfin, les sujets les plus courants portent sur la reconnaissance d'émotions, les interactions homme-machine et les environnements informatiques pour l'apprentissage humain. Avec un peu plus de 20 ans d'existence, l'informatique affective est un domaine en pleine expansion. Avec la volonté de faciliter les interactions entre les hommes et les machines, par l'intermédiaire des émotions, elle rapproche des scientifiques de communautés variées et s'intéresse à de nombreux contextes. L'augmentation des données récoltées par les nouvelles technologies et les avancées des méthodes de traitement de ces informations offrent des perspectives séduisantes pour l'avenir de ce domaine de recherche.

III.2.2 Application au contexte du service client

L'informatique affective est majoritairement dirigée vers le domaine de l'éducation, de la santé, des réseaux sociaux ou de la robotique. Le domaine particulier des services clients n'a été que peu abordé. Pourtant, la gestion de la relation client ou *customer relationship management* (CRM) est particulièrement concernée par la numérisation et l'évolution des technologies de l'information et de la communication (TIC). Un CRM peut être défini de la manière suivante :

« A coherent and complete set of processes and technologies for managing relationships with current and potential customers and associates of the company, using the marketing, sales and service departments, regardless of the channel of communication » (Chalmers, 2006)

Selon cette définition, la gestion de la relation client désigne donc un ensemble de moyens organisationnels et d'outils technologiques pour interagir avec les clients d'une entreprise. L'objectif est de maximiser la satisfaction de ces clients et de répondre efficacement à leurs demandes et besoins. En plus des moyens de communication conventionnels comme la boutique, le téléphone ou le courrier, les CRM se sont diversifiés pour adopter de nombreux canaux. Ainsi, de nouveaux canaux de contact ont émergé au cours des dernières années. En effet, il est désormais possible de communiquer avec une entreprise par le biais d'emails, des réseaux sociaux ou par l'intermédiaire de chats en direct sur ordinateur ou par smartphone.

Les entreprises se dirigent à présent vers l'automatisation de la communication par le biais d'agents conversationnels virtuels ou *chatbots* en anglais. Selon Viot et Bressolles (2012), l'intégration de *chatbots* dans les canaux d'un CRM permettrait de réduire les coûts opérationnels et d'améliorer l'accessibilité de l'information (les *chatbots* étant disponible 24h sur 24). Cependant, bien qu'ils soient capables de répondre à des sollicitations simples, la qualité de leurs réponses et leur capacité à s'adapter aux clients est encore limitée et reste perfectible. Ainsi, selon Julia Velkovska et Beaudouin (2014) bien que les agents conversationnels répondent dans de nombreux cas aux besoins des clients, ils demandent une adaptation supplémentaire de la part de ces derniers pour comprendre leur fonctionnement. À l'inverse les conseillers humains sont capables de s'adapter plus facilement à leur interlocuteur.

Ainsi, la relation client constitue un terrain d'étude à part entière pour l'informatique affective. Identifier et mesurer les émotions des clients ou des conseillers représente un enjeu important, dans l'optique d'améliorer les outils de communication existants. En effet, la prise en compte des émotions est une composante essentielle pour s'adapter aux besoins des clients. Cette démarche pourrait aussi permettre de faire progresser les interactions avec des agents virtuels ou aider les conseillers clients à mieux saisir les émotions des clients dans leurs réponses, quel que soit le canal de communication utilisé.

III.2.3 Expression des émotions dans le cadre des messageries instantanées

Dans le cadre de cette thèse, nous nous intéressons à l'usage du smartphone et plus particulièrement au canal des messageries instantanées (*chat* en anglais) qui se développe. Ahmed Elmorshidy (2013) propose de définir les messageries instantanées de la manière suivante :

« Online chat may refer to any kind of communication over the Internet which offers an instantaneous transmission of text-based messages from sender to receiver. » (Elmorshidy, 2013)

Ce canal offre la possibilité d'établir une communication entre un conseiller et un client ou entre un agent virtuel et un client. Cette modalité est moins engageante qu'une communication par téléphone, tout en permettant de conserver une trace écrite des interactions. De plus, elle permet d'interagir de manière synchrone et donc d'éviter les délais de réponse longs, qui peuvent se produire avec les emails. Enfin, pour les entreprises, la messagerie instantanée est perçue comme un moyen rentable d'augmenter les interactions sociales avec les clients et de personnaliser l'expérience des utilisateurs.

Les messageries instantanées représentent une situation d'interaction sociale. Or, l'expression des émotions est une composante essentielle dans les interactions humaines et selon l'informatique affective, dans les relations homme-machine. Pour Dacher Keltner et Haidt (1999), l'expression des émotions dans les communications dyadiques (entre deux individus) aide les interlocuteurs à se coordonner, en facilitant la compréhension des intentions. Elle permet au récepteur d'obtenir des informations essentielles sur l'émetteur. De plus, cette expression régule les comportements des interlocuteurs, afin d'atteindre le même but. Elle récompense les comportements dirigés vers l'objectif et réprime les comportements contre-productifs. Dans le cadre d'une interaction entre un conseiller et un client, la communication étant dirigée vers un objectif particulier, l'expression d'émotions est d'autant plus marquée et traduit l'intention des deux participants. Ces expressions se concrétisent en partie par des interactions non verbales. Un sourire, par exemple, pouvant récompenser un comportement désirable pour l'atteinte de l'objectif.

Cependant, les communications par l'intermédiaire de messageries instantanées sont particulièrement restreintes pour l'expression d'émotions. Pour Nadia Gauducheau (2008), l'absence de comportements non verbaux et l'anonymat des interlocuteurs sont particulièrement impactants. Elle soulève alors trois problématiques qui découlent de ces caractéristiques :

- L'absence de connaissance de son interlocuteur peut limiter la volonté à s'exprimer émotionnellement. En effet, en l'absence de connaissances sur la personne en face, les normes sociales à appliquer sont plus incertaines.
- L'impossibilité d'exprimer ces émotions de manière non verbale restreint les moyens d'expression d'émotions. Au final, les interlocuteurs ne disposent que de la modalité textuelle pour s'exprimer.
- Conséquence des deux premiers points, les récepteurs ont peu d'indices pour inférer les émotions de l'autre.

Selon la théorie du renforcement des indices sociaux, présentée par Walther (1996), bien que l'expression des émotions soit bien limitée par le format du canal, les individus peuvent tout de même en exprimer une partie. En effet, il avance que les interlocuteurs sont en mesure de mettre en place des stratégies pour compenser la perte d'indices

émotionnels. Premièrement, le format par messagerie instantanée offre la possibilité de préparer et modifier son message à l'avance, afin de choisir des termes appropriés pour transmettre correctement ses intentions et attitudes. Deuxièmement, l'utilisation d'émoticônes est une solution pratique pour exprimer une émotion ou pour ajuster le propos d'une phrase. De la même manière, les réseaux sociaux et applications de messagerie instantanée incluent de nos jours l'intégration d'images ou d'images animées pour exprimer des émotions.

Ces stratégies, même si elles permettent de compenser en partie la perte de la communication non verbale, ne remplacent pas entièrement une communication en face à face. En effet, un temps d'appropriation des TIC est nécessaire pour mettre en place des moyens comme des émoticônes. Or, tous les clients ne sont pas familiers de ce type de communication. De plus, l'intégration de ces stratégies impose une charge cognitive supplémentaire et empêche une expression spontanée des émotions, comme dans le cas du non verbal. Ainsi, il nous semble essentiel d'explorer la possibilité de concevoir et d'intégrer des systèmes de reconnaissance automatique des états émotionnels des interlocuteurs.

III.3 Reconnaissance automatique des émotions

III.3.1 Introduction

Comme nous l'avons présenté dans la sous-section [II.3.1](#), les expériences émotionnelles se traduisent par l'évolution d'un ensemble de sous-systèmes de l'organisme. Nous avons alors décrit chacune de ces composantes et leurs caractéristiques, qui représentent l'ensemble des manifestations provoquées par un épisode émotionnel. Ces manifestations constituent des sources de données potentielles pour concevoir des outils de reconnaissance automatique d'émotions. Le progrès dans ce domaine est porté par les avancées en termes de capteurs, de traitement des signaux et de méthodes d'intelligence artificielle (Poria *et al.*, 2017). Pour créer des bases de données pertinentes, il est nécessaire, d'une part, d'obtenir de larges quantités de données. La captation de données est facilitée par la miniaturisation des capteurs (caméra, micro, capteurs physiologiques...), l'amélioration de leurs performances et leur intégration dans des outils informatiques (smartphones, ordinateurs portables, montres connectées...). D'autre part, ces données doivent être labellisées pour être exploitables. Nous présenterons cela de manière plus détaillée dans la sous-section [III.5.1](#) de ce chapitre. Enfin, les données étant complexes (des matrices de pixels pour des images), il est nécessaire de les prétraiter et d'extraire des descripteurs significatifs, cela demande donc de la puissance de calcul et des méthodes d'intelligence artificielle pour classer les descripteurs.

Selon Soujanya Poria et Anvita Saxena, nous pouvons globalement diviser les sources de données pour la recherche sur la détection d'émotions en quatre modalités (Poria *et al.*, 2017; Saxena *et al.*, 2020) :

- La modalité visuelle : principalement orientée vers la reconnaissance des expressions du visage, mais aussi des gestes et des postures. Les données exploitées proviennent de vidéos ou d'images.
- La modalité audio : portant sur la voix et plus particulièrement sur la parole. Cela peut donc inclure à la fois ce qui est dit et comment cela est dit, en analysant la prosodie du discours.
- La modalité textuelle : à savoir la compréhension des émotions exprimées dans un texte. Cette approche s'intègre dans la discipline du traitement automatique du langage (*Natural Language Processing* en anglais). Elle est plus connue par sous le nom d' « analyse de sentiments » (*sentiment analysis* en anglais).
- La modalité physiologique : ils s'agit de capter les variations des signaux physiologiques au cours d'un épisode émotionnel. Cela comprend donc la conductance électrodermale, la fréquence cardiaque ou les signaux des électroencéphalogrammes.

Selon Poria, les modalités visuelles, audio et textuelles constitueraient 90% de la littérature sur le sujet de la reconnaissance automatique d'émotions. Au vu de la représentation forte de ces trois modalités, il nous semble essentiel de présenter les travaux de ces domaines, leurs avantages ainsi que leurs limites. De même, nous présenterons la modalité physiologique, de plus en plus abordée dans cette discipline.

III.3.2 Modalité visuelle

L'expression faciale est l'objet d'étude le plus représentatif de cette catégorie. Tout comme nous l'avons présenté dans la composante des expressions motrices dans la sous-section II.3.6, il s'agit d'une porte d'entrée naturelle pour étudier les émotions. L'émergence des techniques issues de l'intelligence artificielle a particulièrement accéléré la recherche dans ce domaine. Les domaines d'application sont nombreux : les interfaces hommes-machines, la réalité virtuelle, la réalité augmentée, les systèmes d'assistance à la conduite, la robotique...

Le nombre de bases de données conçues pour la reconnaissance d'émotions à partir des expressions faciales s'est multiplié au cours des dernières années. Dhwani Mehta *et al.* (2018) décrivent les différentes méthodes pour acquérir les données d'expressions faciales et le type de données enregistrées. Ainsi, il existe deux approches pour la création du corpus :

- Les émotions sont simulées par des acteurs. Cela permet de faciliter l'acquisition et l'annotation des données et d'obtenir des expressions faciales plus marquées. En

revanche, les émotions exprimées sont artificielles et ne sont donc pas forcément représentatives de la réalité.

- Les émotions sont exprimées de manière spontanée. Cette approche est beaucoup plus complexe à mettre en place, car elle implique de filmer des sujets pendant de longues périodes et de ne conserver que les instants où des émotions ont été exprimées.

Les deux approches peuvent être combinées pour former des bases de données diversifiées. De plus, il est nécessaire d'inclure dans les corpus des individus de différents genres, couleurs de peau et âges pour limiter les biais.

La grande majorité des corpus de données créés sont labellisés en fonction d'émotions catégorielles, principalement les six émotions basiques d'Ekman et Friesen (1971), ainsi qu'une expression faciale neutre. Néanmoins, certains chercheurs proposent d'annoter les expressions faciales selon des dimensions continues. C'est le cas par exemple de l'*OMG dataset*, dans lequel les émotions sont annotées non seulement à partir de sept émotions basiques, mais aussi en termes de valence et d'activation. (Barros *et al.*, 2018).

Une fois les données enregistrées et annotées dans un corpus, l'objectif est d'entraîner des modèles de classification automatique à distinguer automatiquement des patrons spécifiques entre chaque émotion. Byoung Chul Ko (2018) propose de différencier deux méthodes de classification (représentées sur le schéma III.1) :

- L'approche conventionnelle : elle consiste à identifier les zones du visage et en extraire des descripteurs géométriques. Ces descripteurs peuvent correspondre à des distances entre des points de repère particuliers (des repères sur la bouche ou les yeux, par exemple). Une fois les descripteurs construits, ils sont transformés afin d'être exploitables pour les modèles de classification. Cela comprend à la fois une mise en forme dans un format numérique, une normalisation et la suppression des données incomplètes. Les descripteurs sont ensuite triés pour ne sélectionner que les plus informatifs. Des classifieurs sont ensuite entraînés à distinguer les différents labels d'émotions à partir des descripteurs sélectionnés et sur une partie réduite du corpus. Il existe de nombreux classifieurs pouvant être utilisés pour de l'apprentissage automatique : régression logistique, arbre de décision, machine à vecteur de support, réseaux de neurones... Les classifieurs sont ensuite testés avec le reste du corpus, une performance de classification correcte est évaluée en fonction du nombre d'images ou de vidéos qui ont été classées dans la bonne catégorie émotionnelle.
- L'approche d'apprentissage profond : elle suit globalement le même processus que l'approche conventionnelle. Cependant, elle se distingue par le fait de ne pas avoir à extraire en amont les descripteurs des données d'entrée. En effet, les réseaux de neurones profonds vont calculer des descripteurs par des traitements convolutifs.

Concrètement, à l'image du fonctionnement du cortex visuel dans le cerveau, les premières couches des réseaux de neurones vont réarranger les images selon des angles ou des patterns distinctifs pour former des descripteurs pertinents. Les dernières couches vont ensuite utiliser ces informations pour classer les différentes expressions faciales. Nous définissons alors cette approche comme bout à bout (*end to end* en anglais), car les réseaux de neurones profonds traitent des données à l'état brut jusqu'à l'entraînement du modèle.

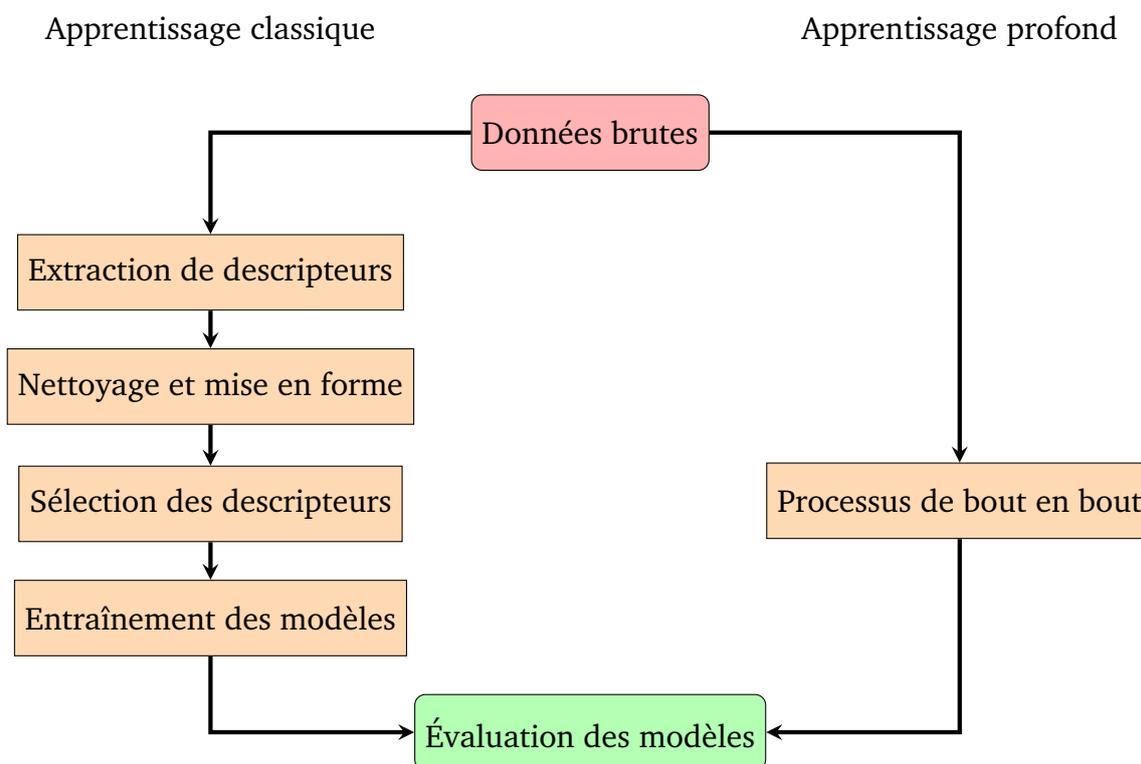


FIGURE III.1 – Description des différentes approches d'apprentissage automatique

Bien que la majorité des études sur la modalité visuelle porte sur les expressions faciales, il existe aussi des travaux sur les gestes ou les postures. Le principal inconvénient étant une quantité de données récupérées plus élevée et une plus grande difficulté à établir des points de référence, à cause notamment des vêtements (Dzedzickis *et al.*, 2020). Des caméras avec capteurs de profondeur permettent de réaliser des captures des mouvements gestuels ou des postures. Elles sont notamment devenues plus accessibles avec des modèles comme Kinect, développé par Microsoft (Kinect, 2020). Leur utilisation permet d'extraire le squelette 3D d'un participant et de calculer des descripteurs à partir des points de repère. En captant des mouvements en continu il est alors possible de calculer des descripteurs à partir des accélérations des points.

Par exemple, Stefano Piana et son équipe ont calculé des descripteurs à partir des mouvements des jointures des mains, des épaules ou des coudes. (Piana *et al.*, 2014).



FIGURE III.2 – Caméra kinect de l'entreprise Microsoft (Kinect, 2020), licence creative commons [CC BY 3.0](#)

À l'image des études des expressions faciales, les données sur les gestes et postures sont généralement annotées à partir d'une liste d'émotions discrètes. La démarche pour la classification automatique est similaire à celle appliquée aux expressions du visage. Une fois les descripteurs extraits et sélectionnés, ils entraînent des modèles de classification. Les gestes et expressions du visage peuvent d'ailleurs être combinés afin de former des bases de données plus complètes (Caridakis *et al.*, 2007).

Pour Andrius Dzedzickis, les modalités visuelles présentent l'avantage d'être des méthodes de reconnaissance d'émotion sans contact. Les données sont relativement simples à enregistrer avec des caméras, qui sont de plus en plus accessibles et perfectionnées (Dzedzickis *et al.*, 2020). Il est envisageable de réaliser des applications de détection d'émotions en temps réel ou en différé. En revanche, la quantité de données nécessaire pour ces traitements implique des besoins en stockage et en ressources importants. De plus, la modalité visuelle des émotions est plus pertinente pour des émotions fortement marquées, les corpus existants sont limités par le manque d'émotions exprimées de manière spontanée. Néanmoins, ce domaine de recherche est en pleine expansion et s'avère offrir des perspectives intéressantes en couplant ces données avec d'autres informations émotionnelles.

III.3.3 Modalité audio

La voix est connue pour véhiculer de nombreuses informations sur une personne et sur son état. À partir d'un extrait d'enregistrement vocal, il est possible d'estimer l'origine culturelle de la personne en utilisant des informations linguistiques, d'inférer son âge ou son genre, voire de prédire son état émotionnel. Ainsi, Monorama Swain et ses collègues divisent les caractéristiques de la voix en trois facteurs : « ce qui est dit », « comment cela est dit » et « qui le dit » (Swain *et al.*, 2018). L'ensemble de ces facteurs permet d'obtenir de nombreuses informations sur une personne.

De la même manière que pour les modalités visuelles, des corpus de données ont été créés afin d'entraîner des modèles à identifier les émotions à partir d'extraits de voix. Selon Swain, il existe alors trois moyens pour créer des bases de données de ce type : de manière naturelle, de manière simulée ou de manière induite. Dans le premier cas, les extraits audio peuvent être récupérés dans les médias, des centres d'appels ou des enregistrements de discussions entre des patients et des médecins... Les bases de données simulées sont enregistrées à partir de performances d'acteurs professionnels qui reproduisent des émotions. Les émotions sont alors plus marquées, mais ne sont pas naturelles. Enfin, les données issues d'émotions induites sont créées en provoquant artificiellement des émotions chez des participants enregistrés. Les émotions sont annotées en majorité avec une liste d'émotions catégorielles, bien qu'il existe aussi des corpus annotés sur des dimensions continues comme la valence, l'activation et la puissance. De plus, les bases de données varient en fonction du nombre de langues, du nombre de participants et de leurs caractéristiques en termes d'âge, de genre ou d'accent.

Étant donné que le signal de la voix évolue dans le temps, il est nécessaire de segmenter les enregistrements en échantillon de taille fixe pour ensuite extraire les descripteurs. Une fois les descripteurs extraits, la démarche est alors similaire à l'approche conventionnelle présentée pour la modalité visuelle (voir sous-section III.3.2), des classifieurs sont entraînés et testés à partir des descripteurs les plus pertinents. En parallèle, les techniques d'apprentissage profond se développent aussi, notamment par l'intermédiaire des réseaux de neurones dits récurrents, particulièrement adaptés pour les analyses de données temporelles. Ces réseaux sont en mesure d'extraire des descripteurs en prenant en compte les informations précédentes ou consécutives d'un instant, permettant d'identifier des patterns (Poria *et al.*, 2017).

Les applications sur les caractéristiques émotionnelles de la voix sont nombreuses : des outils de diagnostic pour la santé, des outils de traduction automatique, les communications avec des agents conversationnels vocaux et plus spécifiquement dans le cas des services clients avec les centres d'appel. Donn Morrison Morrison *et al.* (2007) ont réalisé des classifications automatiques sur des bases de données de voix provenant de centres d'appels et annotés selon des émotions. Ils obtiennent des résultats proches des 77% de bonne classification pour distinguer les voix de personnes énervées ou neutres.

Tout comme la détection d'émotions à partir des visages ou des gestes, la voix constitue un objet d'étude à part entière avec ses propres avantages, inconvénients et problématiques. Des données sont disponibles en grande quantité. Néanmoins, actuellement les bases de données sont majoritairement issues de corpus d'émotions simulées et avec des listes d'émotions catégorielles. De plus, elles ne comprennent souvent qu'un nombre réduit de locuteurs et de langues représentées. Les travaux sur la voix sont donc confrontés à des limites similaires aux recherches sur les modalités

visuelles, à savoir la difficulté d'obtenir des modèles généralisables à des populations diversifiées. Cependant, l'étude de la voix est complémentaire aux modalités visuelles et permet d'acquérir des informations utiles dans des contextes d'échanges vocaux. Elle peut être combinée avec le contenu textuel des conversations afin d'obtenir des informations subjectives ou contextuelles.

III.3.4 Modalité textuelle

Depuis l'essor de l'utilisation des réseaux sociaux, des notations dans les sites de e-commerce ou la création de blogs, la quantité de texte disponible est très importante et accessible. Ces corpus de texte sont complexes à synthétiser, mais offrent des sources de données pertinentes dans de nombreuses situations. Les différentes techniques servant à analyser de manière automatique des bases de données textuelles sont regroupées sous l'étiquette de *Traitement Automatique du Langage* (TAL). Une partie de ces traitements est tournée vers l'analyse des émotions des auteurs des textes. Dans le cadre du TAL, on parle alors d'analyse du sentiment, car il s'agit de la composante exprimée des émotions, associée donc au sentiment subjectif. Plus spécifiquement, il peut s'agir de comprendre l'opinion des auteurs. L'opinion se définit d'après Bing Liu (2017), comme un ensemble composé d'un sentiment et d'un objet vers lequel est tourné ce sentiment. L'agrégation et la synthèse des corpus permettent alors d'estimer les opinions des internautes envers un événement, un objet ou une personne. Des sites de réseaux sociaux comme Twitter sont particulièrement exploités pour réaliser des analyses de sentiment envers un mot-clé (*hashtag* en anglais).

Selon Erik Cambria, nous pouvons diviser l'analyse du sentiment en plusieurs tâches (Cambria *et al.*, 2017) :

- La reconnaissance des émotions présentes dans le texte : il peut alors s'agir d'extraire des labels émotionnels ou plus généralement de réaliser des classifications binaires, positives versus négatives. On parle alors de polarité pour définir la valence mesurée dans un texte.
- Le degré de polarité : une granularité plus précise peut être calculée sur un axe de polarité. En effet, une phrase ou paragraphe peut contenir à la fois des termes négatifs ou positifs. Calculer un degré de polarité permet donc de faire une moyenne de tous ces éléments.
- La détection de l'objet du sentiment : un autre aspect du contexte est la prise en compte du sujet vers lequel est tourné le sentiment. Un corpus de texte peut contenir plusieurs objets, avec des sentiments positifs ou négatifs. De même, certaines phrases peuvent être hors sujet, par rapport au sujet principal. Identifier l'objet est donc un élément important pour donner un contexte aux sentiments identifiés.

On peut encore répertorier d'autres tâches applicables à un corpus de texte comme l'extraction de l'objet du sentiment, la détection des noms propres ou la détection de personnalité... Au vu de la quantité des données disponibles, il est nécessaire de mettre en place des méthodes automatiques pour réaliser les tâches. Nous pouvons lister trois approches d'analyse des sentiments (Cambria *et al.*, 2017 ; Medhat *et al.*, 2014) :

- L'approche lexicale : cette approche consiste à extraire des termes présentant une signification émotionnelle non ambiguë, à partir d'un dictionnaire préétabli. L'inconvénient de cette approche est le manque de rapprochement des termes sélectionnés avec le contexte de la phrase, ce qui limite fortement la précision de la reconnaissance de sentiment. Cette démarche peut donc être améliorée avec un corpus de référence incluant des règles syntaxiques simples. Les adjectifs d'une phrase peuvent ainsi être évalués plus finement en tenant compte des adverbes ou des connecteurs, pour déterminer l'orientation du texte. Nous pouvons lister plusieurs bibliothèques en ligne utilisant cette approche pour réaliser de l'analyse de sentiment : text-blob qui possède notamment une version française Loria, 2018, SentiWordNet (Esuli & Sebastiani, 2007) ou Word Net Affect (Strapparava, Valitutti *et al.*, 2004).
- L'approche statistique : des modèles d'apprentissage automatique sont entraînés avec des corpus de texte, annotés avec des catégories d'émotions. Les modèles vont alors apprendre à déterminer les mots-clés ayant une valeur affective et les fréquences de cooccurrence des termes, et à identifier des mots-clés souvent associés aux termes affectifs. En revanche, l'approche par apprentissage nécessite des corpus de texte importants pour être en mesure d'apprendre à reconnaître des patterns d'émotions significatifs. En effet, la prise en compte de la sémantique étant très faible, la quantité du texte doit être suffisamment importante pour identifier les cooccurrences et les fréquences des termes.
- L'approche hybride : cette approche implique l'utilisation combinée des deux démarches précédentes. Cela permet donc d'introduire des règles linguistiques ou des dictionnaires des termes annotés pour faciliter l'apprentissage des modèles de classification statistiques.

L'analyse automatique des textes est une tâche complexe. En effet, il existe des formats très diversifiés, comme de très courts avis sur un site de e-commerce, des articles de plusieurs pages ou des dialogues entre plusieurs personnes. Certaines situations posent de véritables défis aux chercheurs du domaine du TAL. Saif Mohammad (2017) recense plusieurs difficultés à surmonter telles que la compréhension des figures de style comme l'ironie, les métaphores ou la traduction entre différentes langues. Selon Erik Cambria, la discipline du TAL (Cambria *et al.*, 2017) a encore plusieurs paliers à atteindre : le niveau syntaxique, le niveau sémantique et le niveau pragmatique. Beaucoup des systèmes actuels restent encore limités au niveau syntaxique, avec une granularité limitée à l'analyse de la signification des mots. Les prochains paliers

consistent donc à dépasser cette granularité pour prendre en compte des concepts avec le niveau sémantique, voire des récits avec le niveau pragmatique.

III.3.5 Modalité physiologique

Nous avons déjà évoqué dans la sous-section II.3.4, le fait que les épisodes émotionnels sont à l'origine de nombreuses manifestations physiologiques. Ces manifestations sont des indices externes mesurables avec des capteurs appropriés. Ces mesures constituent des données complexes, qui, à l'aide de méthodes d'apprentissage automatique, permettent de créer des systèmes de classification automatique des émotions. Nous pouvons lister les signaux les plus couramment utilisés dans la recherche expérimentale dans ce domaine (Jerritta *et al.*, 2011 ; Shu *et al.*, 2018) :

- Au niveau du système cardiovasculaire : la variabilité de la fréquence cardiaque (VFC) et l'intervalle interbattement cardiaque (IIC) peuvent être mesurés à partir d'électrocardiogramme (ECG) ou de photopléthysmographe (PPM) (voir figure III.3). De nombreux descripteurs fréquentiels, temporels et statistiques peuvent être calculés à partir d'un extrait d'enregistrement de VCF (Egger *et al.*, 2019), couramment utilisé dans les études d'informatique affective. À partir de ce signal, Han-Wen Guo *et al.* (2016) sont parvenus à classer avec une exactitude de 71,6% des états de valence positive ou négative chez des participants ayant regardé des extraits vidéos.
- Au niveau de l'activité électrodermale : la conductance électrodermale de la peau (CE) ou la réponse galvanique de la peau (RGP). Des descripteurs statistiques calculés à partir de la RGP ont permis à Wen *et al.* (2014) de classer avec une exactitude de 74% cinq états émotionnels induits chez des participants par l'intermédiaire de vidéos.
- Au niveau de l'activité cérébrale : récemment, les signaux des électroencéphalogrammes (EEG) sont particulièrement exploités dans la recherche en informatique affective. Cependant, la mesure de ces signaux nécessite la mise en place de nombreux capteurs sur le cuir chevelu, ce qui ne facilite pas l'utilisation hors laboratoire (voir figure III.3). Pour simplifier l'utilisation des EEG, Shiyi Wu *et al.* (2017) ont réalisé une classification d'état de valence à partir des signaux de deux uniques électrodes placées à l'avant du crâne. À partir des signaux de ces deux électrodes, ils ont été en mesure de classer des niveaux de valence avec une exactitude de 75,18%.

En plus des signaux cités précédemment, on peut noter la pertinence de la pupillométrie, à savoir la mesure de la taille des pupilles. Il s'agit d'une pratique récente en informatique affective. Mesurer la taille de la pupille est possible avec l'utilisation d'*eye trackers* ou oculomètres en français. Ces dispositifs impliquent généralement

FIGURE III.3 – Appareils de mesures physiologiques



que les participants restent statiques, ce qui peut convenir à certaines situations d'interaction homme-machine. Les auteurs d'une étude ont classé trois catégories de valence (négative, neutre et positive) avec une exactitude moyenne de 71,7%. Pour cela ils ont utilisé des données sur la taille des pupilles, les mouvements de saccades et les clignements des yeux (Aracena *et al.*, 2015). Plus récemment, des oculomètres sous forme de lunettes ont été développés, qui permettent de réaliser des mesures dans des contextes plus variés (utilisation de smartphones, activités sportives, déplacements dans un magasin, voir figure III.4).



FIGURE III.4 – Lunettes d'occulométrie Tobii Glasses 2 développées par l'entreprise Tobii (Tobii, 2021), licence creative commons CC BY 3.0

À l'image des traitements des modalités visuelles ou audio, il existe deux méthodes principales pour réaliser des modèles de classification automatique des états émotionnels. Premièrement, l'approche par apprentissage automatique classique, qui nécessite une phase d'extraction des descripteurs en fonction des signaux étudiés, pour entraîner des modèles de classification (régression logistique, arbre de décision, machine à vecteur de support). Deuxièmement, l'approche par apprentissage profond, qui, à

l'aide de réseaux de neurones extrait automatiquement des descripteurs pertinents et apprend à différencier les enregistrements. De la même manière que pour l'étude de la voix, les données sont au format temporel, des réseaux de neurones récurrents sont donc particulièrement appropriés.

Les mesures physiologiques sont pendant longtemps restées réservées à l'utilisation en laboratoire avec des outils de mesures souvent encombrants. La miniaturisation des capteurs a permis de réaliser des outils de mesure beaucoup plus accessibles et moins invasifs. Nous pouvons noter l'existence de nombreuses montres dédiées au suivi de l'activité sportive, qui enregistrent les mouvements des utilisateurs, ainsi que des données physiologiques comme le rythme cardiaque avec des PPM. Des montres, disponibles au grand public, ont ainsi été utilisées pour réaliser des expériences avec induction d'émotions (Harper & Southern, 2019). Plus spécifiquement, la montre Empatica E4 a été conçue pour la recherche. Elle inclut, en plus d'un capteur pour la VCF, un accéléromètre, un capteur de température de la peau et un capteur de CE (voir figure III.5). Elle a été, entre autres, utilisée pour recueillir des données physiologiques pendant des épisodes émotionnels induits, afin de réaliser des outils de reconnaissance automatique d'émotions (Zhao *et al.*, 2018). Bien que moins précise qu'un holter cardiaque, elle s'avère capable d'enregistrer des données comparables 85% du temps testé, tout en étant moins invasive (McCarthy *et al.*, 2016).

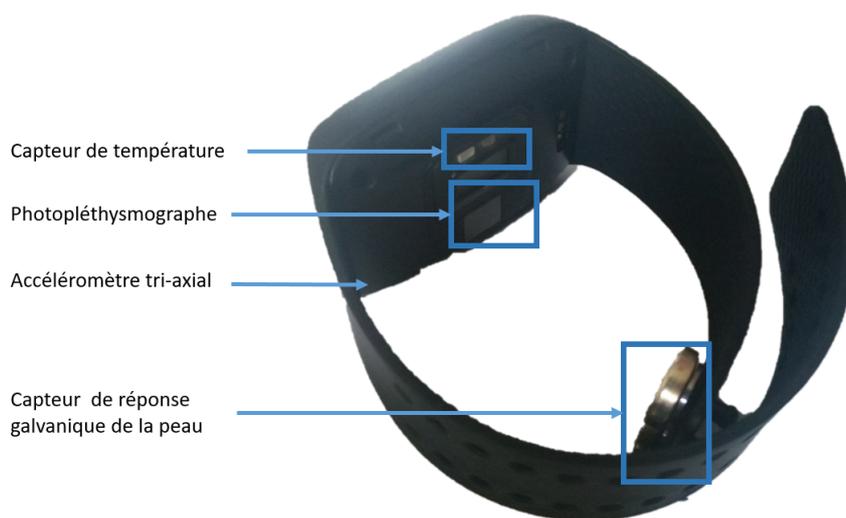


FIGURE III.5 – Montre E4 développée par l'entreprise Empatica (Empatica, 2021)

Des bases de données ont été réalisées à partir de plusieurs mesures de capteurs physiologiques. Nous pouvons citer la base MPED labellisée avec six émotions basiques plus un état neutre. Elle inclut des enregistrements de signaux EEG, de la CE, de la respiration et de la VFC (Zheng *et al.*, 2019). De la même manière, Philip Schmidt *et al.* (2018) ont créé la base de données WESAD, en induisant des états affectifs de stress,

d'amusement ou de neutralité avec des activités spécifiques (activité de présentation en public, de visionnage de vidéos humoristiques et de repos). Les mouvements, la respiration, la CP, la température de la peau et la VCF ont été enregistrés. Des bases de données avec des annotations continues ont été constituées. C'est le cas notamment des travaux de Sharma *et al.* (2018) qui ont enregistré la température de la peau, la VCF, la CP, la fréquence respiratoire et des signaux EMG des muscles du visage. Ces signaux ont été mis en correspondance avec des annotations sur des dimensions de valence et d'activation.

Selon Lin Shu, les études sur la reconnaissance d'émotions à partir des signaux physiologiques ont suscité un grand intérêt et de nombreux travaux (Shu *et al.*, 2018). Cependant, elle identifie plusieurs difficultés et limites à ces recherches. En effet, la majorité des études sont réalisées dans un contexte maîtrisé en laboratoire. Cela est nécessaire pour obtenir des données exploitables, mais pose des problèmes quant à l'utilisation de ces systèmes dans des contextes écologiques. De plus, il existe une grande variabilité interindividuelle dans la manière de réagir aux émotions, il est donc complexe de sélectionner des descripteurs et de concevoir des systèmes de reconnaissance automatique d'émotions généralisables. Cela est d'autant plus vrai, que les études de physiologie comprennent généralement un faible échantillon de sujets.

III.3.6 Approche multimodale

Ainsi, il existe de nombreuses méthodes pour obtenir des données. L'inclusion d'un maximum de mesures et de données semble propice à la réalisation de systèmes de reconnaissance automatique. Certaines recherches utilisent donc des approches multimodales pour tenter d'augmenter la précision des systèmes de reconnaissance automatique d'émotions. En revanche, l'approche multimodale implique une fusion des modalités pour être exploitable pour une prise de décision finale. Selon Soujanya Poria *et al.* (2017), il existe trois principales méthodes de fusion des modalités :

- La fusion au niveau des descripteurs : les données sont fusionnées en amont avant d'être utilisées par des classifieurs. La difficulté est de parvenir à synchroniser temporellement les données (par exemple une expression faciale avec une variation de mesures physiologiques) pour les faire correspondre avec un label ou une condition expérimentale. De plus, les différentes modalités doivent être adaptées dans un même format. Étant donné qu'un seul classifieur doit être développé, cette approche est plus économe en termes de ressources.
- La fusion au niveau décisionnel : chacune des modalités avec ses données a été classée indépendamment. Ainsi, il y a une décision sur l'état émotionnel détecté par modalité. Un système peut alors être mis en place pour statuer d'une décision finale sur l'état émotionnel identifié par les différentes modalités. L'avantage de cette méthode est que chaque modalité bénéficie du système de classification le

plus adapté à son format. En revanche, cette approche est plus consommatrice en ressources et en temps.

- La fusion multimodale hybride : elle implique un mélange des deux approches précédentes. Elle présente donc l'avantage de mettre en commun des classifieurs pour des données similaires. Martin Wöllmer *et al.* (2013) ont, par exemple, réalisé de l'analyse de sentiment à partir de vidéos. Ils ont exploité, d'une part les textes des vidéos avec un modèle SVM (machine à vecteurs de support) et d'autre part la vidéo et l'audio, fusionnés ensemble, avec de l'apprentissage profond. La décision des deux classifieurs est alors fusionnée pour fournir une estimation finale des sentiments exprimés dans les vidéos.

Certaines bases de données multimodales ont été conçues. Nous pouvons prendre l'exemple de la base de données SEMAINE, créée par McKeown *et al.* (2011), qui comprend des enregistrements vidéo et audio de conversations entre des sujets et des agents virtuels. Ces enregistrements sont annotés sur huit émotions. De même, à partir de vidéos postées sur le site *YouTube*, Morency *et al.* (2011) ont réalisé une base de données annotée sur trois états de valence (positive, neutre, négative) avec trois modalités de données (visuelle, audio et textuelle). Pour Poria *et al.* (2017), les approches multimodales ont tendance à donner de meilleurs résultats pour la reconnaissance d'émotions. La principale difficulté est de parvenir à fusionner des données différentes de manière efficiente, pour réaliser des systèmes utilisables en temps réel.

III.3.7 Difficultés en reconnaissance d'émotions

Les moyens disponibles pour la reconnaissance d'émotions sont variés. Chacune des modalités présentées constitue un champ de recherche à part entière. L'usage de modèles d'apprentissage automatique, et plus récemment d'apprentissage profond, est à l'origine de nombreux progrès pour identifier des descripteurs pertinents pour discriminer différents états. Il reste toutefois des défis à relever. R. A. Calvo et D'Mello (2010) ont mis en avant plusieurs problématiques sur le sujet de la reconnaissance automatique d'émotions :

- La différence entre l'expression et l'expérience : en effet, un sourire n'indique pas nécessairement la joie, tout comme froncer les sourcils n'implique pas forcément la colère. Les systèmes basés sur de la détection d'expressions faciales ne peuvent donc pas conclure sur l'état émotionnel d'un individu à partir de cette seule information. Néanmoins, cela fournit une information sur les réactions du participant.
- La cohérence des composantes d'un épisode émotionnel : selon les théories sur les émotions, les composantes évoluent de manière synchronisée lors d'un épisode émotionnel. Cependant, cela n'est pas forcément évident à observer sur le terrain, sauf pour des émotions intenses ou simulées. Des épisodes émotionnels moins intenses n'entraînent donc pas nécessairement des réponses synchronisées

et cohérentes des composantes. La conception de systèmes de reconnaissance multimodaux est donc confrontée à la complexité de traiter des mesures qui ne suivent pas la même évolution. La prise en compte de la synchronisation des données mesurées constitue en soi une perspective de recherche pour estimer l'intensité d'un état émotionnel.

- La faible quantité d'expressions d'émotions : il reste rare dans des situations écologiques que les individus expriment leurs émotions avec intensité. Des systèmes de reconnaissance entraînés à partir de mesures d'émotions fortes seraient donc limités dans un usage quotidien.
- Le manque de contexte : en effet, le contexte apporte des informations essentielles pour expliquer le déclenchement et la nature d'un épisode émotionnel. Ces informations sont compliquées à intégrer, d'autant plus pour des systèmes de reconnaissance ne prenant en compte qu'une seule modalité. Ainsi, détecter un sourire peut être très difficilement associé à de la joie, sans avoir l'élément à l'origine de ce sourire (il peut s'agir de gêne, d'un sourire de politesse...). De même, dans le cas de l'analyse des données textuelles, une remarque ironique ne pourra pas être perçue sans le contexte, ce qui cause de nombreuses erreurs de classification. Le traitement des informations contextuelles demande un niveau de cognition élevé. Cet aspect représente donc un défi important à relever pour la recherche.
- Le choix de classification catégorielle ou dimensionnelle : les deux possibilités ont été abordées parmi les études que nous avons présentées. D'une manière générale, le choix du type de représentation d'émotions est un débat de fond et non résolu en informatique affective. Il semblerait que certaines approches soient privilégiées pour certaines modalités, comme les études sur les modalités visuelles et plus particulièrement les expressions faciales, qui adoptent généralement des représentations catégorielles.
- L'évaluation des systèmes de reconnaissance d'émotions : la diversité des systèmes rend complexe la comparaison des performances des systèmes. Les types de données, le nombre de classes, le contexte d'utilisation constituent des paramètres qui spécialisent des modèles de classification. Ainsi, un système classant avec une bonne performance des données recueillies en laboratoire ne sera pas forcément aussi reproductible dans un contexte écologique.

Reconnaître des émotions est une tâche complexe pour les humains et les machines. En effet, nous tentons d'inférer les émotions d'autrui en cherchant des informations pertinentes dans le contexte ou dans les expressions visibles de la personne. Nous pouvons donc faire des erreurs, si nous n'avons pas toutes les données en notre possession. En ce sens, les systèmes automatiques sont encore très limités, car nous ne pouvons, actuellement, leur fournir que des informations restreintes à certaines modalités. La fusion de données variées est une perspective attirante pour la recherche, bien que présentant de nombreuses difficultés. En revanche, les systèmes automatiques

sont en mesure d'obtenir des données inaccessibles aux humains grâce à des capteurs innovants. Nous pouvons penser, par exemple, aux images des caméras infrarouges ou aux signaux EEG.

Dans le contexte du service client, certaines données peuvent s'adapter à certains canaux. Nous pouvons penser, notamment, aux analyses de la voix pour les canaux téléphoniques. La modalité textuelle est utilisée pour réaliser de l'analyse de sentiment sur les réseaux sociaux, les emails ou bien les messageries instantanées, particulièrement pour l'usage d'agents conversationnels. Nous pouvons aussi envisager l'usage des modalités visuelles en boutique, voire en visioconférence. Les modalités physiologiques sont beaucoup plus complexes à utiliser dans ce cadre-là. En revanche, elles peuvent constituer une source de données complémentaire dans le cadre d'expérimentations, afin de vérifier les émotions ressenties par les participants sur la composante psychophysologique.

Cependant, les modalités conventionnelles présentent des limites. Comme nous l'avons évoqué, les systèmes de reconnaissance basés sur la modalité textuelle ont plus de difficulté à reconnaître le sarcasme ou l'ironie. De plus, l'enregistrement de données fondées sur les expressions faciales, la voix ou des contenus textuels soulève des questions éthiques. En effet, ces informations sont susceptibles de porter atteinte à la vie privée des clients. Il peut donc être peu acceptable d'utiliser ces données dans le contexte de la relation client. C'est pourquoi nous considérons qu'il est essentiel d'explorer des modalités alternatives, moins intrusives et moins critiques pour la vie privée des utilisateurs. Les interactions avec des interfaces homme-machine nous semblent être une piste pertinente à explorer.

III.4 Reconnaissance automatique d'émotions et interfaces homme-machine

III.4.1 Introduction

La littérature en informatique affective porte majoritairement sur les modalités présentées dans la section précédente [III.3](#). Néanmoins, au vu de l'augmentation de l'usage des technologies numériques, ou technologies de l'information et de la communication (TIC), de nombreuses données peuvent être extraites à partir de nos interactions. Selon l'INSEE, en 2019 83% des Français ont utilisé un ordinateur, une tablette ou un smartphone pour accéder à internet contre 65% en 2009. Les TIC sont désormais des éléments incontournables de la vie quotidienne, que cela soit pour réaliser des activités professionnelles, communiquer, partager, consommer, se divertir... L'ensemble de ces interactions constitue une source de données supplémentaire pour accéder à nos

comportements et donc pour inférer nos états émotionnels. C'est pourquoi, la discipline de l'informatique affective s'intéresse à l'étude des interfaces homme-machine ou IHM. On peut aussi parler « d'interface utilisateur », définie par Michel Beaudouin-Lafon de la manière suivante :

« L'interface utilisateur contient les éléments logiciels et matériels dédiés à la capture des entrées de l'utilisateur, à la restitution des sorties du système et à la logique de l'interaction. » (Beaudouin-Lafon, 1997)

Ainsi, les périphériques d'entrée informatique comme les souris, les claviers et les écrans tactiles constituent des interfaces pour interagir avec des systèmes informatiques comme des smartphones, des ordinateurs ou des tablettes. L'étude de ces périphériques permet de recueillir des données sur l'interaction avec les utilisateurs. Afin de bien distinguer la différence entre ces dispositifs dont les frontières peuvent être floues, Lugtig et Toepoel (2016) proposent de les différencier en prenant comme paramètres l'interface et la taille d'écran, représentés sur le tableau III.1.

TABLE III.1 – Différenciation des dispositifs numériques les plus courants (Lugtig & Toepoel, 2016)

Dispositif	Interface	Taille d'écran
Ordinateur (portable et de bureau)	Souris et clavier	Supérieure à 6 pouces
Tablette	Écran tactile	Supérieure à 6 pouces
Smartphone	Écran tactile	Inférieure à 6 pouces

Selon une étude de Chritinia Bröhl *et al.* (2018), les ordinateurs sont privilégiés pour les activités suivantes : la rédaction de lettres, d'emails ou le partage d'informations confidentielles. Tandis que les smartphones sont préférés pour les usages suivants : la prise de note, la navigation web, l'utilisation de messageries instantanées, le partage de photos, l'usage des réseaux sociaux...

Dans cette sous-partie, nous présenterons les études en informatique affective portées sur les interactions avec des périphériques d'ordinateurs (souris et clavier) et sur les interactions avec des smartphones et leurs capteurs. Nous pouvons associer l'ensemble de ces modalités à la composante de l'expression motrice de la théorie de l'évaluation cognitive, car elle implique les mouvements et gestes des utilisateurs.

III.4.2 Ordinateur

Il existe deux principales sources pour la collecte de données sur ordinateur : la dynamique de frappe au clavier et les mouvements de souris (Kořakowska, 2013). L'enregistrement de ces données présente l'avantage de ne pas nécessiter de matériel particulier, des logiciels de capture des entrées peuvent être développés pour enregistrer les

interactions de l'utilisateur. De la même manière que pour les modalités plus classiques, comme les modalités visuelles (voir sous-section III.3.4) ou audio (voir sous-section III.3.3), une des difficultés est de pouvoir labelliser les données recueillies. Il existe deux approches principales, qui s'appliqueront aussi à l'étude des données de smartphone, pour faire correspondre les données enregistrées avec des labels émotionnels :

- L'expérimentation en laboratoire : les émotions sont généralement induites et les conditions expérimentales sont contrôlées.
- L'expérimentation de terrain : les participants utilisent des ordinateurs enregistrant leurs données et doivent auto-évaluer leurs émotions au cours de leur quotidien.

III.4.2.1 Dynamique de frappe au clavier

Clayton Epp propose de définir la dynamique de frappe de la manière suivante :

« Keystroke dynamics is the study of the unique timing patterns in an individual's typing, and typically includes extracting keystroke timing features such as the duration of a key press and the time elapsed between key presses. » (Epp *et al.*, 2011)

L'étude de la dynamique de frappe au clavier n'est pas récente. En effet, elle a déjà fait l'objet de recherches sur la reconnaissance automatique des utilisateurs, pour des utilisations dans le domaine de la sécurité (Joyce & Gupta, 1990). Les données brutes récoltées à partir du clavier sont constituées : de la touche pressée, de l'évènement d'appui et de relâchement de la touche, ainsi que le moment où ces évènements se sont produits (Epp *et al.*, 2011).

Deux types de descripteurs peuvent être calculés à partir de ces données (Kołakowska, 2013) :

- Des descripteurs temporels : à savoir la durée d'appui d'une touche ou entre plusieurs touches, la durée d'une séquence de plusieurs évènements et la vitesse de frappe globale au cours d'une session.
- Des descripteurs fréquentiels : ils correspondent au nombre de fois où les touches sont pressées sur une durée. Cela s'applique plus particulièrement pour des touches correspondant à des caractères spéciaux, comme la touche « effacer » ou des touches de ponctuation.

Preeti Khanna et Sasikumar (2010) ont réalisé une étude en laboratoire, dans laquelle les participants devaient recopier, par l'intermédiaire d'un clavier d'ordinateur, des textes avec des contenus émotionnels. Trois modalités de valence étaient proposées : négative, neutre ou positive. À partir de huit descripteurs extraits de la dynamique

de frappe au clavier, les auteurs sont parvenus avec de l'apprentissage automatique à une exactitude de classification de 89% pour distinguer les enregistrements négatifs des enregistrements neutres et de 88,88% pour distinguer les enregistrements positifs des enregistrements neutres. De plus, ils observent une vitesse de frappe moyenne plus faible pour les états neutres par rapport aux états positifs chez 83% des participants, pour 30 participants.

Clayton Epp *et al.* (2011) ont conduit une étude de terrain à l'aide d'un logiciel enregistrant les données de claviers sur les ordinateurs des participants. Le logiciel proposait régulièrement aux participants, lorsqu'ils interagissaient avec leur ordinateur, de s'auto-évaluer avec un questionnaire sur quinze états émotionnels et d'enregistrer les 10 minutes de données de clavier précédant cette évaluation. Ils parviennent à obtenir une exactitude de classification allant jusqu'à 87,8% pour distinguer certaines émotions par rapport aux autres.

III.4.2.2 Usage de la souris

L'utilisation des données issues des souris est moins courante dans la littérature. Néanmoins, il existe des études examinant le lien entre émotions et mouvements d'un curseur de souris. Selon Anna Kołakowska (2013), il existe deux types de descripteurs qui peuvent être extraits des interactions avec une souris d'ordinateur :

- Les descripteurs issus des mouvements appliqués au périphérique : la longueur du parcours, la direction, la vitesse ou l'accélération.
- Les descripteurs des actions : la fréquence de clics sur un bouton de la souris, l'utilisation de la molette, les actions de glisser-déposer ...

Hibbeln *et al.* (2017) font l'hypothèse que les mouvements de la souris sont moins précis lorsque l'utilisateur fait l'expérience d'une émotion négative. À partir d'une expérimentation avec des tâches induisant des émotions négatives ou neutres chez différents groupes de participants, ils observent que la présence d'émotions négatives diminue la vitesse d'utilisation de la souris et augmente les distances parcourues. À partir de ces données, ils conçoivent un système d'apprentissage automatique en mesure de classer les émotions négatives, des états neutres avec une exactitude de 81,7%.

III.4.3 Smartphone

Comme nous l'avons présenté, les smartphones, ou téléphones intelligents en Français, représentent une part de marché de plus en plus importante chez les utilisateurs de TIC, au point de dépasser les ventes d'ordinateurs personnels dans certaines régions

du monde. Ils permettent d'accomplir de nombreuses tâches : navigation sur internet, communication (emails, SMS, messagerie) ou applications de jeu... avec l'atout de la mobilité. Ces dispositifs représentent une concrétisation de l'informatique ubiquitaire, qui désigne un usage de l'informatique mobile, connectée et embarquée dans des objets de la vie quotidienne (Weiser, 1994). Depuis l'apparition des premiers smartphones dans le milieu des années 90, avec l'exemple de Simon développé par l'entreprise IBM, les smartphones ont considérablement évolué avec une augmentation de leur puissance de calcul, de leur mémoire et l'intégration de capteurs diversifiés (appareil photo, GPS, accéléromètre...). Le premier iPhone, développé par l'entreprise Apple en 2007, a marqué le début des téléphones intelligents modernes, avec une progression exceptionnelle de son usage jusqu'à nos jours (Politou *et al.*, 2017).

Anna Kołakowska *et al.* (2020) ont listé les capteurs des smartphones les plus courants (voir figure III.6) :

- Les caméras et appareils photo : ces capteurs sont présents sur l'avant et l'arrière du téléphone et permettent de filmer ou de photographier. La qualité de ces capteurs n'a pas cessé de s'améliorer au cours de ces dernières années pour fournir des photos et vidéos de haute qualité. La caméra à l'avant des smartphones est particulièrement adaptée pour des communications en visioconférence ou des prises de photos de visages (*selfie*). Ainsi, ces capteurs peuvent être rapprochés de la modalité visuelle pour la reconnaissance d'émotions, exposée dans la sous-section III.3.2.
- Les microphones : essentiels pour les activités de communication. Ils permettent d'enregistrer la voix ou le contexte sonore. Nous pouvons rapprocher ce capteur de la modalité audio de la reconnaissance d'émotions, voir sous-section III.3.3.
- Les écrans tactiles : très spécifiques aux smartphones et aux tablettes, les écrans tactiles ont fondamentalement modifié notre manière d'interagir, par rapport aux périphériques des ordinateurs de bureau. Certains gestes, généralement intuitifs, permettent de réaliser des actions pour naviguer et utiliser les fonctionnalités des applications : zoomer, balayer (*swipe* en anglais) ou dérouler (*scroll* en anglais). Les coordonnées des points de contact entre les doigts et l'écran peuvent être enregistrées. Ces enregistrements permettent donc de déduire des descripteurs sur les gestes effectués par les utilisateurs, comme des vitesses, des accélérations ou des directions. De plus, dans le cas de l'utilisation d'un clavier virtuel, la dynamique de frappe au clavier peut être recueillie à partir des enregistrements de l'écran tactile. La manière d'interagir avec un écran tactile peut donc être associée avec l'expression motrice. Dans le cadre de cette thèse, la prise en compte des informations de frappe au clavier sur un écran tactile est particulièrement intéressante. En effet, cela constitue une donnée exploitable dans le cadre de communications via des smartphones dans des applications de messagerie pour des contacts de service client. Ce point sera détaillé dans la sous-section III.4.3.1.
- Les capteurs inertiels : les accéléromètres, les gyromètres et les magnétomètres

sont des capteurs qui enregistrent les mouvements appliqués sur le téléphone et son orientation. Ces capteurs ont été initialement utilisés afin de positionner l'écran en mode vertical ou horizontal, en fonction de l'orientation du smartphone. De plus, ils peuvent servir pour des applications de jeux ou sportives. Ils sont notamment utilisés pour détecter l'activité physique des utilisateurs. Tout comme pour les données de l'écran tactile, les données issues des capteurs inertiels retranscrivent des gestes et des mouvements des utilisateurs et sont donc à associer à la composante de l'expression motrice. Nous détaillerons leur fonctionnement et leur utilisation pour l'informatique affective dans la sous-section III.4.3.2.

- Les capteurs de géolocalisation : la plupart des smartphones sont équipés de GPS (Global Positioning System) permettant de calculer la localisation de l'utilisateur à partir des données des satellites. Ces données permettent donc de suivre les parcours des utilisateurs, que ce soit lors de leurs trajets quotidiens ou leurs activités sportives. Il est donc envisageable d'utiliser ces données de localisation pour avoir des informations complémentaires sur le contexte dans lequel évolue un individu pendant un état émotionnel.

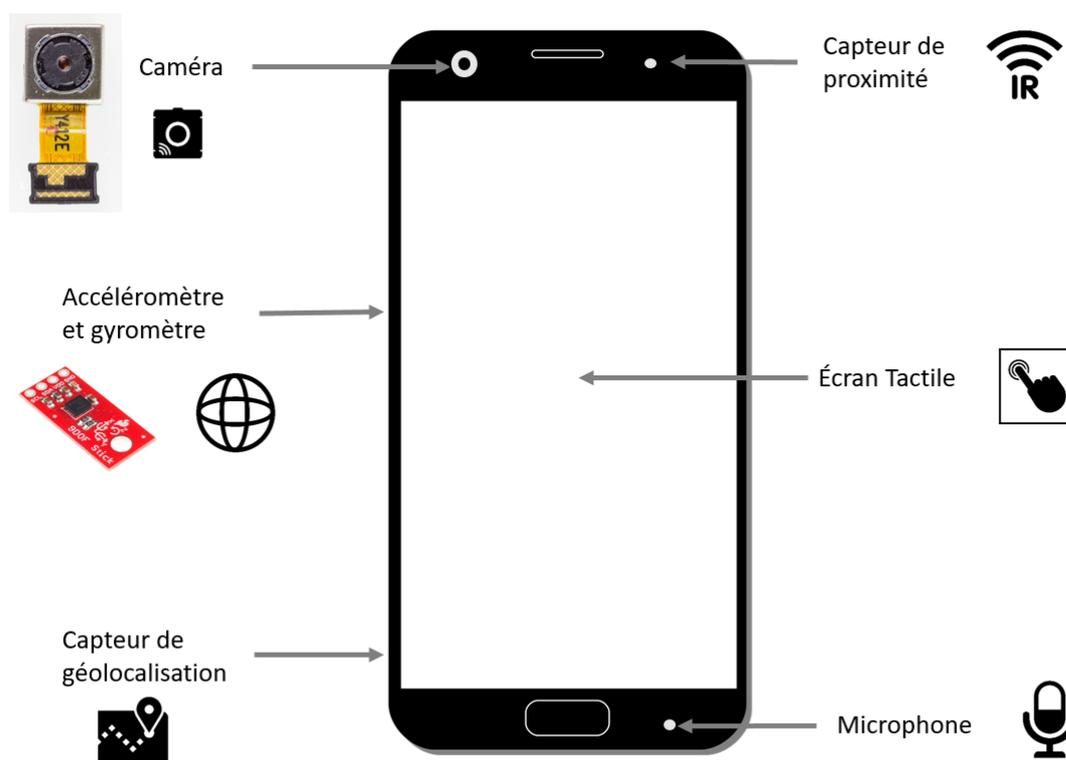


FIGURE III.6 – Smartphone et principaux capteurs associés. (Sparkfun, 2021) licence creative commons CC BY 2.0 / ©Raimond Spekking / CC BY-SA 4.0

De nombreux autres capteurs peuvent être intégrés aux smartphones selon le niveau

de gamme et l'âge des modèles. On peut, par exemple, noter l'ajout de capteurs de proximité pour évaluer la distance de l'utilisateur avec l'appareil, de capteurs de lumière ambiante pour ajuster le niveau de luminosité de l'écran ou de scanner d'empreintes digitales. On peut imaginer détourner l'usage de ces capteurs pour récupérer des informations sur le contexte d'utilisation.

La majorité des études de reconnaissance d'émotions à partir des données d'interaction avec des smartphones se concentrent sur les données d'écran tactile ou des capteurs inertiels. Dans le cadre de cet état de l'art nous présenterons particulièrement les modalités de mouvement et de frappe au clavier sur écran tactile, qui nous semblent être les données les plus à même de représenter le comportement d'un utilisateur dans le cadre d'un usage d'une application de messagerie sur smartphone. Tel que nous l'avons présenté pour les autres modalités, il est nécessaire de constituer des bases de données annotées sur des catégories ou dimensions émotionnelles. De la même manière que pour les études sur ordinateurs, nous pouvons considérer deux types de protocoles principaux : les expérimentations en laboratoire et les expérimentations sur le terrain. Les premières nécessitent généralement une induction d'émotions pour générer des états émotionnels chez les sujets de l'expérience. Tandis que la deuxième implique que l'utilisateur auto-évalue régulièrement ses émotions dans le cadre d'un usage quotidien du smartphone. Comparés à un ordinateur personnel, les smartphones, de par leur mobilité, facilitent la mise en place d'études de terrain. Le choix du type de protocole dépendra du contexte d'utilisation étudié.

III.4.3.1 Données des écrans tactiles et claviers virtuels

Les écrans tactiles permettent de rédiger des textes par l'intermédiaire de claviers virtuels. L'étude de la frappe au clavier sur smartphone est similaire à celle sur les claviers physiques d'ordinateur. En effet, il est possible d'extraire à partir de la dynamique de frappe au clavier, des descripteurs temporels et fréquentiels (voir la sous-section III.4.2). L'auteur Surjya Ghosh a réalisé de nombreuses études sur la dynamique de frappe au clavier et les états émotionnels (Ghosh *et al.*, 2017a, 2017b; Ghosh *et al.*, 2019). Il définit l'*inter-tap distance* comme étant la durée entre deux événements claviers consécutifs, cette distance est calculée à partir de l'horodatage (*timestamp* en anglais) de ces événements. À partir de cette distance, il est possible de calculer des descripteurs statistiques comme la moyenne, la médiane, le maximum ou le minimum pour des sessions d'utilisation du clavier. De plus, des descripteurs fréquentiels peuvent être extraits d'une session, comme la fréquence d'utilisation de la touche « effacer », qui met en évidence des corrections de la part de l'utilisateur ou l'utilisation de caractères spéciaux.

Avec son équipe, il réalise ainsi une étude de terrain avec une application développée spécifiquement pour enregistrer les événements claviers. Pendant trois semaines,

il recueille des sessions d'utilisation de clavier, qui se définissent comme une période pendant laquelle l'utilisateur reste sur une même application nécessitant un clavier virtuel. Un formulaire d'auto-évaluation d'émotions est alors proposé selon certaines conditions (minimum de texte produit, changement d'application et durée minimum de la session). Les participants peuvent alors choisir de s'annoter selon les quatre catégories suivantes : joie, relaxation, tristesse, stress. Les chercheurs ont alors conçu des modèles d'apprentissage automatique pour chaque participant. Ils obtiennent une exactitude moyenne de classification des quatre états déclarés de 73%.

Des gestes plus sophistiqués que la dynamique de frappe au clavier peuvent aussi être extraits et associés à des labels émotionnels. La pression appliquée, la vitesse et la durée des mouvements enregistrés sont autant de données fournissant des informations sur les gestes des participants. Balducci *et al.* (2020) ont réalisé une application de déverrouillage de smartphone par mouvement sur écran tactile. Ils ont proposé à des sujets de réaliser des séquences de gestes sur leur téléphone et de s'auto-évaluer sur trois états psychologiques : l'anxiété, le stress et la dépression. Les chercheurs ont mis en évidence l'importance de certains descripteurs, comme la variance des points du vecteur de mouvements par rapport à des lignes verticales et horizontales, avec les états déclarés.

III.4.3.2 Données des capteurs inertiels

Les capteurs inertiels sont composés de l'accéléromètre et du gyromètre, auxquels nous pouvons ajouter le magnétomètre. L'accéléromètre est un capteur de force. Par simplification, certaines explications présentent ce capteur comme enregistrant directement l'accélération linéaire appliquée à un appareil. En réalité, un accéléromètre enregistre les forces appliquées à l'appareil, une formule permet ensuite de calculer l'accélération enregistrée sur les axes : x, y et z, son unité est le m/s^2 . De fait, un accéléromètre enregistre notamment la force de la gravité, bien que celle-ci ne soit pas une accélération. Le gyromètre enregistre les rotations et les mouvements angulaires appliqués au téléphone. À l'image des capteurs sur un avion, il enregistre les rotations autour des trois axes de l'espace : le tangage autour de l'axe x, le roulis autour de l'axe y et le lacet autour de l'axe z. L'unité de mesure de la vitesse angulaire est le rad/s. Les axes enregistrés par les accéléromètres et gyromètres sont représentés sur la figure III.7. Le magnétomètre, bien que plus rarement utilisé dans la littérature en informatique affective, constitue aussi une source de données supplémentaire. Le magnétomètre est un compas électronique, qui à partir des mesures des champs magnétiques, peut déterminer la direction de l'appareil dans l'espace.

La plupart des smartphones sont désormais équipés d'un accéléromètre, bien que tous n'aient pas nécessairement un gyromètre et un magnétomètre. Ces capteurs ont été particulièrement utilisés pour réaliser de la reconnaissance de l'activité humaine,

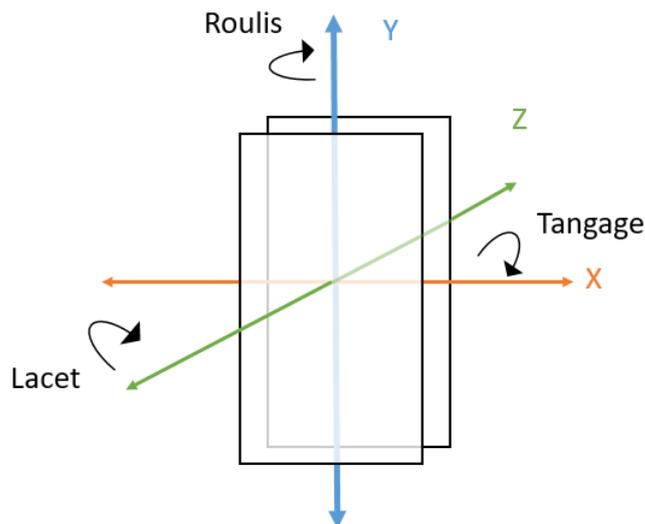


FIGURE III.7 – Représentation spatiale des axes d'accélération enregistrés par un accéléromètre et des axes de rotation enregistrés par un gyromètre

discipline consistant à déterminer les actions réalisées par une personne. En effet, des activités physiques comme courir, marcher ou monter un escalier produisent des mouvements discernables par les capteurs inertiels d'un smartphone situé sur le corps d'un utilisateur (Anguita *et al.*, 2013). Cela est particulièrement approprié pour des applications mobiles sportives ou de santé, qui peuvent ainsi à partir des ces informations, déterminer la durée et les intensités des activités physiques pratiquées par les utilisateurs.

Pour être exploitables, les signaux bruts enregistrés par les capteurs inertiels nécessitent un certain nombre de traitements (Kołakowska *et al.*, 2020) :

- Une réduction du bruit : cela peut être réalisé par l'intermédiaire d'un filtre moyennant sur une fenêtre glissante, afin de lisser les enregistrements extrêmes.
- De la même manière que pour les données temporelles issues des modalités audio et physiologiques, les données doivent être échantillonnées en segments de taille fixe. Afin de conserver un maximum de données, cela peut être effectué avec une fenêtre glissante.
- À partir des échantillons extraits, il est alors nécessaire de calculer des descripteurs pour résumer le signal et pour permettre d'entraîner des modèles d'apprentissage automatique. Les descripteurs peuvent se diviser en signaux temporels et en signaux fréquentiels. À noter, qu'il est aussi envisageable d'appliquer des modèles d'apprentissage profond directement sur les signaux filtrés pour une approche bout à bout.

Une explication plus détaillée du traitement des données et des descripteurs calculés à partir des capteurs inertiels sera présentée au chapitre V.

Dans la lignée de la reconnaissance de l'activité humaine, des chercheurs ont analysé les signaux des capteurs inertiels, en fonction des émotions induites chez des participants pratiquant une activité physique. Liqing Cui *et al.* (2015) ont montré à des participants des vidéos pour induire de la joie ou de la colère et leur ont demandé de marcher ensuite avec des smartphones attachés à leurs chevilles et leurs poignets. Après avoir filtré les signaux d'accéléromètre, les auteurs ont calculé 114 descripteurs et ont entraîné des modèles d'apprentissage automatique (SVM, arbres de décision, réseaux de neurones...). Les résultats ont montré que les smartphones attachés à la cheville fournissaient des signaux plus informatifs pour discriminer les états émotionnels. Ils ont réussi ainsi à obtenir une exactitude de classification de 90%. Muhammad Arslan Hashmi *et al.* (2020) ont réalisé une expérimentation en laboratoire similaire avec des smartphones attachés sur les torsos des participants. Cependant, ils ont induit six émotions différentes par mémoire autobiographique. Ils obtiennent une exactitude moyenne de classification de 86,45%. Enfin, Andreas Olsen et Torresen (2016) ont réalisé des travaux comparables, mais de manière non contrôlée. Les participants étaient incités à auto-évaluer leurs émotions lors d'activités de marche au quotidien, sur la dimension de la valence et de l'activation. Bien que le nombre de participants soit faible, des modèles d'apprentissage ont permis d'obtenir une exactitude de classification atteignant les 75%.

Des études ont cherché à analyser des gestes plus fins à partir des capteurs inertiels. Céline Coutrix et Mandran (2012) ont demandé à 12 participants de réaliser des gestes sur l'écran tactile de leur smartphone personnel, à partir d'une application dédiée. Ces gestes devaient être réalisés au moment où les participants ressentaient des émotions particulières ou à des moments aléatoires de la journée. À chaque usage de l'application les participants devaient auto-évaluer leurs émotions sur les dimensions de la valence, de l'activation et de la dominance. Pour chaque utilisation, l'application a recueilli les enregistrements de l'accéléromètre. L'extraction de descripteurs et la mise en correspondance avec les dimensions annotées a permis de mettre en évidence des corrélations significatives. Par exemple, la dimension de valence a été corrélée positivement avec des descripteurs de secousses sur l'axe z. Cependant, ce sont surtout les niveaux de la dimension de l'activation qui ont été les plus corrélés avec le signal d'accélération, avec 97 descripteurs corrélés significativement.

Des contextes plus particuliers comme des activités de jeux ont aussi été étudiés. Mottelson et Hornbæk (2016) ont ainsi créé une application de jeu basée sur l'écran tactile. En plus des données de mouvements enregistrées par l'écran tactile, ils ont enregistré les signaux de l'accéléromètre et du gyromètre. Avant la réalisation de chaque tâche, les participants devaient auto-évaluer leur état émotionnel avec un formulaire sur les dimensions de valence et d'activation. Les participants devaient réaliser des tâches impliquant des mouvements spécifiques de zoom, de tracé de ligne

ou pointage avec les doigts. L'apprentissage automatique réalisé avec des modèles de SVM a permis d'obtenir une exactitude de 81,7% pour distinguer une faible valence déclarée d'une forte valence.

Des activités plus spécifiques à l'usage du smartphone peuvent aussi être étudiées par l'intermédiaire de ces capteurs. Mintra Ruensuk *et al.* (2019) ont, ainsi, étudié les comportements des utilisateurs sur des tâches de navigation et de messagerie, au travers des capteurs inertiels (accéléromètre et gyroscope) et de l'écran tactile. Dans le cadre d'une expérience contrôlée en laboratoire, les sujets ont été incités à interagir avec des médias, dans un format similaire à des sites de réseaux sociaux et susceptible de provoquer des réactions émotionnelles. Les utilisateurs devaient ensuite annoter leur émotion sur des degrés de valence et d'activation et dialoguer avec un agent conversationnel à propos des médias visionnés. Les descripteurs extraits à partir des capteurs ont été classés en fonction de leur apport d'informations pour distinguer les niveaux des dimensions émotionnelles et en fonction de la tâche effectuée. Les mouvements de lacet et de roulis, rotations sur l'axe z et l'axe y, ont par exemple été particulièrement informatifs dans la tâche de navigation. Tandis que seul le mouvement de lacet était particulièrement informatif pour la tâche de conversation. Combinés avec les données des contacts sur l'écran tactile, des apprentissages automatiques permettent d'obtenir une exactitude de 81% pour distinguer des valences positives et négatives, lors de la tâche de navigation. Les résultats pour la tâche de conversation sont plus faibles avec une exactitude de 69%, montrant une plus grande difficulté à traiter les informations provenant de ce type d'activité.

III.4.4 Discussion

L'étude des données issues des IHM constitue une perspective sérieuse pour étudier les émotions des utilisateurs. L'attention de ce domaine se porte de plus en plus sur les dispositifs mobiles comme les smartphones, car ils intègrent un large panel de capteurs et peuvent être utilisés dans des contextes variés. Dans leur revue de littérature, Anna Kołakowska *et al.* (2020) estiment que depuis 2010, le nombre de publications portées sur de la reconnaissance d'émotions à partir des données de smartphones a augmenté. Cette démarche est d'autant plus pertinente que les smartphones sont amenés à s'imposer comme le moyen de télécommunication et de traitement informatique le plus utilisé.

Bien qu'étant une approche prometteuse, Euginia Politou *et al.* (2017) recensent un certain nombre de points durs pour la recherche en informatique affective centrée sur les usages de technologies. Premièrement, la vie privée des utilisateurs est particulièrement fragilisée en ce qui concerne l'usage de smartphones. Politou *et al.* (2017) catégorisent le niveau de criticité et de risques pour la vie privée des données recueillies sur des appareils mobiles, voir tableau III.2. Ainsi, de nombreux capteurs

III.4. RECONNAISSANCE AUTOMATIQUE D'ÉMOTIONS ET INTERFACES HOMME-MACHINE

peuvent collecter des données sensibles, telles que la localisation avec les données GPS, le visage avec la caméra ou la voix avec le microphone. À l'inverse, certaines données présentent moins de risques d'identification pour les utilisateurs, ce qui est notamment le cas des données des capteurs inertiels (accéléromètre et gyromètre) ou des interactions avec l'écran. Cela démontre donc l'importance d'explorer des données moins sensibles en parallèle des modalités plus classiques qui impliquent des données plus sensibles.

TABLE III.2 – Criticité des données recueillies sur smartphone (Politou *et al.*, 2017)

Niveau de criticité des données	Faible	Intermédiaire	Élevé
Données	Accéléromètre Gyromètre Magnétomètre Niveau de batterie Capteur de lumière Contact d'écran tactile	Usage d'application Usage de wifi Usage du bluetooth	Microphone Caméra Trafic internet Géolocalisation Appels SMS E-mails

Deuxièmement, les ressources de traitement des appareils, surtout portables, sont limitées. Or, l'enregistrement des données, le traitement et l'utilisation de données pour de l'intelligence artificielle impliquent l'utilisation de ressources conséquentes. Ces problématiques ne sont pas forcément bloquantes dans le cadre d'expérimentations scientifiques. En revanche, la mise en production de systèmes de reconnaissance représente une forte complexité pour des appareils limités comme des smartphones. De plus, l'usage de systèmes énergivores en ressources a des conséquences sur la durée de vie des batteries de ces dispositifs. Un travail conséquent doit donc être réalisé pour rendre ces technologies optimisées et acceptables dans des usages mobiles.

Dans le cadre de ce projet de recherche, les réalisations existantes dans ce domaine et particulièrement sur smartphone ont servi de base de réflexion pour concevoir une méthodologie appropriée. L'objectif est de construire des jeux de données représentatifs d'une activité de communication par messagerie sur smartphone. Dans cette optique, les données des capteurs inertiels et de la dynamique de frappe au clavier nous semblent prometteuses. En effet, elles présentent un risque faible pour la vie privée des clients par rapport à l'utilisation des modalités visuelles ou audio. De plus, la multimodalité étant préférable pour inférer l'émotion d'un utilisateur, ces informations pourraient s'avérer complémentaires aux analyses de sentiment par le texte. Ainsi, ces données pourraient être analysées afin d'aider les conseillers ou les agents conversationnels à s'adapter aux clients et à leurs émotions. Afin de réaliser des systèmes de reconnaissance d'émotions exploitant ces signaux, il est nécessaire de créer une base de données annotées selon des états émotionnels. Dans la prochaine partie, nous expliciterons les méthodes permettant de mesurer et labelliser des états

émotionnels dans un contexte expérimental.

III.5 Informatique affective et contexte expérimental

III.5.1 Protocoles expérimentaux

Afin de réaliser des outils de reconnaissance automatique d'émotions, il est nécessaire de rassembler des données labellisées. Deux alternatives s'ouvrent alors, soit il est possible de tester des outils de classification sur une base existante, soit de collecter de nouvelles données. Dans le cas des interactions homme-machine et plus particulièrement des capteurs des smartphones, il n'existe pas de base de données déjà réalisée et correspondant à une tâche de communication par messagerie. La collecte de nouvelles données est donc nécessaire. Pour cela, il existe plusieurs approches possibles, basées sur des expérimentations ou sur de la collecte de données issues de la réalité. Nous pouvons lister trois approches :

- Les études en laboratoire : il s'agit de la méthode permettant de recueillir des données avec la plus forte maîtrise des conditions expérimentales. De nombreux capteurs peuvent être déployés pour enregistrer des données. Il est cependant nécessaire d'établir un protocole permettant de provoquer des émotions chez les participants au cours de l'expérimentation. Cette approche a l'avantage de permettre de recueillir des données diversifiées sur un nombre assez large de participants. Afin de vérifier la bonne induction d'émotions, des auto-évaluations subjectives ou des mesures complémentaires peuvent être utilisées. La durée des expérimentations doit cependant rester restreinte. De plus, la validité écologique est limitée par le contexte de laboratoire et l'induction artificielle d'émotions. Cette approche est néanmoins l'une des plus utilisées par les chercheurs, car elle est l'approche la plus pragmatique pour tester de nouveaux types de capteurs.
- Les études de terrain : cette approche consiste à enregistrer des données au cours du quotidien des participants. Des outils spécifiques, comme des applications sur smartphone ou ordinateur, peuvent être développés pour recueillir les informations désirées (Epp *et al.*, 2011). Afin d'annoter les données enregistrées, il est nécessaire d'utiliser des systèmes d'auto-évaluation d'émotions se déclenchant régulièrement ou après détection d'un contexte approprié. L'avantage de cette approche est de permettre d'enregistrer des émotions avec une plus forte validité écologique, car elles sont expérimentées dans des situations réelles. En revanche, il est plus complexe d'inclure des outils de mesure comme des capteurs physiologiques. De plus, pour obtenir suffisamment de données annotées, les études doivent être prévues sur un temps suffisamment long, pour que les participants

expérimentent des émotions. Au vu de la complexité de cette approche, le nombre de sujets est généralement faible.

- L'exploration de données (*data mining* en anglais) : il est envisageable de rassembler des bases de données par l'intermédiaire des données accessibles sur internet. Cette approche est particulièrement utilisée en analyse de sentiment par TAL, à partir de réseaux sociaux ou pour les expressions faciales avec les sites de partage de vidéos (Morency *et al.*, 2011 ; Wöllmer *et al.*, 2013). Cette approche permet de collecter de nombreuses données. Cependant, la difficulté est de parvenir à labelliser les données selon des catégories ou des dimensions. Cela peut être réalisé de manière manuelle avec des observateurs ou des outils de classification fiables. Cette approche a donc l'avantage de permettre de récupérer une large quantité de données, avec une forte validité écologique en étudiant des émotions spontanées. En revanche, il est complexe d'annoter des données comme des enregistrements de capteurs ou des enregistrements physiologiques sans contexte.

Le tableau III.3 synthétise les informations des trois approches de recueil de données.

TABLE III.3 – Type de protocole pour la collecte de données

Type de protocole	Étude de terrain	Étude en laboratoire	Exploration de données
Nombre de participants	Faible (généralement moins de 20)	Intermédiaire (entre 20 et 100 participants)	Élevée (jusqu'à plusieurs milliers)
Quantité de données par participant	Élevée	Intermédiaire	Faible
Durée	Élevée (plusieurs jours) plusieurs semaines)	Intermédiaire (quelques minutes à plusieurs heures)	Variable (exemple : un message sur un réseau social ou une vidéo)
Source des émotions	Spontanée	Induite	Spontanée
Contrôle des conditions	Faible	Fort	Faible

III.5.2 Induction d'émotions

L'induction d'émotions est une tâche particulièrement délicate, l'objectif étant de faire ressentir aux sujets une émotion réelle. Selon Elise Dan-Glauser (2008) on peut diviser les méthodes d'induction d'émotions en deux catégories : les inductions internes et les inductions externes. Les premières impliquent un processus mnésique chez les participants pour qu'ils se remémorent des épisodes émotionnels de leur vie. On appelle alors cette méthode le rappel autobiographique. La deuxième nécessite l'utilisation de stimuli pour provoquer des émotions chez les participants. Sandrine Gil (2009) et

Elise Dan-Glauser (2008) listent les modalités suivantes pour induire des émotions : les images, les vidéos, la musique et même les odeurs. L'utilisation de stimuli est une des approches les plus courantes. Néanmoins, nous pouvons ajouter à ces modalités perceptives, la réalisation d'activités particulières comme des jeux, des tests de logique ou des interactions avec d'autres individus pour provoquer des émotions.

L'utilisation de stimuli, comme des images ou des vidéos, nécessite l'utilisation de bases de données testées, annotées par un nombre significatif et diversifié de sujets. Ainsi, des scientifiques ont réalisé des corpus de ressources conçus pour induire des émotions déterminées. La variabilité interindividuelle étant toutefois forte en matière d'émotions, les médias annotés ne représentent que des moyennes des effets globaux sur de grands échantillons de population.

III.5.2.1 Images

Les images sont des médias très largement utilisés dans la littérature scientifique pour induire des émotions en laboratoire. *L'international Affective Picture System* (IAPS) est l'un des recueils d'images les plus utilisés (Lang *et al.*, 1997), il est composé d'environ 1000 images, annotées sur les dimensions de la valence, de l'activation et du plaisir sur une échelle de 1 à 9. De plus, il comporte des contenus très variés impliquant des humains, des animaux, de la nourriture... En complément, Joseph Mikels *et al.* (2005) ont reproduit l'annotation du IAPS sur des émotions catégorielles. Enfin, certaines images ont été testées sur un public francophone par une étude de Catherine Bungener *et al.* (2016). Codispoti *et al.* (2001) ont démontré que des images du IAPS provoquaient des réactions physiologiques, mesurées à partir de la fréquence cardiaque, de la conductance électrodermale et de l'activation musculaire.

D'autres banques d'images annotées, plus récentes, sont disponibles, telles que la base OASIS (Kurdi *et al.*, 2017), composée uniquement d'images disponibles en *open source*. Cette base est labellisée sur les dimensions de la valence et de l'activation.

III.5.2.2 Vidéos

L'utilisation de vidéos, qu'il s'agisse de séquences de films ou d'extraits de vidéos amateurs, est une pratique répandue pour induire des émotions chez des sujets. Selon plusieurs études, les vidéos seraient en mesure de provoquer des réactions physiologiques et subjectives significatives (Schaefer *et al.*, 2010). En utilisant des extraits vidéos, Frazier *et al.* (2004) ont mesuré des corrélations significatives entre la valence ressentie et l'intervalle inter-battement cardiaque. De même, ils enregistrent une corrélation en l'activation déclarée et les variations de conductance de la peau.

Plusieurs banques de vidéos ont été réalisées et testées sur des participants. Gross et Levenson (1995) ont assemblé des séquences de films de cinéma, les ont catégorisées selon huit émotions cibles et ont sélectionné les deux extraits les mieux annotés pour chaque émotion cible selon les annotations des participants. Une expérience similaire a été réalisée par Schaefer *et al.* (2010) sur une population francophone à partir de séquences de films français ou étrangers. Les auteurs ont ainsi rassemblé une base annotée sur les dimensions de la valence, l'activation et des émotions discrètes. Étant donné que ces bases sont réalisées à partir de films avec des droits d'auteur, l'usage des bases peut être restreint. C'est pourquoi, Baveye *et al.* (2013) ont constitué une base à partir de films non soumis aux droits d'auteur et sous une licence permissive. Cette banque est composée de 9800 extraits issus de 160 films, a été annotée sur les dimensions de la valence et de l'activation.

Cependant, l'utilisation de films issus du cinéma constituerait des stimuli éloignés de la réalité. Ainsi, il peut être souhaitable d'utiliser des vidéos issues de situations réelles et n'impliquant pas des personnages et situations fictives. Samson *et al.* (2016) ont construit une base à partir de vidéos amateurs accessibles sur des sites de partages sur internet. Ces sites de partage de vidéos facilitent l'accès à ce type de contenu. Environ 200 extraits ont ainsi été évalués sur les dimensions de la valence, de l'activation, ainsi que des émotions discrètes comme l'amusement, le dégoût, la colère... Enfin, les vidéos ont été classées dans quatre catégories de valence générale : positive, négative, mixte et neutre.

III.5.2.3 Manipulation de la difficulté

Une des limites de l'utilisation de stimuli externes comme des vidéos ou des images est la validité écologique de la situation. En effet, les participants ne sont pas directement impliqués et ne subissent pas les conséquences de la situation. Ainsi, certaines approches proposent d'impliquer les sujets dans des activités à réaliser. Afin de susciter des émotions, des auteurs ont proposé de manipuler les résultats obtenus par rapport aux résultats attendus.

Nummenmaa et Niemi (2004) nomment ce type de protocole : manipulation du succès et de l'échec. L'objectif est de créer des conditions expérimentales sur une tâche accentuant les risques d'échec ou de succès selon les conditions et par rapport à un objectif à atteindre. Ce type de protocole peut être réalisé pour de nombreuses activités : tests de logique, examens, activités sportives... Selon une méta-analyse de l'auteur, les succès sont à l'origine d'émotions positives comme la joie, la fierté, tandis que des échecs sont à l'origine d'émotions comme la honte ou l'hostilité. D'une manière générale, ce type de protocole est plus particulièrement adapté pour influencer sur la dimension de la valence.

Par exemple, Schuch *et al.* (2017) ont réalisé des expérimentations, auprès d'étudiants, à partir de tests logiques (la matrice progressive de Raven *et al.* (2003)). Pour provoquer des succès et des échecs, les auteurs ont créé deux conditions expérimentales :

- Une condition succès : avec 2 items faciles du test et 10 items difficiles.
- Une condition d'échec : avec 3 items faciles et 9 items difficiles.

Avant de résoudre les tests, les participants étaient entraînés avec des items faciles, ils étaient aussi informés qu'en moyenne les étudiants résolvaient 50% des items. À la fin du test, il était indiqué aux sujets s'ils avaient réussi ou non, à être au-dessus de la moyenne. Les résultats ont montré une différence significative d'émotions entre les deux conditions. Selon les auteurs de cette publication, ce type de protocole présente trois avantages. Premièrement, sa validité écologique par rapport à l'utilisation de stimuli, car les participants sont directement impliqués. Deuxièmement, ce type de test rend l'induction d'émotions moins évidente et évite donc que les participants pensent à cet aspect. Enfin, le protocole provoque une comparaison sociale avec le groupe d'appartenance, les étudiants dans ce cas, ce qui renforce les émotions ressenties.

La difficulté de la manipulation du succès et de l'échec repose sur la bonne adaptation et équilibrage des différentes conditions à la population. Des conditions bien trop difficiles ou faciles pourraient rendre les participants suspicieux sur l'honnêteté du test. De plus, le test doit correspondre aux caractéristiques de la population et susciter la motivation des participants en impliquant l'ego par exemple, ou la comparaison sociale.

III.5.2.4 Comparaison des méthodes

Toutes les méthodes d'induction d'émotions ne se valent pas. Ainsi, des auteurs ont cherché à comparer l'efficacité des différentes approches à leur disposition. Westermann *et al.* (1996) ont réalisé une méta-analyse sur 250 articles pour comparer les impacts de 11 procédures d'induction d'émotions. Leurs résultats montrent que d'une manière générale, les inductions de valence négative sont plus efficaces que les inductions de valence positive. De plus, ils montrent que les films ou la présentation d'histoires sont les modalités les plus efficaces pour provoquer à la fois des émotions de valence positive et négative. Ferrer *et al.* (2015) réalisent une étude sur l'induction d'émotions à distance, par le biais d'Internet. Ils comparent les procédures d'induction par images et par vidéos de 26 études. De même que pour l'étude de Westermann, ils trouvent un effet significativement plus fort pour les inductions par vidéo. En outre, ils montrent aussi que l'induction d'émotions négatives comme la colère ou la tristesse est plus efficace d'une manière générale. Il faut cependant nuancer ces recherches qui se fondent majoritairement sur des études qui utilisent comme donnée principale l'auto-évaluation des participants. En effet, comme nous l'avons évoqué, le sentiment subjectif est complexe à interpréter pour les individus et n'est représentatif que d'une

part limitée de l'expérience vécue.

III.5.3 Mesures déclaratives

Afin de s'assurer de l'efficacité d'une induction d'émotions ou de permettre aux sujets d'auto-évaluer leur état émotionnel lors d'une étude de terrain, des outils de mesures déclaratives ont été réalisés par des chercheurs. Comme nous l'avons évoqué dans la sous-section II.3.7, le sentiment subjectif est difficile à exprimer. Bien qu'il ne reflète pas l'intégralité de l'épisode émotionnel vécu, le mesurer permet au moins d'estimer comment un stimulus ou une situation ont été perçus. De plus, il permet de réaliser des statistiques pour comparer l'effet de plusieurs stimuli. C'est pourquoi il est essentiel d'utiliser des outils adaptés et intuitifs pour faciliter la démarche d'auto-évaluation subjective des sujets d'une expérimentation.

Un des outils les plus courants dans la littérature est le PANAS, *Positive Affect and Negative Affect Schedule*. Développé par Watson *et al.* (1988), il se compose d'un questionnaire permettant d'évaluer 20 émotions sur une échelle de 1 à 5. Il permet d'évaluer globalement les émotions ressenties sur une période donnée : instant présent, journée, jours passés... À partir des différents items, il permet d'établir une mesure globale de l'état affectif positif ou négatif des participants.

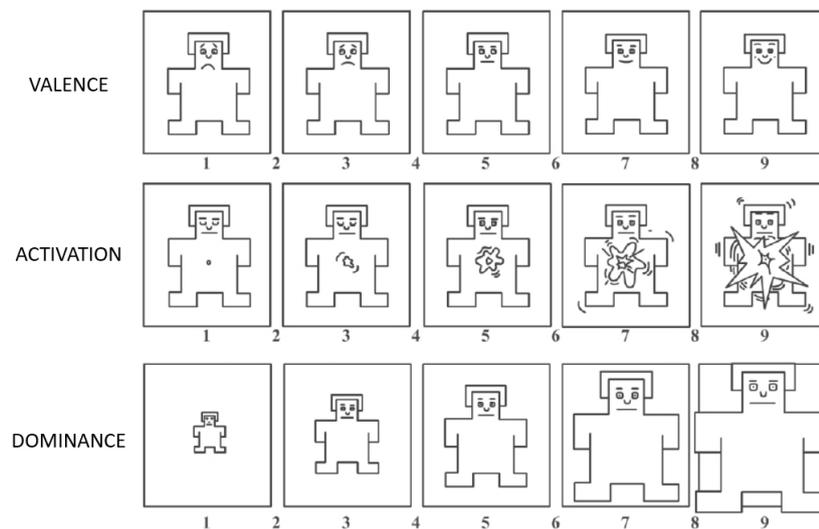


FIGURE III.8 – Représentation de l'échelle SAM (*Self Assessment Manikin*) (Bradley & Lang, 1994)

Dans une autre démarche, Bradley et Lang (1994) proposent l'outil SAM, *Self Assessment Manikin* pour évaluer ses émotions à partir d'images (voir figure III.8). Cet

outil offre la possibilité d'annoter ses émotions avec des échelles de 1 à 9 (ou de 1 à 5 selon les versions) sur trois dimensions : la valence, l'activation et la dominance. Selon Gil, 2009, l'utilisation d'échelles graphiques plutôt que des échelles textuelles présente l'avantage d'éviter certaines difficultés d'interprétation des termes émotionnels. De plus, les représentations graphiques sont plus faciles à utiliser avec des enfants ou à diffuser dans différentes cultures.

Sacharin *et al.* (2012) et Scherer *et al.* (2013) ont développé un outil prenant en compte à la fois des émotions discrètes et des dimensions émotionnelles : la *Geneva Emotion Wheel*, voir figure III.9. Cet outil a pour objectif de disposer sur un cercle les termes les plus représentatifs pour englober les grandes catégories d'expérience émotionnelle. Pour cela, les auteurs ont mis en œuvre une méthode, permettant à un échantillon de population d'annoter une liste de termes selon leur appartenance à une dimension et de sélectionner les plus représentatifs à partir des résultats. La taille des cercles correspond à l'intensité de l'état émotionnel ressenti. Les axes de l'espace représentent la dimension de la valence en horizontal et du contrôle (ou dominance) en vertical.

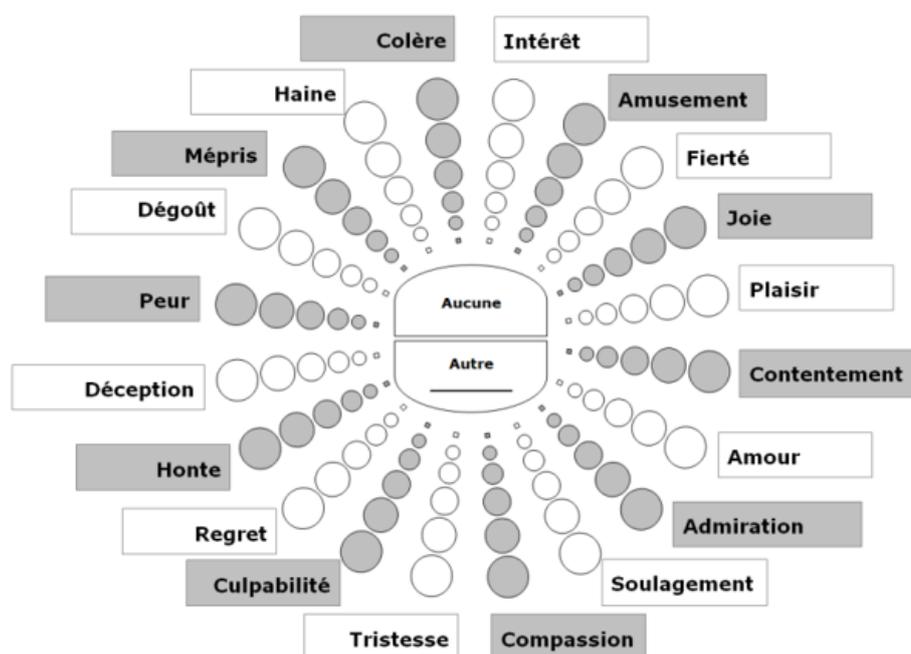


FIGURE III.9 – Représentation de la Geneva Emotion Wheel (Sacharin *et al.*, 2012)

III.5.4 Discussion

La collecte de données pour la création d'outils de reconnaissance d'émotions est une tâche complexe. D'une manière générale, le point dur est de parvenir à annoter correctement les émotions ressenties. Cela demande donc dans le cas des études en

laboratoire de créer des conditions expérimentales à même de générer des émotions attendues à l'aide de stimuli ou d'une activité appropriée. De plus, l'annotation subjective des sujets est nécessaire, soit pour vérifier une induction d'émotion, soit pour annoter une émotion spontanée dans le cas d'une étude de terrain. Cependant, il faut garder à l'esprit que les méthodes d'induction et de mesures ne sont pas infaillibles et que les variations inter et intraindividuelles sur l'expression des émotions sont fortes. L'utilisation de méthodes d'induction, même éprouvée et globalement efficace, ne garantit pas le déclenchement d'un épisode émotionnel marqué (Gil, 2009). Il peut être pertinent d'ajouter des mesures complémentaires comme des mesures physiologiques ou comportementales (expressions faciales, voix) pour observer les expressions des émotions des participants sur plusieurs composantes.

III.6 Conclusion

Ce chapitre nous a permis d'introduire la discipline de l'informatique affective, qui considère la prise en compte des émotions comme essentielle lors de la conception de technologies, afin d'améliorer les interactions. En effet, les TIC sont désormais parties prenantes de nos usages quotidiens et sont à la fois source et vecteur d'émotions. Elles sont notamment largement intégrées dans le domaine de la relation client, chaque canal de communication pouvant conditionner les expressions d'émotions. Nous considérons qu'appliquer l'approche de l'informatique affective à ce contexte est pertinent. Plus particulièrement, nous voulons explorer la possibilité de pallier le manque de communication non verbale dans les messageries instantanées. Ainsi, la mise en place d'un système de reconnaissance automatique des états émotionnels pourrait être une première étape pour mettre en place des stratégies, afin de faciliter la communication et l'adaptation entre interlocuteurs.

Les technologies nous offrent des moyens innovants pour recueillir des données sur les états émotionnels et pour tenter de les mesurer. Ainsi, nous avons présenté les moyens les plus étudiés dans la littérature pour reconnaître automatiquement des émotions, à savoir les modalités visuelles, auditives, textuelles et physiologiques. De nombreuses solutions et bases de données existent déjà à l'heure actuelle. Cependant, bien que certaines solutions soient largement utilisées dans la littérature, il existe encore de nombreuses difficultés à soulever pour fiabiliser ces systèmes. Ainsi, nous avons listé des problématiques inhérentes à cette démarche, telle que la différence entre l'expression et l'expérience ou le manque de contexte. Fusionner les modalités de données peut constituer une perspective intéressante pour compenser les faiblesses de chaque modalité. Cependant, la question de la criticité des données pour la vie privée est aussi essentielle, ainsi les modalités visuelles ou auditives sont par exemple particulièrement critiques pour des raisons de vie privée.

Dans le cadre de cette thèse, il nous semble intéressant d'étudier de nouvelles sources de données, issues des IHM et plus particulièrement des interactions avec des smartphones. En effet, cette nouvelle modalité peut s'associer à des sources de données plus conventionnelles. Le cadre du service client nous pousse à étudier ces interactions dans un contexte d'utilisation d'applications de messagerie instantanée. Spécifiquement, les données issues des capteurs inertiels et des claviers virtuels permettent d'obtenir des informations pertinentes, tout en étant non invasives et peu critiques sur la vie privée des utilisateurs. Ces données étant encore peu exploitées et ce contexte peu étudié, il est nécessaire de constituer une base de données composée des enregistrements des interactions avec un smartphone et annotée sur des états émotionnels, pour mettre en place un système de reconnaissance automatique d'émotions par apprentissage automatique.

Plusieurs méthodes sont envisageables pour assembler un corpus de données. Nous avons explicité la possibilité de réaliser des études de terrain ou en laboratoire afin de collecter les données nécessaires. Pour ces dernières, l'induction d'émotion est une approche privilégiée pour provoquer des états émotionnels contrôlés. La revue des méthodes d'induction d'émotions montre que l'utilisation de médias comme des vidéos ou des images, ainsi que la réalisation de tâches spécifiques sont les approches les plus courantes en laboratoire. Enfin, les mesures subjectives constituent une des méthodes les plus accessibles pour, d'une part, vérifier l'efficacité d'une induction d'émotions et d'autre part pour annoter des données enregistrées selon des états émotionnels.

Dans le chapitre [IV](#), nous détaillerons le matériel et la méthode mis en œuvre pour recueillir les données nécessaires pour nos recherches.

Chapitre IV

Matériel et méthode

Sommaire

IV.1	Introduction et positionnements méthodologiques	113
IV.2	Conception des outils d'expérimentation	116
IV.2.1	Présentation des outils	116
IV.2.2	Fonctionnement général de l'application d'expérimentation . .	117
IV.2.3	Interface de l'application d'expérimentation	118
IV.2.4	Adaptation de la <i>Geneva Emotion Wheel</i> en format mobile . . .	120
IV.3	Pré-expérimentation	122
IV.3.1	Présentation et objectif	122
IV.3.2	Protocole expérimental	122
IV.3.3	Hypothèse générale	126
IV.3.4	Passation	127
IV.4	Expérimentation principale	127
IV.4.1	Présentation et objectif	127
IV.4.2	Protocole expérimental	128
IV.4.3	Hypothèses générales	133
IV.4.4	Passation	134
IV.5	Expérimentation complémentaire	136
IV.5.1	Présentation et objectif	136
IV.5.2	Protocole expérimental	136
IV.5.3	Passation	138
IV.5.4	Hypothèse générale	139
IV.6	Méthode de reconnaissance automatique d'émotions	140
IV.6.1	Introduction	140
IV.6.2	Démarche générale	141
IV.6.3	Extraction des signaux bruts	143
IV.6.4	Prétraitement des signaux	144

IV.6.5	Génération de descripteurs	146
IV.6.6	Mise en œuvre de l'apprentissage automatique	153
IV.6.7	Discussion	164
IV.7	Conclusion	166

IV.1 Introduction et positionnements méthodologiques

Le chapitre précédent nous a permis d'identifier les sources de données les plus appropriées pour recueillir des informations émotionnelles sur les utilisateurs sur smartphone, de manière non intrusive et non invasive. De plus, nous avons pu lister les approches pour labelliser ces données selon des états émotionnels, notamment par l'intermédiaire d'échelles de mesure subjective.

À partir de ces éléments, nous pouvons ainsi réaliser des choix méthodologiques. Premièrement, en matière de collecte de données, nous avons pu mettre en évidence la pertinence de la collecte de données via des capteurs du smartphone et plus particulièrement les enregistrements des capteurs inertiels et de la frappe clavier (voir sous-section III.4.3). En effet, ils présentent l'intérêt de représenter une partie de l'expression comportementale de l'utilisateur. De plus, leur fonctionnement est totalement intégré au smartphone et n'implique pas de dispositifs supplémentaires. Ils pourraient donc permettre d'enregistrer des interactions entre un utilisateur et l'appareil mobile de manière écologique. Ces enregistrements constituent donc notre principale source de données pour la constitution d'une base de données, afin de concevoir un système de reconnaissance automatique des émotions.

Deuxièmement, nous faisons le choix de réaliser des expérimentations en laboratoire. En effet, il s'agit de l'approche permettant le meilleur contrôle des conditions expérimentales (voir sous-section III.5.1). Cela facilite d'une part la labellisation des données et d'autre part la vérification du bon fonctionnement du protocole expérimental. Ainsi, nous cherchons à créer des conditions expérimentales permettant d'induire des états émotionnels chez les participants de l'étude. Les modèles de représentation d'émotions, présentés dans le chapitre II (voir II.4), constituent une base théorique pour diviser les conditions expérimentales en types d'états émotionnels à induire. Plus spécifiquement, nous divisons les conditions expérimentales à partir des dimensions de la valence et/ou de l'activation, ce qui permet d'obtenir des données labellisées sur ces axes. L'utilisation de médias comme des vidéos ou des images et l'implication des sujets dans des tâches constituent des moyens privilégiés pour induire des émotions de manière contrôlée et efficace. Nous utilisons des banques de vidéo annotées sur les dimensions de la valence et de l'activation afin d'induire des émotions. De plus, nous avons intégré des tests pour impliquer les participants.

Enfin, pour vérifier la robustesse d'un protocole, plusieurs modalités de données peuvent être prises en compte. Plus particulièrement, des auto-évaluations de la part des sujets de l'expérimentation constituent des mesures pertinentes pour estimer l'efficacité d'un protocole expérimental. Nous utilisons donc des outils de mesures

subjectives adaptés, déjà présents dans la littérature scientifique (voir sous-section III.5.3).

L'objectif général est de concevoir des modèles de classification, permettant de discriminer des dimensions émotionnelles, telles que des niveaux de valence ou d'activation, à partir des données d'interaction sur smartphone. Nous pouvons donc redéfinir plus précisément la problématique générale énoncée dans le chapitre I : **les signaux des capteurs inertiels et de la frappe au clavier de smartphone permettent-ils d'obtenir des modèles de classification robustes pour discriminer des dimensions émotionnelles**? Bien que de nombreuses bases de données existent en informatique affective sur de nombreuses modalités, il n'y a pas, à notre connaissance, de corpus rassemblant ces données. Une réflexion particulière doit donc se diriger vers la réalisation de protocoles expérimentaux robustes et généralisables, pour smartphone. Les questions fondamentales suivantes peuvent donc être formulées :

- **Question Q_Induction** : La mise en œuvre de méthodes d'induction d'émotions sur smartphone nous permet-elle d'obtenir des résultats mesurables à partir d'auto-évaluation ?
- **Question Q_Interindividu** : Les variations inter-individuelles des signaux enregistrés sont-elles assez fiables pour généraliser le processus de détection des émotions à un grand nombre de personnes ?
- **Question Q_Intraindividu** : Les variations intra-individuelles des signaux enregistrés sont-elles assez fiables pour généraliser le processus de détection des émotions au niveau d'une seule et même personne ?
- **Question Q_Equivalence_Physio** : Les signaux issus de capteurs physiologiques permettent-ils d'obtenir des robustesses de classification similaires aux modèles de classification issus des capteurs de smartphones ?

Dans un premier temps, nous devons obtenir des conditions expérimentales à même de provoquer des états émotionnels et recueillir des données pour vérifier et calibrer l'induction d'émotion. Ainsi, nous pouvons lister les objectifs suivants :

- La création d'outils expérimentaux pour enregistrer des données d'interaction sur smartphone (capteurs inertiels et frappe au clavier).
- La mise en place de protocoles pour induire des états émotionnels conditionnés et mesurables, pour labelliser les données précédentes.
- L'intégration ou le développement d'outils pour mesurer l'effet de l'induction d'émotions.

Une application d'expérimentation sur smartphone est développée, prenant le format d'une messagerie instantanée et intégrant un agent conversationnel. Ainsi, l'application reprend un format similaire à une situation d'interaction. Cependant, l'interaction avec l'agent conversationnel reste limitée. Ce dernier permet surtout d'adresser

automatiquement des questionnaires, de diffuser les médias et les tâches à réaliser, pour obtenir des conditions expérimentales contrôlées. Les instruments de mesure subjective ont été intégrés en format numérique. En outre, les mesures des capteurs du smartphone et de la frappe au clavier sont enregistrées, en parallèle de l'utilisation. Trois expérimentations ont été réalisées par l'intermédiaire de cette application :

- Une pré-expérimentation afin de tester la possibilité de faire passer des expériences sur une application mobile sur smartphone, dans un format de messagerie instantanée. Nous avons utilisé des extraits vidéos et avons proposé aux sujets une tâche de mémorisation et d'attention pour susciter leur engagement et induire des émotions.
- Une expérimentation principale, pour laquelle nous avons amélioré l'interface de l'application mobile. Des extraits vidéo ont aussi été utilisés pour induire des émotions. Les sujets devaient décrire avec le clavier virtuel les extraits vidéo visualisés. Les données des capteurs du smartphone ainsi que la frappe au clavier ont été enregistrées pour créer une base de données.
- Une expérimentation supplémentaire a été réalisée, afin de tester une autre forme de protocole expérimental pour susciter des émotions : la manipulation de la difficulté de tests de logique.

Finalement, la mise en place de modèles de reconnaissance automatique d'émotions permet d'évaluer la possibilité de discriminer différents états émotionnels. Pour cela un travail particulier a été réalisé pour concevoir un système de reconnaissance automatique, utilisant des enregistrements de l'expérimentation principale jusqu'à l'entraînement de modèles de classification. Ces étapes impliquent de traiter des signaux temporels, en fonction de leur nature, pour obtenir des ensembles de descripteurs adaptés pour résumer des morceaux de signaux. Nous décrirons ainsi l'ensemble des briques de code développées pour concevoir un système de reconnaissance automatique.

Dans ce chapitre, nous présenterons premièrement la conception des outils d'expérimentation. Cela comprend principalement l'application d'expérimentation, qui a fait l'objet de plusieurs variantes pour s'adapter aux besoins des différents protocoles. De plus, un travail de développement particulier a été réalisé pour adapter en format numérique l'instrument de mesure *Geneva Emotion Wheel* (Sacharin *et al.*, 2012; Scherer *et al.*, 2013). Deuxièmement, nous présenterons les protocoles expérimentaux et les spécificités des trois expérimentations. Enfin, nous détaillerons la méthodologie pour mettre en oeuvre un système de reconnaissance automatique d'émotions, à partir des données enregistrées dans l'expérimentation principale.

IV.2 Conception des outils d'expérimentation

IV.2.1 Présentation des outils

Pour répondre au mieux à nos besoins, nous avons développé une application pour smartphone. Cette application doit répondre à plusieurs exigences :

- Enregistrer les données des capteurs inertiels et de frappe au clavier, afin de collecter des informations sur l'expression motrice dans le cadre de l'interaction homme-machine avec le smartphone.
- Proposer une situation d'interaction, de type messagerie instantanée, pour se rapprocher d'une situation de service client.
- Diffuser des médias tels que des extraits vidéo, des images et du texte, afin d'induire des émotions.
- Administrer des questionnaires, pour recueillir des données sur les utilisateurs : âge, genre, habitudes d'usage ou plus spécifiquement sur l'impulsivité émotionnelle.
- Enregistrer l'ensemble des résultats sur le smartphone ou sur un serveur à distance.

Globalement, notre objectif est de concevoir une application flexible, permettant à la fois de réaliser des expérimentations en laboratoire, mais aussi à distance pour permettre une plus large diffusion de l'expérimentation, si nécessaire. Nous avons développé cette application par l'intermédiaire du *framework React-Native* (React-Native, 2021) permettant de réaliser des applications sur smartphone pour Android.

Afin de soumettre des questionnaires et le contenu des expérimentations, nous avons opté pour l'intégration d'un agent conversationnel (*chatbot* en anglais). De cette manière, les participants interagissent avec une application de messagerie instantanée et suivent un scénario guidé par un agent conversationnel. L'application a suivi plusieurs évolutions pour s'adapter aux besoins des expérimentations réalisées, c'est pourquoi plusieurs versions ont été développées. L'ensemble des versions développées proposent un fonctionnement similaire en terme de gestion des données et de déroulement global des scénarios d'expérimentation. Les différentes variantes se différencient par le contenu diffusé pour induire des émotions, les instruments d'auto-évaluation d'émotions utilisés et les activités que devaient réaliser les participants.

En plus de cette application, un développement particulier a été réalisé pour adapter la *Geneva Emotion Wheel* en format numérique. Cette adaptation a permis la réalisation à la fois d'une application indépendante et d'une intégration sur l'application d'expérimentation. Cette application sera présentée de manière plus détaillée sur la sous-section IV.2.4.

IV.2.2 Fonctionnement général de l'application d'expérimentation

La figure IV.1 représente le schéma fonctionnel et la gestion des données. Des scénarios sont stockés en format JSON, dans un serveur dédié (serveur A sur la figure IV.1). Un scénario correspond au contenu de l'expérimentation qui sera adressé au participant par l'intermédiaire de l'application mobile. Un scénario contient les éléments suivants :

- Des questionnaires : il s'agit de questions adressées à l'utilisateur par le *chatbot*. Les questions peuvent être fermées, pour des échelles de Likert ou des questionnaires sur les caractéristiques des participants (âge, genre...) ou ouvertes pour laisser les participants s'exprimer.
- Des activités : il s'agit du contenu principal de l'expérimentation. Il peut s'agir de l'intégration de médias : extraits vidéo, images ou extraits audio, liés à une action que le sujet doit réaliser, pour interagir. Chaque activité est associée à une émotion cible à induire qui peut être définie par un niveau de valence, d'activation ou de dominance, selon les conditions expérimentales.
- Des auto-évaluations : afin de mesurer le sentiment subjectif des participants, des auto-évaluations des émotions peuvent être incluses dans l'application. Elles peuvent être intégrées après chaque activité ou après un ensemble d'activités. Plusieurs types de méthodes d'annotation d'émotions peuvent être incluses comme la *Geneva Emotion Wheel* numérique ou des formats plus classiques de questionnaires, tels que le SAM (Bradley & Lang, 1994) ou le PANAS (Watson *et al.*, 1988).

Lorsque l'application est lancée, une demande est formulée, par l'intermédiaire d'une API hébergée sur le serveur A (voir figure IV.1), pour télécharger le scénario d'expérimentation. L'application démarre ensuite une session contenant le scénario d'expérimentation. Au cours de l'expérimentation, l'application enregistre les réponses de l'utilisateur aux questionnaires, aux tâches et aux auto-évaluations. En parallèle, les capteurs du smartphone sont enregistrés, l'application vérifie au préalable si le capteur est présent sur l'appareil. La bibliothèque *react-native-sensors* (React-Native-Sensors, 2021), permet d'enregistrer les signaux des capteurs suivants : accéléromètre, gyroscope et magnétomètre. De plus, les événements d'interaction sur le clavier virtuel sont enregistrés, si celui-ci est utilisé lors d'une tâche. Ces enregistrements sont horodatés et associés aux différents éléments d'une session, et plus particulièrement aux tâches, pour faire correspondre les signaux avec les situations émotionnelles.

À la fin de la session, les données sont envoyées au serveur B, par l'intermédiaire d'une API. Elles sont enregistrées dans une base de données MongoDB (mongoDB, 2021) et peuvent être extraites en fichier CSV (*Comma-Separated Values*). Les résultats peuvent aussi être enregistrés directement sur le téléphone, en format CSV, en cas de problème de connexion internet. Le fonctionnement global de l'application permet donc de proposer plusieurs types de scénarios et de s'adapter à des expérimentations

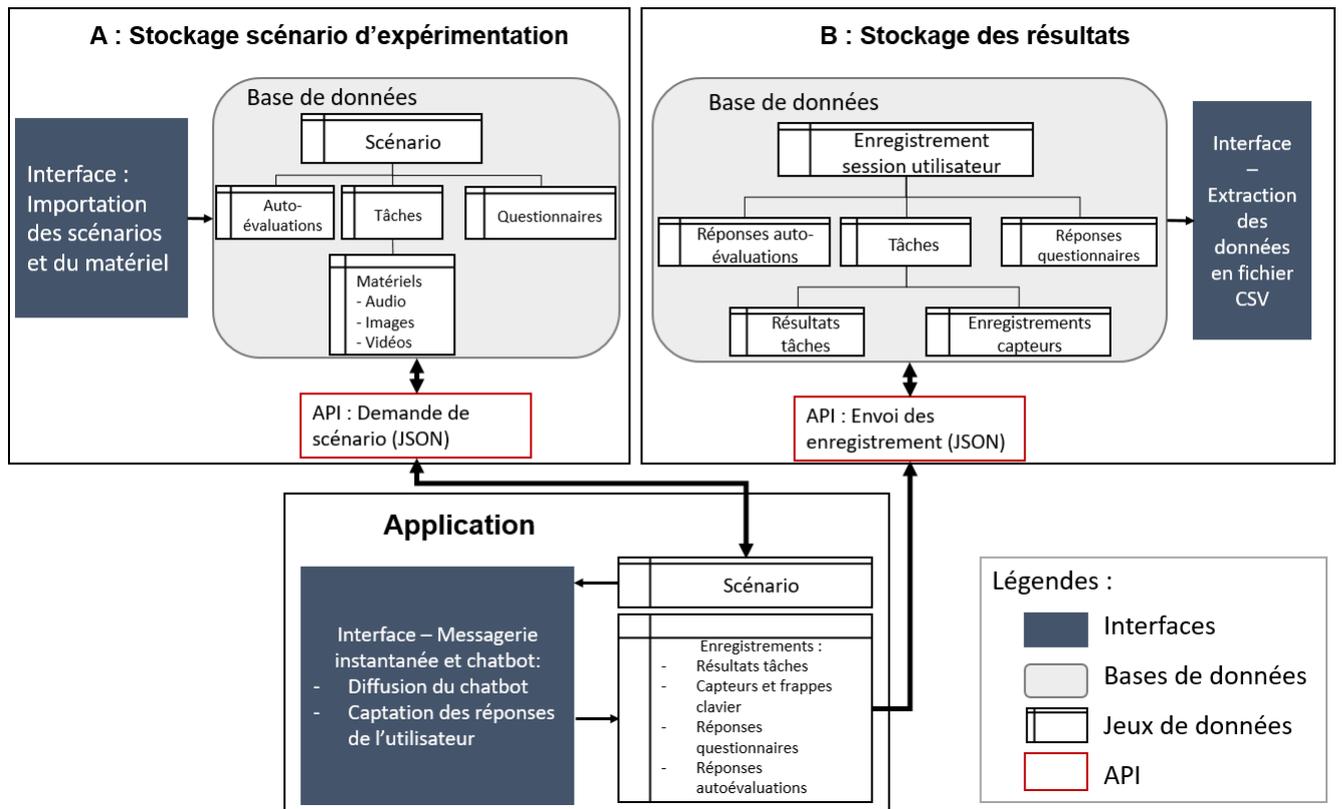


FIGURE IV.1 – Schéma fonctionnel

en laboratoire ou à distance.

Les serveurs utilisés pour les API et le stockage des scénarios sont hébergés au sein de l'entreprise Orange. Ils sont situés en France et doivent donc respecter la législation européenne sur le traitement des données : Règlement Général sur la Protection des Données (RGPD).

IV.2.3 Interface de l'application d'expérimentation

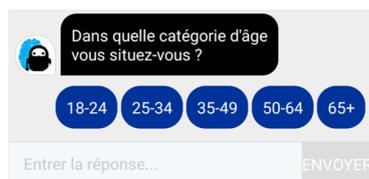
Au niveau de l'interface, l'application prend la forme d'une messagerie instantanée entre l'utilisateur et un chatbot. Les messages du chatbot sont chargés à partir du contenu du scénario téléchargé. Les messages peuvent contenir du texte ou afficher des médias. De plus, des interfaces spécifiques ont été développées pour certaines activités. Le chatbot enchaîne automatiquement les différentes étapes du scénario.

IV.2.3.1 Questionnaires

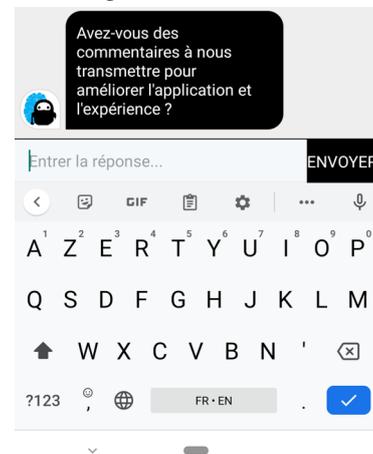
Le chatbot peut proposer des questions fermées, auxquelles les participants répondent en cliquant directement sur la réponse désirée. Pour les questions ouvertes, les utilisateurs ont accès au clavier virtuel pour rédiger leur message. Les formats des questionnaires proposés par l'agent conversationnel sont visibles sur la figure IV.2.

FIGURE IV.2 – Captures d'écran des interactions avec l'agent conversationnel par questionnaire

Question fermée avec les options



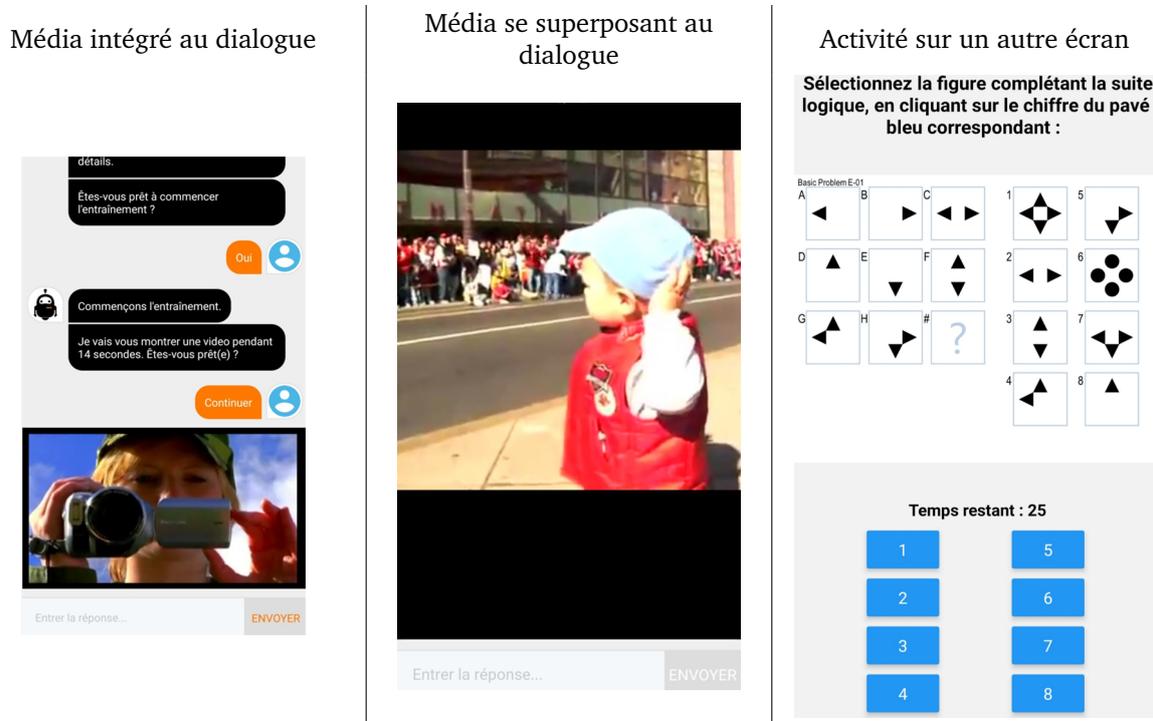
Question ouverte



IV.2.3.2 Tâches

Afin de créer des situations expérimentales, l'application peut diffuser des médias. De plus, l'agent conversationnel peut proposer des activités à réaliser, comme des questions auxquelles il faut répondre sur les médias diffusés, et où un temps maximum peut être imposé. Selon l'interaction nécessaire, le contenu peut être affiché dans une bulle de discussion de l'agent conversationnel ou dans un autre écran temporaire se superposant à la conversation du chatbot. Au cours de chacune des tâches, les réponses du sujet, les données des capteurs inertiels et si possible les frappes au clavier sont enregistrées. L'affichage d'exemples de médias et d'activités est représenté sur la figure IV.3.

FIGURE IV.3 – Captures d'écran des médias et activités pouvant être intégrés à l'application



IV.2.3.3 Autoévaluation

La mesure déclarative est l'un des moyens les plus utilisés pour vérifier l'efficacité d'une induction émotionnelle. À l'image des tâches, les auto-évaluations peuvent être intégrées au dialogue ou sur un autre écran.

IV.2.4 Adaptation de la *Geneva Emotion Wheel* en format mobile

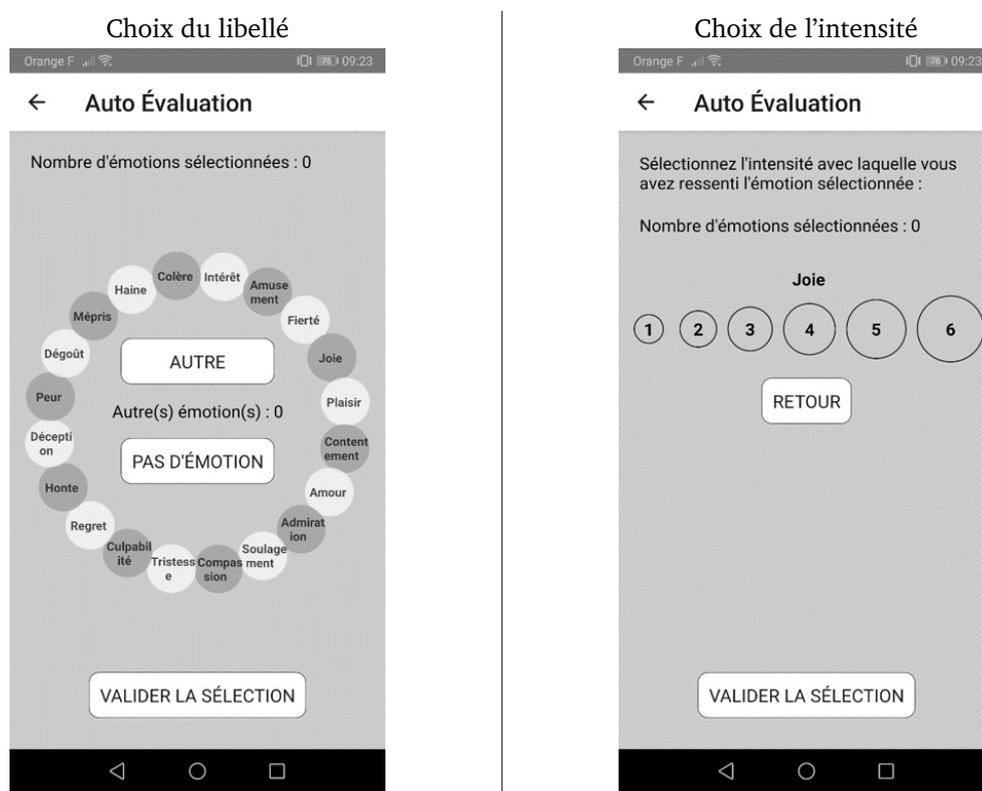
Pour les besoins de nos expérimentations, nous avons utilisé la *Geneva Emotion Wheel* (GEW) (Scherer *et al.*, 2013). Cependant, l'usage initial de la GEW est sur papier, les participants devant cocher avec un stylo les cercles du libellé émotionnel et de l'intensité désirée. Afin d'intégrer l'instrument de mesure à nos expérimentations, nous avons adapté la GEW pour un format mobile. Nous avons développé une application indépendante : la *Geneva Emotion Wheel Mobile Interface* (GEWMI), dont des éléments de code peuvent s'intégrer à l'application d'expérimentation. Cette application devait remplir les exigences suivantes :

- Afficher les instructions de l'utilisation de l'instrument.
- Permettre de sélectionner les libellés émotionnels et les niveaux d'intensité de manière tactile.

- S'adapter aux écrans de smartphones.
- Paramétrer le nombre minimum et maximum d'émotions sélectionnables.
- Permettre d'ajouter d'autres émotions (item « Autre »).
- Pouvoir modifier les libellés, afin de s'adapter aux différentes versions de roue (les créateurs de la GEW proposent une méthodologie pour créer des roues avec de nouveaux libellés pour s'adapter à des contextes différents).
- Enregistrer les résultats dans un fichier CSV sur le téléphone.

Le format d'un smartphone étant limité, nous avons divisé la sélection du libellé et de l'intensité en deux étapes distinctes. Les participants choisissent d'abord le libellé du sentiment qu'ils veulent déclarer, ensuite, ils détaillent le niveau d'intensité ressenti. La figure IV.4 représente les écrans de sélection des émotions. Les utilisateurs peuvent indiquer d'autres émotions en cliquant sur « Autre » ou directement indiquer s'ils n'ont pas ressenti d'émotions en cliquant sur « Pas d'émotion ».

FIGURE IV.4 – Écrans de sélection du label d'émotion et de l'intensité



Les captures des autres écrans de l'application GEWMI sont disponibles en annexes A. L'application en version indépendante est proposée en accès *open source* sur demande. Cette application a été conçue pour être utilisée dans de nombreux contextes

pour recueillir les émotions des sujets d'une étude. Le contenu de la roue et son aspect sont personnalisables.

IV.3 Pré-expérimentation

IV.3.1 Présentation et objectif

Réaliser des expérimentations sur smartphone représente un contexte scientifique particulier. Nous avons donc mis en place un premier protocole expérimental pour confronter l'application d'expérimentation à des participants avec une méthodologie d'induction d'émotions. L'ensemble de l'expérimentation est guidé par un agent conversationnel, ce qui nous permet aussi d'explorer cette modalité comme moyen pour adresser automatiquement des tests.

L'objectif de cette expérimentation est la vérification de l'induction d'émotions par l'intermédiaire d'une mesure d'auto-évaluation. Comme nous l'avons vu dans la sous-section III.5.2.4, les extraits vidéo permettent d'induire des émotions de manière efficace. C'est donc cette méthode que nous avons retenue pour ce premier protocole expérimental. La difficulté est toutefois que l'émotion induite atteigne une intensité suffisante. Pour renforcer l'engagement des participants dans la tâche, nous proposons de les solliciter dans une tâche d'attention et de mémoire pour chacune des vidéos visionnées.

Avec cette expérimentation, nous appliquons un premier protocole d'induction et vérifions ses effets par l'intermédiaire d'une mesure d'auto-évaluation de la composante du sentiment subjectif. En complément, cette pré-expérimentation nous a permis de tester l'ergonomie et le fonctionnement global de l'application, afin de réaliser des améliorations et des correctifs pour les expérimentations suivantes. Les données de l'accéléromètre, du gyroscope et de la frappe au clavier ont été enregistrées pour tester le bon fonctionnement de l'application et vérifier la mise en correspondance des émotions avec les enregistrements.

IV.3.2 Protocole expérimental

IV.3.2.1 Induction d'émotions et mesure subjective

Nous avons pris comme référence un modèle multidimensionnel : le circomplexe de Russell, présenté dans la sous-section II.4.4 est constitué des dimensions de la valence et de l'activation. Ce modèle nous a permis de créer des conditions expérimentales, correspondant à des zones de l'espace (voir figure IV.5). Un modèle dimensionnel permet

de prendre en compte des ensembles d'émotions plutôt que des émotions basiques spécifiques, offrant une plus grande flexibilité. En outre, ce choix a été guidé aussi par la simplicité du modèle, permettant d'éviter la création d'un nombre trop important de conditions expérimentales.

- Condition 1 : Valence et activation moyennes.
- Condition 2 : Valence élevée et activation élevée.
- Condition 3 : Valence élevée et activation faible.
- Condition 4 : Valence faible et activation élevée.
- Condition 5 : Valence faible et activation faible.

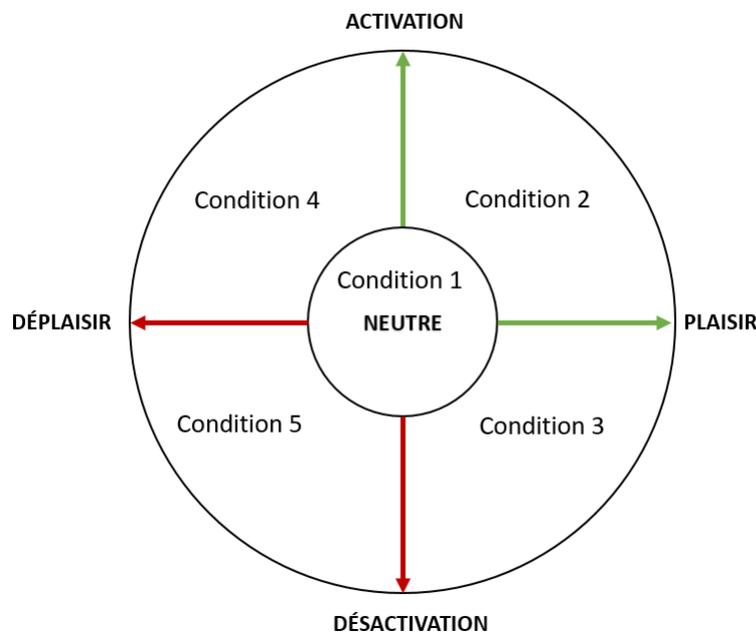


FIGURE IV.5 – Représentation des conditions expérimentales sur le circomplexe de Russell (Russell, 1980)

Pour correspondre aux conditions expérimentales, nous avons sélectionné une base de données d'extraits vidéo, annotés sur les dimensions de valence et d'activation. La base de données LIRIS-ACCEDE créée par Baveye *et al.* (2013) remplit cette condition. Elle contient des extraits de films en libre accès, annotés par un panel de participants sur la valence et l'activation avec une échelle de Likert de 1 à 5. La base est accessible sur le site du laboratoire LIRIS. Les extraits durent en moyenne 20 secondes. Nous avons sélectionné 10 vidéos de chaque catégorie avec l'objectif de trouver des vidéos

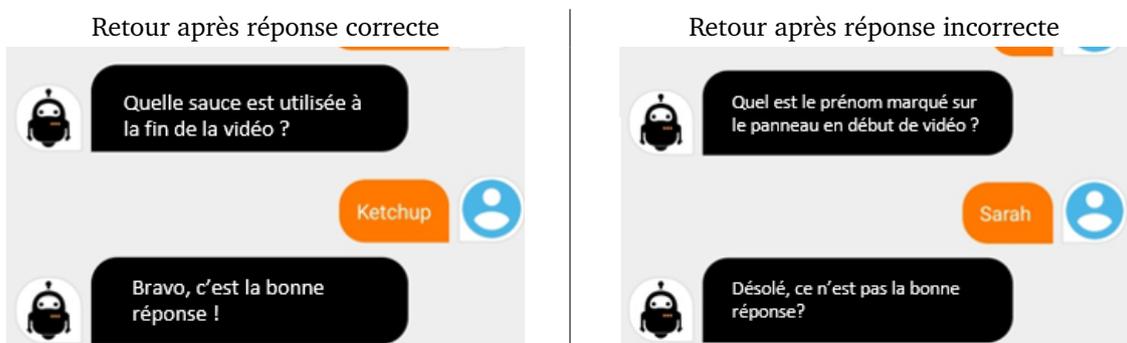
TABLE IV.1 – Moyenne et écart-type des annotations de valence et d’activation des vidéos sélectionnées (Baveye *et al.*, 2013)

Condition expérimentale	Valence moyenne (1 à 5)	Écart-type	Activation moyenne	Écart-type
Neutre	2,48	0,31	2,65	0,32
Valence Positive / Activation Positive	3,55	0,01	3,88	0,28
Valence Positive / Activation Négative	3,55	0,14	1,41	0,14
Valence Négative / Activation Positive	1,40	0,14	3,99	0,16
Valence Négative / Activation Négative	1,41	0,14	1,40	0,14

correspondant aux cinq conditions expérimentales (voir tableau IV.1). À savoir, des vidéos présentant des valeurs d’activation et de valence faibles, intermédiaires ou élevées.

Les extraits vidéo sont intégrés parmi les messages du chatbot. Pour chaque vidéo, une question est posée au participant sur un évènement ou un élément présent dans le média. La difficulté des questions varie en fonction du niveau de valence. Les vidéos avec une valence faible sont associées à des questions plus compliquées pour augmenter le risque d’échec et donc de valence négative. Inversement, les extraits vidéo avec une valence élevée sont associés à des questions plus simples pour augmenter les chances de réussite et donc les valences positives. Pour chaque réponse, l’agent conversationnel réagira en fonction de la réponse correcte ou incorrecte (voir la figure IV.6). Un score est enregistré tout au long de l’expérimentation pour engager les participants.

FIGURE IV.6 – Exemple de feedback de la part du chatbot



Afin de vérifier l'induction d'émotions, des auto-évaluations avec l'instrument *Self-Assement Manikin* (SAM), présenté en sous-section III.5.3. Les participants doivent évaluer leur sentiment subjectif sur les dimensions de la valence et de l'activation. Nous avons sélectionné cet outil, pour sa simplicité d'utilisation et d'intégration à notre application. Il permet aux participants de directement s'évaluer sur les conditions expérimentales que nous avons conçues. Une version du SAM avec une échelle sur 5 points a été intégrée (voir figure IV.7).

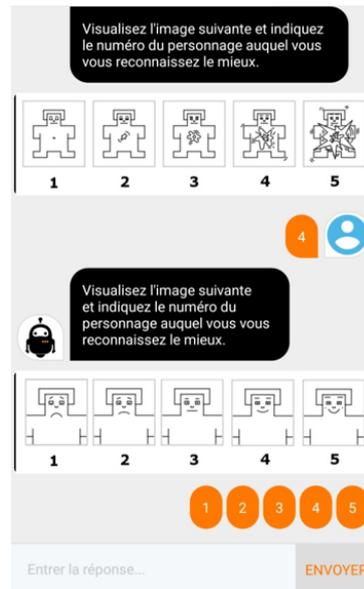


FIGURE IV.7 – Intégration de l'instrument SAM dans l'application d'expérimentation (Bradley & Lang, 1994)

IV.3.2.2 Plan expérimental

Au préalable de l'expérimentation, des instructions sont données aux participants par l'intermédiaire de l'agent conversationnel. Le *chatbot* explique aux participants en quoi consiste l'expérimentation. L'accent est porté sur la complétion des tests de mémoire et d'attention sur les vidéos, pour ne pas biaiser les participants sur l'aspect émotionnel. Les émotions ne sont pas évoquées pour ne pas influencer sur les réactions des sujets. Il est demandé aux participants de rester assis tout au long de l'expérience pour ne pas perturber les enregistrements des capteurs. Des questionnaires sont proposés par l'agent conversationnel pour enregistrer des informations générales sur la population de sujets :

- Questionnaire sur les caractéristiques personnelles : âge, genre, main dominante (voir annexe C.I).
- Questionnaire sur les habitudes d'usage des smartphones : temps d'usage quotidien de smartphone et d'applications de messagerie instantanée (voir C.II).

Suite aux instructions et questionnaires, un entraînement est proposé. Deux vidéos avec une valence et une activation neutres sont montrées aux participants, ainsi qu'une première auto-évaluation du sentiment subjectif avec le SAM. Une fois l'entraînement terminé, les sujets ont le choix de le recommencer autant de fois qu'ils le souhaitent ou de débiter le test. Les dix extraits vidéo de chaque condition expérimentale sont rassemblés en blocs. Les participants passent par chacun des cinq blocs, dans un ordre aléatoire pour compenser les effets de fatigue ou d'apprentissage. Au sein de chaque bloc, l'ordre des extraits vidéo est aussi aléatoire. Après chaque bloc, une auto-évaluation avec l'instrument SAM sur la valence et l'activation est demandée aux participants.

Un champ libre est proposé à la fin du test pour laisser les participants s'exprimer librement sur l'application et suggérer des améliorations possibles. La figure IV.8 résume l'ensemble de l'expérimentation. Celle-ci dure entre 40 minutes et 1 heure, selon les participants.

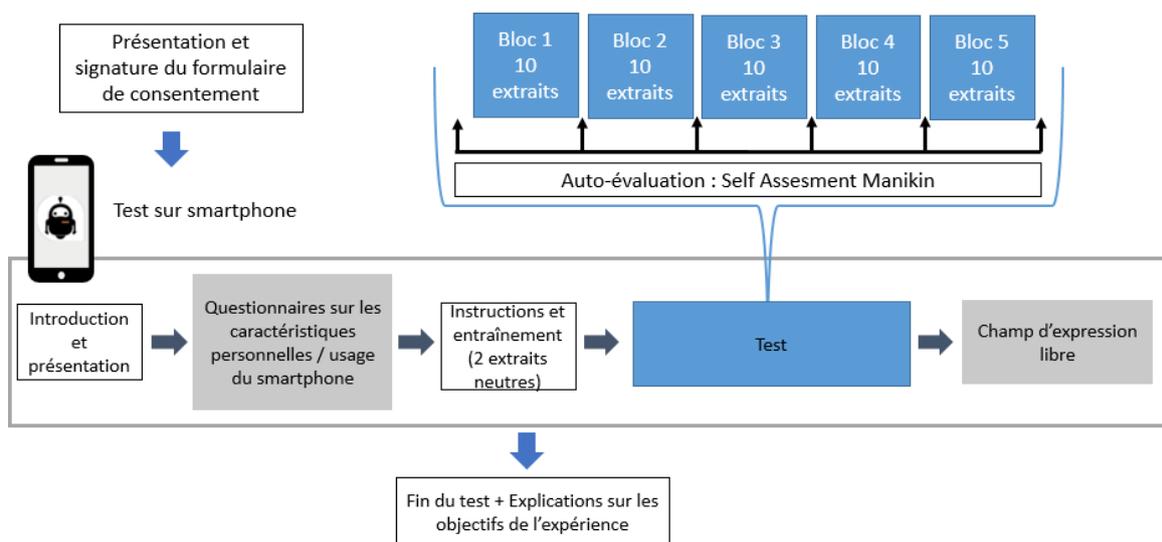


FIGURE IV.8 – Déroulement global de l'expérimentation

IV.3.3 Hypothèse générale

L'objectif de cette expérimentation est de déterminer si les participants ont été suffisamment engagés pour ressentir des émotions avec une valence positive, négative ou neutre et une activation faible, neutre ou élevée. Nous cherchons donc à donner des éléments de réponse à la question **Question Q_Induction**.

Ainsi, nous postulons l'hypothèse générale suivante : **pour l'ensemble des participants, les méthodes d'induction d'émotions de cette expérimentation permettent**

d’obtenir des effets sur la valence et l’activation, en correspondance avec les conditions expérimentales proposées. Ces effets sont mesurables par l’intermédiaire des auto-évaluations de la valence et de l’activation sur l’instrument SAM, permettant de mesurer quantitativement le sentiment subjectif des participants.

Les hypothèses opérationnelles qui découlent de cette hypothèse générale et les analyses sont disponibles dans la sous-section [V.2.2](#).

IV.3.4 Passation

La passation de cette expérimentation s’est déroulée en juillet 2019. Une déclaration au sein du groupe Orange a été formulée, pour le respect du Règlement Général sur la Protection des Données (RGPD) au préalable. Les participants ont tous passé le test sur le même smartphone : un Huawei P Smart. Une salle d’expérimentation a été spécifiquement prévue pour que les participants puissent passer le test dans les mêmes conditions (luminosité, bruits...). Au préalable de chaque test, les participants doivent lire et signer un formulaire de consentement (voir l’annexe [B](#)).

40 employés d’Orange Labs Services ont passé l’expérimentation, 24 hommes et 16 femmes. La répartition des âges est la suivante :

- 15 sujets de 18 à 35 ans
- 14 sujets de 35 à 49 ans
- 11 sujets de 50 à 64 ans

94% des sujets ont déclaré être droitiers. 50% des participants ont déclaré utiliser leur smartphone plus 30 minutes par jour.

IV.4 Expérimentation principale

IV.4.1 Présentation et objectif

Comme nous l’avons vu au chapitre [V](#), les nouvelles technologies de la communication permettent de capter des signaux susceptibles de fournir des informations sur les émotions induites. L’objectif de cette nouvelle expérimentation est ainsi de recueillir des données des signaux des capteurs des smartphones et de la frappe au clavier, labellisées selon des émotions induites et déclarées. De plus, cette expérimentation a aussi pour objectif d’améliorer la méthodologie d’induction d’émotions dans un contexte de smartphone, en vérifiant les effets des vidéos par l’intermédiaire de la GEW, de verbatim et de capteurs physiologiques.

Cette expérimentation suit un schéma similaire à la pré-expérimentation. Les sujets interagissent avec un agent conversationnel qui propose un scénario guidé, avec des médias et des tâches à réaliser pour induire certains types d'émotions. Afin de simplifier le protocole expérimental, seule la dimension de la valence a été étudiée lors de cette expérimentation. La dimension de la valence nous a semblé prioritaire à étudier par rapport à l'activation pour des situations de conversation instantanée et plus particulièrement pour le service client. En effet, la valence nous semble être plus représentative que la satisfaction de clients en rapport à l'objectif à atteindre par l'échange. De plus, avec moins de dimensions à intégrer, le protocole expérimental est simplifié. L'induction d'émotions a été réalisée uniquement à partir d'extraits vidéo. Cette méthode d'induction a été conservée, en raison de sa forte représentation dans la littérature.

Les retours sur la pré-expérimentation ont permis d'améliorer la visibilité des extraits vidéo sur l'écran de smartphone. Plusieurs modifications ont été apportées sur l'application :

- La taille des vidéos a été augmentée pour occuper la majorité de l'écran.
- Un fond noir se superpose au reste de l'application et derrière la vidéo pour améliorer le contraste.
- Les vidéos bénéficient aussi d'une meilleure résolution.

De plus, des améliorations diverses ont été apportées à la partie conversation, en particulier les temps de réponse de l'agent conversationnel. La partie « instructions » a été améliorée avec l'ajout d'images pour que les participants comprennent plus aisément les consignes affichées au cours de l'expérimentation.

L'utilisation du SAM nous a semblé limitante dans la pré-expérimentation, c'est pourquoi nous avons cette fois intégré la *Geneva Emotion Wheel* dans l'application pour les auto-évaluations. Cet instrument permet d'étudier les émotions déclarées à la fois sur la dimension de la valence, mais aussi avec plus de granularité avec le choix de labels et d'intensité. Avec vingt libellés proposés de manière visuelle et la possibilité de ne pas déclarer d'émotions ou de proposer d'autres émotions, cet outil a été choisi pour sa flexibilité et sa polyvalence.

IV.4.2 Protocole expérimental

IV.4.2.1 Induction d'émotions et mesure subjective

Nous avons divisé la valence en trois modalités : positive, négative et neutre. Pour chaque modalité quatre extraits vidéo ont été sélectionnés. Les extraits vidéo ont été choisis à partir de la base de données réalisée par Samson *et al.* (2016). Les

extraits vidéo, ainsi que les annotations moyennes sur plusieurs dimensions (valence et activation) et des émotions basiques, sont disponibles sur demande sur le site de l'université de Stanford. Les extraits sont sélectionnés à partir de vidéos amateurs partagées sur l'Internet. Les extraits ne dépassent pas les 30 secondes, ils dépeignent tous des situations réelles, filmées par des amateurs et disponibles sur des sites internet de partage de vidéos. Les vidéos avec une valence négative représentent des accidents relativement violents. Les vidéos avec une valence positive représentent des enfants ou des animaux amusants. Enfin, les vidéos neutres représentent des situations de la vie quotidienne : une terrasse de café ou la préparation d'un repas. Le détail des extraits vidéo sélectionnés est disponible en annexe D.. Les extraits ont été sélectionnés de manière à obtenir des conditions expérimentales avec des niveaux de valence distincts. De plus, les vidéos ont été sélectionnées pour présenter de la variabilité. Par exemple, pour la condition négative, des accidents d'origines diverses ont été sélectionnés pour éviter les répétitions. La valence moyenne des vidéos de chaque condition est indiquée sur le tableau IV.2.

TABLE IV.2 – Moyenne et écart-type des annotations de valence et d'activation des vidéos sélectionnées (Samson *et al.*, 2016)

Condition expérimentale	Valence moyenne (1 à 7)	Écart-type
Valence Négative	1,85	0,34
Valence Positive	4,97	0,27
Valence Neutre	3,19	0,17

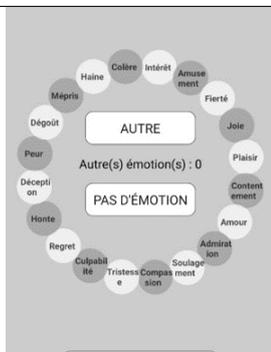
Le contenu de l'expérimentation consiste donc à visualiser les 12 extraits vidéo. Pour chaque vidéo un ensemble d'actions doit être réalisé (voir le tableau IV.3) :

- Pour impliquer les participants, il leur est demandé d'accomplir une tâche plus neutre par rapport à celle de la pré-expérimentation. Les participants doivent décrire les extraits de vidéos visionnés. Aucune consigne particulière n'est donnée pour la description. Les sujets peuvent donc s'exprimer comme ils le souhaitent, tant qu'ils saisissent au minimum une phrase. Il n'y a pas de limite de temps pour décrire la vidéo. Cette tâche permet de rendre les participants actifs et engagés, tout en gardant une vue globale de la vidéo et sans se concentrer sur un aspect trop précis. De plus, cette tâche permet de recueillir d'une part des données des événements claviers et d'autre part les corpus de textes rédigés sur les médias visualisés. Les événements claviers permettent de calculer des descripteurs sur la dynamique de frappe au clavier. Les textes enregistrés peuvent être étudiés par traitement automatique du langage pour en extraire la polarité. Cette phase est particulièrement importante pour l'analyse des données car elle correspond à une expression de la part des participants, directement après l'induction d'émotions réalisée par la vidéo.

- Après chaque vidéo, les participants peuvent déclarer les émotions ressenties sur la GEWMI (voir sous-section IV.2.4) intégrée à l'application d'expérimentation. Les participants peuvent déclarer jusqu'à 2 émotions avec leur intensité ou n'en déclarer aucune. Plusieurs informations peuvent être étudiées à partir de ces auto-évaluations :
 - Le nombre d'émotions sélectionnées (entre 0 et 2)
 - L'intensité des émotions sélectionnées (entre 1 et 6)
 - Les libellés sélectionnés et plus particulièrement la valence (représentée par l'axe horizontal).
 - La proposition d'émotions autres.

Les données des capteurs des smartphones sont enregistrées tout au long de la tâche de la visualisation de la vidéo jusqu'à la fin de l'auto-évaluation. De plus, chaque transition d'une des phases du test est horodatée.

TABLE IV.3 – Enchaînement d'actions pour chaque extrait vidéo

Phases	Phase visionnage	Phase de description	Phase d'auto-évaluation
Captures d'écran			
Données enregistrées	Capteurs inertiels	Capteurs inertiels, Frappe clavier, Verbatim	Capteurs inertiels Émotion déclarée

IV.4.2.2 Enregistrement des capteurs du smartphone

Les capteurs suivants ont été enregistrés au cours de chacune des tâches (voir le tableau IV.4). Chaque capteur est enregistré avec un horodatage à la milliseconde près pour pouvoir être synchronisé précisément avec les phases du test réalisées par le sujet. La fréquence d'échantillonnage des capteurs inertiels a été augmentée à 50Hz, soit un enregistrement toutes les 20 millisecondes. Cette fréquence d'échantillonnage a été sélectionnée en correspondance avec le corpus d'activité humaine enregistré avec

un smartphone réalisé par Davide Anguita *et al.* (2013). Des fréquences similaires ou inférieures sont utilisées dans des études de reconnaissance d'émotions sur smartphone (Olsen & Torresen, 2016; Ruensuk *et al.*, 2020). Les détails des enregistrements sont disponibles sur le IV.4.

TABLE IV.4 – Données enregistrées

Capteurs	Fréquence d'échantillonnage	Mesures
Accéléromètre	50Hz	Accélération sur trois axes : x, y et z (m/s^2)
Gyroscope	50Hz	Mouvement angulaire sur trois axes : x, y et z (rad/s)
Magnétomètre	50Hz	Champ magnétique sur trois axes : x, y et z (G)

Les évènements de frappe au clavier sont enregistrés avec le contenu de la touche et l'horodatage de l'appui.

Le traitement de ces données est explicité dans la section

IV.4.2.3 Enregistrement des mesures physiologiques

La montre Empatica E4 a été utilisée au cours de l'expérimentation afin de recueillir des données physiologiques sur les participants. Les variables enregistrées sont listées sur le tableau IV.5.

TABLE IV.5 – Capteurs enregistrés par la montre Empatica E4

Capteur	Fréquence d'échantillonnage	Mesures
Accéléromètre	32Hz	Accélération sur trois axes : x, y et z (m/s^2)
Photopléthysmographe (BVP)	64Hz	Volume de pulsation sanguine (nW)
Capteur de conductance électrodermale (CE)	4Hz	Réponse galvanique de la peau (RGP) (Microsiemens μS)
Thermopile infrarouge	4Hz	Température de la peau ($^{\circ}C$)

La montre Empatica E4 se porte sur le poignet, une application Android dédiée se synchronise avec la montre en *Bluetooth* pour enregistrer les données de la montre. De la même manière que pour les capteurs du smartphone, les données sont horodatées pour être mises en correspondance avec les actions réalisées par les participants de l'expérimentation.

IV.4.2.4 Plan expérimental

Avant l'expérimentation, les participants devaient installer la montre Empatica sur leur main non dominante. Afin de calibrer la montre Empatica E4, les participants ont dû visualiser une vidéo de 15 minutes (Fowles *et al.*, 1981). Les créateurs de la montre recommandent cette phase de calibration pour ajuster la captation de la conductance électrodermale. Un extrait de documentaire avec une valence neutre a été sélectionné pour cela. Cette phase permet notamment de conduire les participants vers un état initial neutre au début de l'expérimentation. Les participants restent assis tout au long de l'expérimentation.

Une fois la vidéo de calibration terminée, l'expérimentation sur smartphone débute. Le *chatbot* explique aux participants le principe général de l'expérimentation, à savoir qu'ils vont devoir visionner des extraits vidéo et interagir avec l'interface. Le terme « émotion » n'est pas évoqué pour ne pas biaiser les réactions des participants. L'agent conversationnel adresse les questionnaires sur les caractéristiques personnelles et sur l'usage du smartphone (voir les annexes C.I et C.II), au préalable. Des instructions expliquent aux sujets le déroulement d'une tâche. Une tâche d'entraînement avec une vidéo neutre et une auto-évaluation est proposée, permettant au participant de se familiariser avec l'interface du clavier et de GEWMI. L'entraînement peut être répété plusieurs fois si nécessaire. Une fois que le participant valide avoir compris les instructions, le test peut débuter.

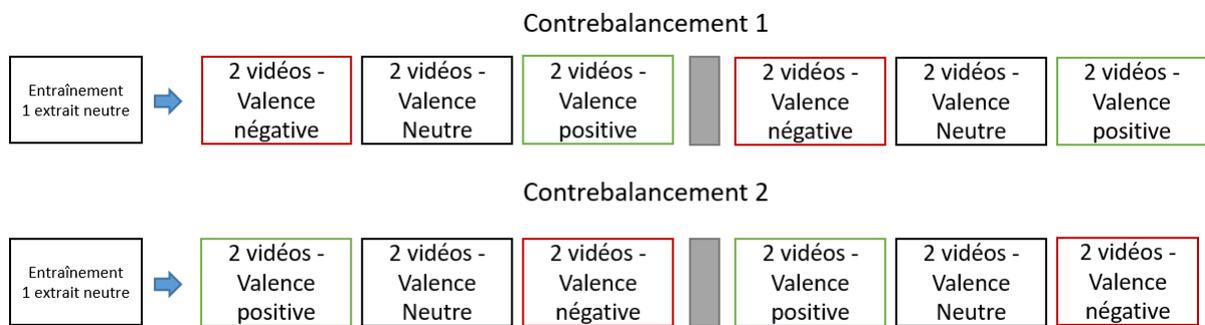


FIGURE IV.9 – Contre-balancements

Les extraits vidéo et les tâches, explicités dans le tableau IV.3, s'enchaînent ensuite automatiquement. Deux contre-balancements sont proposés pour éviter les effets d'ordre ou de fatigue sur les modalités de valence (voir figure IV.9). Au sein des conditions expérimentales (valence positive, négative ou neutre) les vidéos sont distribuées de manière aléatoire. À la moitié des vidéos diffusées, l'agent conversationnel propose une courte pause au participant avant de reprendre sur le reste des vidéos.

Après la fin du test, deux questionnaires complémentaires sont adressés aux sujets :

- Un questionnaire pour évaluer la facilité d’usage de l’outil GEWMI pour l’expression d’émotions. Ce questionnaire est adapté du questionnaire proposé par Siegert *et al.* (2011). Il est composé de quatre questions, disponibles en annexe C.III.
- Afin d’évaluer les profils émotionnels des participants, dans l’optique de réaliser des analyses complémentaires, un questionnaire d’impulsivité émotionnelle a été intégré. Le questionnaire a été constitué à partir de sept items du questionnaire d’expressivité émotionnelle Berkeley (Gross *et al.*, 1995). Une version francophone a été utilisée, proposée par Caroline Ouellet *et al.* (2012). Les questions sont accessibles en annexe C.IV.

Un champ libre a été laissé à la toute fin pour permettre aux participants d’exprimer leurs avis sur l’ergonomie de l’application et le déroulement de l’expérimentation. L’ensemble des différentes étapes de l’expérimentation est résumé sur le schéma de la figure IV.10.

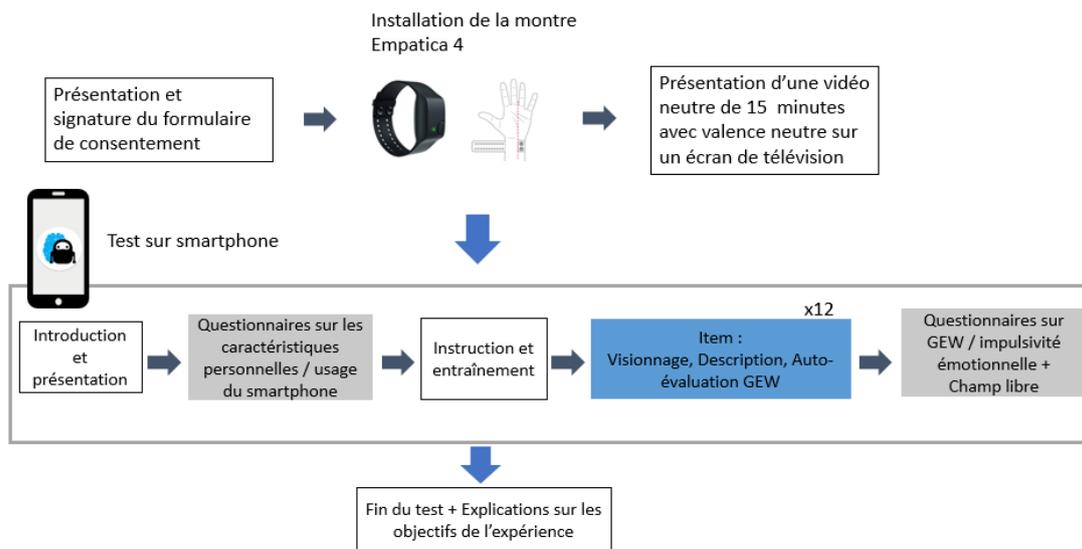


FIGURE IV.10 – Déroulement global de l’expérimentation

IV.4.3 Hypothèses générales

Premièrement, nous nous intéressons à nouveau à la question fondamentale **Q_Induction**. Ainsi, nous formulons l’hypothèse générale associée à cette question de recherche de la manière suivante : **pour l’ensemble des participants, la méthode d’induction d’émotions de cette expérimentation par des extraits vidéo permet d’obtenir des effets sur la valence et l’intensité, en correspondance avec les**

conditions expérimentales proposées. Pour vérifier cette hypothèse, nous mesurons les auto-évaluations de l'instrument GEW. Ainsi, nous postulons qu'en fonction des conditions expérimentales des vidéos, le nombre d'émotions, l'intensité et la valence des auto-évaluations sur la GEW sont affectés. En complément, nous analysons aussi la polarité des verbatims des descriptions. Un ensemble d'hypothèses opérationnelles est proposé et analysé dans la sous-section [V.2.3](#).

Deuxièmement, nous désirons apporter des réponses aux questions fondamentales **Q_Interindividu** et **Q_Intraindividu**. Les données d'interaction entre l'utilisateur et le smartphone correspondent à l'enregistrement des capteurs inertiels et de la frappe au clavier. Ces données servent à l'apprentissage du système automatique de reconnaissance d'émotions. Plusieurs types de modèles peuvent être conçus : des modèles généraux ou interpersonnels, qui utilisent les données de tous les individus, et des modèles personnalisés ou intra-individuels, qui s'entraînent à partir des données d'un seul individu. Nous postulons les hypothèses générales suivantes :

- **L'utilisation des données de tous les individus pour entraîner des modèles inter-individuels permet de classer automatiquement des modalités de valence déclarées, avec une performance supérieure au hasard.**
- **L'utilisation des données d'un seul individu pour entraîner des modèles intra-individuel permet de classer automatiquement des modalités de valence déclarées, avec une performance supérieure au hasard et avec une plus forte robustesse que les modèles inter-individuels.**

Les analyses de ces hypothèses générales sont décrites dans la section [V.3](#).

Enfin, en ce qui concerne la question fondamentale **Question Q_Equivalence_Physio**, nous avons pu voir dans le chapitre [III](#) que les données des capteurs physiologiques peuvent aussi servir à entraîner des modèles de reconnaissance automatique d'émotions. Nous pouvons ainsi comparer les robustesses des modèles entraînés à partir de ces données avec les modèles de classification basés sur les interactions avec un smartphone. Nous postulons l'hypothèse générale suivante : **Les performances de classification des modèles entraînés avec des données physiologiques sont similaires aux modèles entraînés avec des données issues des interactions avec un smartphone.** Pour cela, nous utilisons les données recueillies avec la montre Empatica E4 au cours de l'expérimentation. Les analyses sont disponibles dans la section [V.3](#).

IV.4.4 Passation

L'expérimentation s'est déroulée en septembre et octobre 2020, dans les différents locaux d'Orange Labs Services : Pessac, Cesson-Sévigné et Lannion. Le protocole

expérimental a été validé par le comité opérationnel d'évaluation des risques légaux et éthiques (COERLE) du laboratoire INRIA Sud-Ouest. Un formulaire de consentement a été adressé à chaque participant en amont des expérimentations (visible en annexe B.). En raison de la pandémie du Covid-19, une attention particulière a été portée pour le respect des règles d'hygiène. Le matériel a été désinfecté entre chaque passation. Les participants ont passé le test sur un smartphone Huawei P Smart. Les participants ont été filmés au cours l'expérimentation (voir figure IV.11). Cependant, les participants ont dû porter le masque tout au long de l'expérimentation.

Des employés d'Orange Labs Services et des étudiants ont passé l'expérimentation, pour un total de 40 participants, dont 21 hommes et 19 femmes. La distribution des âges est la suivante :

- 10 sujets entre 18 et 24 ans.
- 7 sujets entre 25 et 34 ans.
- 14 sujets entre 35 et 49 ans.
- 9 sujets entre 50 et 64 ans.



FIGURE IV.11 – Capture de l'enregistrement vidéo de l'expérimentation

94% des sujets ont déclaré être droitiers. De plus 97,5% ont estimé qu'ils utilisaient leur smartphone plus de 30 minutes par jour en moyenne et 60% ont déclaré utiliser des applications de messagerie instantanée plus de 30 minutes par jour en moyenne.

IV.5 Expérimentation complémentaire

IV.5.1 Présentation et objectif

Les expérimentations précédentes ont principalement porté sur la diffusion d'un média pour induire des émotions. Bien qu'il s'agisse d'une approche particulièrement privilégiée en science des émotions pour provoquer des épisodes émotionnels, elle reste conditionnée par l'empathie ressentie envers les personnages ou situations présentées. Les sujets peuvent se sentir affectés par l'aspect attrayant ou aversif des situations, mais ne sont pas directement concernés.

Selon la théorie de l'évaluation cognitive, les processus émotionnels intègrent une composante cognitive et une composante motivationnelle. La réussite ou l'échec d'objectifs personnels sont des critères à l'origine d'épisodes émotionnels. Par exemple, le modèle OCC Ortony *et al.* (1990) décrit des émotions comme des actions réalisées par soi-même, qui peuvent causer de la fierté ou de la honte en fonction des objectifs.

Le protocole expérimental de la manipulation de difficulté, présenté dans la sous-section III.5.2.3, permet de simuler des situations avec un objectif à atteindre. La manipulation des conditions expérimentales favorise l'apparition de réussites ou d'échecs avec pour objectif d'induire des émotions à valence positive ou négative. Ce type de protocole nous semble se rapprocher des émotions de frustration ou de satisfaction ressenties dans le cadre de service client, dans lesquels un but précis est recherché par les clients.

L'objectif de cette expérimentation est donc de vérifier la pertinence d'un protocole de manipulation de difficulté dans le contexte d'un usage de smartphone. Des tests logiques ont donc été utilisés pour créer les conditions expérimentales avec un objectif à atteindre. Les analyses ont ensuite porté sur les déclarations des participants sur l'instrument GEWMI.

IV.5.2 Protocole expérimental

IV.5.2.1 Induction d'émotions et mesure subjective

Le protocole d'induction d'émotions a été inspiré des travaux de Schuch *et al.* (2017), présentés dans la sous-section III.5.2.3. Deux conditions expérimentales ont été réalisées : une condition facile et une condition difficile. Chaque sujet ne passe qu'une des deux conditions expérimentales. Dans les deux cas de figure, les sujets doivent résoudre 12 items.

Des matrices progressives de Raven pour adultes ont été utilisées pour concevoir les conditions expérimentales (Raven *et al.*, 2003). Deux types d'items ont été sélectionnés :

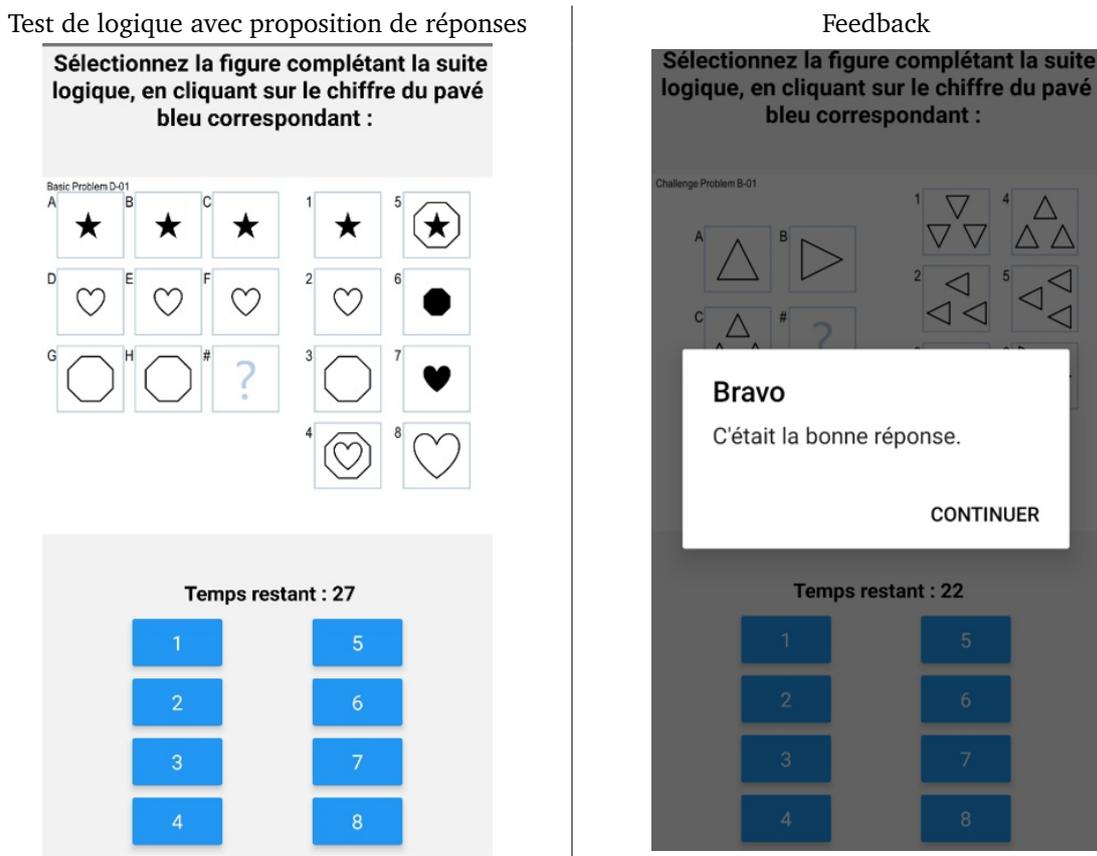
- Des tests de logique difficiles provenant de la liste de tests avancés.
- Des tests de logique plus accessibles provenant de la liste de tests basiques.

Les conditions expérimentales ont été assemblées de la manière suivante :

- Pour la condition facile : neuf items faciles et trois items difficiles.
- Pour la condition difficile : trois items faciles et neuf items difficiles.

Au sein de chaque condition, les tests de logique sont répartis de manière aléatoire. Pour chaque item de test, il n'y a qu'une réponse possible. Une limite de 30 secondes est imposée pour fournir une réponse. S'il n'y a pas de réponse dans les 30 secondes, l'item est considéré comme échoué. Un feedback est donné dans une fenêtre indépendante et par le chatbot en fonction de la validité de la réponse. La figure IV.12 représente l'interface proposée pour résoudre les tests de logiques. Le choix a été fait de séparer l'interface avec des tests logiques de la fenêtre de conversation pour une meilleure visibilité et pour des raisons techniques (intégration d'un décompte facilité).

FIGURE IV.12 – Déroulement du test de logique



IV.5.2.2 Plan expérimental

Au préalable, l'agent conversationnel explique l'objectif de l'expérience, à savoir tester une application pour adresser des tests de logique de manière automatique. Le *chatbot* précise que les performances des utilisateurs sont évaluées lors de cette expérimentation. Les émotions ne sont pas évoquées pour ne pas biaiser les réponses des participants.

Les questionnaires sur les caractéristiques personnelles sur l'usage de smartphone sont adressés. L'agent conversationnel donne ensuite les instructions pour interagir avec les tests de logique et pour utiliser l'instrument GEWMI. Quatre items faciles sont proposés en entraînement, suivis d'une première auto-évaluation sur la GEW. L'agent conversationnel indique aux participants qu'en moyenne les étudiants résolvent plus de la moitié des items et qu'ils doivent faire de même pour réussir le test. Une fois que le sujet valide la bonne compréhension des consignes, le test peut débuter. Les 12 items s'enchaînent automatiquement. Un compteur indique le score en cours et le nombre total d'items.

À la fin de la phase de test, le chatbot résume les résultats obtenus par les participants et indique si l'objectif est atteint ou pas. Après cette information, il est demandé aux participants d'évaluer leurs propres émotions sur GEWMI. Des questionnaires sur l'impulsivité émotionnelle et sur l'usage de l'outil d'auto-évaluation sont ensuite adressés, comme lors de l'expérimentation précédente. À la fin du test, les expérimentateurs expliquent aux sujets l'objectif des recherches. La figure IV.12 résume l'ensemble de l'expérimentation.

IV.5.3 Passation

En raison du confinement d'octobre à décembre 2020, les expérimentations ont dû être réalisées à distance. Le fonctionnement de la gestion de données de l'application a permis de mettre en place cette expérimentation sans modifications importantes. Les expérimentateurs se sont connectés par visioconférence avec les sujets pour leur expliquer comment installer l'application sur leur téléphone et pour leur transmettre des conditions générales pour passer le test.

La population de sujets se compose exclusivement d'étudiants de l'université de Bordeaux et de l'École Nationale Supérieure de Cognitique de Talence. La distribution des âges est donc différente par rapport aux autres expérimentations :

- 36 sujets entre 18 et 24 ans.
- 4 sujets entre 25 et 34 ans.

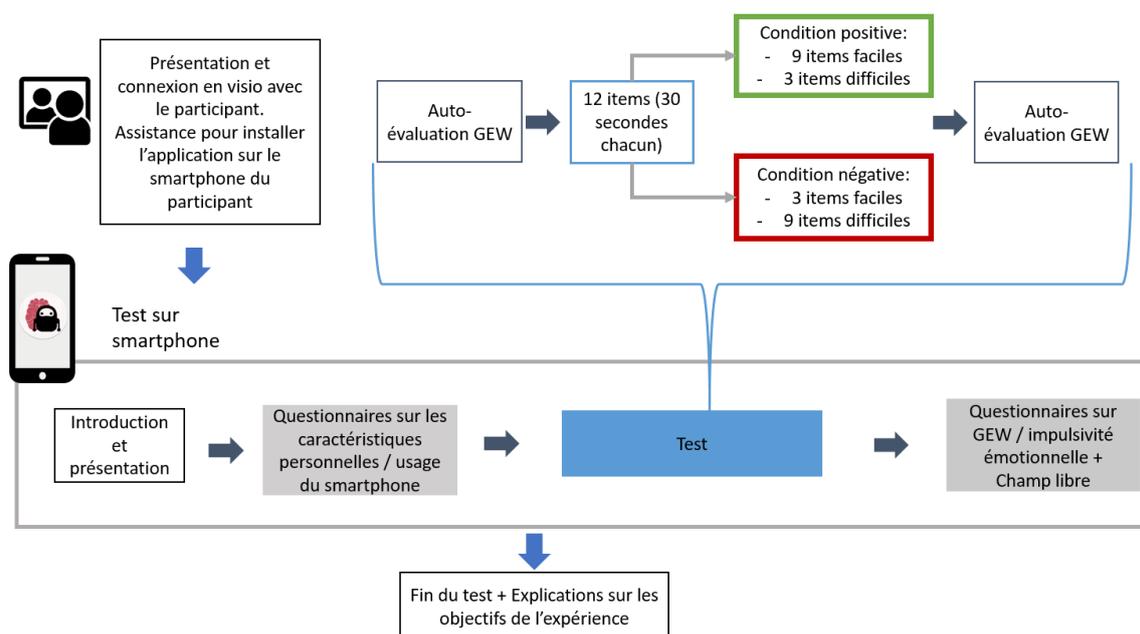


FIGURE IV.13 – Déroulement global de l'expérimentation

Parmi les sujets, 24 sont des hommes et 16 des femmes. Tous les participants ont déclaré utiliser leur smartphone plus de 30 minutes par jour, la moitié ayant déclaré l'utiliser plus de deux heures. La moitié des sujets est passée sur la condition facile et la deuxième est passée sur la condition difficile.

IV.5.4 Hypothèse générale

Avec ce protocole expérimental, nous portons notre attention à la question fondamentale **Question Q_Induction**, afin d'apporter de nouveaux éléments de réponse. Ainsi, nous proposons l'hypothèse générale suivante : **la méthode d'induction d'émotions par manipulation de la difficulté permet d'obtenir des effets mesurables sur la valence, en correspondance avec les conditions expérimentales**. Les auto-évaluations du sentiment subjectif, à partir de la GEW, nous permettent d'estimer les effets du protocole sur les sujets. Les analyses et les réponses aux hypothèses opérationnelles sur cette expérimentation sont disponibles en sous-section [V.2.4](#).

Nous pouvons aussi postuler que le choix des libellés sur la roue a un effet sur les résultats par rapport à l'expérimentation principale. En raison du type de protocole, impliquant plus personnellement les participants et leur performance.

IV.6 Méthode de reconnaissance automatique d'émotions

IV.6.1 Introduction

Les données collectées au cours de l'expérimentation principale forment un corpus permettant de construire un système de reconnaissance automatique des émotions. L'utilisation d'apprentissage automatique est préférable pour traiter l'ensemble des données collectées. En effet, il n'est pas envisageable de déduire des règles ou patrons liés à certains types d'émotions de manière manuelle, au vu de la quantité de données et de leur complexité. Des modèles d'apprentissage automatique sont donc appropriés pour tenter de détecter automatiquement des informations pertinentes pour discriminer des états émotionnels. En l'état, les données enregistrées ne sont pas directement exploitables. Il est donc nécessaire d'appliquer des traitements au préalable pour générer des descripteurs (ou *features* en anglais) afin d'entraîner les modèles de classification. Une alternative serait d'utiliser des modèles d'apprentissage profond, qui sont en mesure de générer eux-mêmes des descripteurs de manière automatique (la différence entre ces deux approches a été présentée dans la sous-section III.3.2 dans la figure III.1). Cependant, l'apprentissage profond nécessite d'avoir un jeu de données de très grande taille et requiert des ressources matérielles importantes. De plus, le fonctionnement en boîte noire, rend plus complexe la compréhension des paramètres de ces modèles d'apprentissage profond. Il est donc plus pertinent d'utiliser, dans un premier temps, des approches plus simples. L'apprentissage automatique classique est d'ailleurs largement utilisé dans les recherches sur la reconnaissance d'émotions en informatique affective (Coutrix & Mandran, 2012; Cui *et al.*, 2015; Hashmi *et al.*, 2020; Kołakowska *et al.*, 2020; Mottelson & Hornbæk, 2016; Olsen & Torresen, 2016).

Afin d'utiliser les enregistrements des capteurs de smartphones ou d'appareils de mesure physiologiques, nous pouvons noter deux étapes principales :

- Premièrement, un ensemble de traitements doit être mis en œuvre afin de transformer des enregistrements bruts en jeux de données.
- Deuxièmement, des modèles de classification sont entraînés et évalués à partir de ces jeux de données.

Dans cette section, nous détaillerons la démarche pour traiter les enregistrements des capteurs du smartphone et de la montre Empatica E4. L'exploitation des données issues des capteurs de smartphones constitue notre principal objectif, afin de déterminer la robustesse de modèles de reconnaissance automatique se fondant sur ces données. Les données physiologiques représentent un moyen supplémentaire pour inférer sur l'induction d'émotions chez les participants de l'expérimentation. Nous ne présenterons que les données enregistrées lors de l'expérimentation principale, afin de simplifier l'explication de la démarche de traitement et d'exploitation des données. L'ensemble

du code permettant d'obtenir ces résultats a été développé en langage Python.

IV.6.2 Démarche générale

Pour réaliser un système de reconnaissance automatique, les enregistrements doivent être mis en forme. Les données issues des capteurs inertiels et physiologiques sont des séries temporelles. Il s'agit donc de valeurs numériques qui évoluent sur un axe temporel. Ces données ont fait l'objet de traitements visant à produire un ensemble de descripteurs numériques, permettant de résumer les patterns d'un échantillon de signal. Ces traitements permettent l'utilisation des modèles d'apprentissage automatique. L'ensemble de cette démarche a été brièvement abordée dans la sous-section [III.4.3](#) pour les capteurs de smartphones et dans la sous-section [III.3.5](#) pour les capteurs physiologiques. La liste des étapes à appliquer est la suivante :

- **Extraction des signaux bruts** : les enregistrements sont extraits de leur format d'origine pour être convertis dans un format lisible en langage Python.
- **Prétraitement des signaux** : des filtres sont appliqués afin de retirer des données inutiles aux enregistrements. De plus, certains signaux peuvent servir de base pour générer d'autres signaux afin d'obtenir des informations complémentaires.
- **Génération des descripteurs** : les signaux filtrés font ensuite l'objet d'un échantillonnage afin de générer des bouts de signaux de durée égale. À partir de ces échantillons, des signaux supplémentaires pourront aussi être générés pour obtenir des représentations complémentaires de la série temporelle. Enfin, un ensemble de descripteurs est calculé pour chaque échantillon, afin de constituer un jeu de données complet. Des descripteurs provenant de différents signaux peuvent être fusionnés afin de créer un jeu de données plus conséquent.
- **Apprentissage automatique** : les échantillons sont associés à des labels (par exemple, une auto-évaluation d'un participant ou une condition expérimentale). Les modèles de classification se configurent de manière à trouver les paramètres les plus optimisés. Les résultats doivent ensuite être évalués pour déterminer la robustesse des modèles de classification.

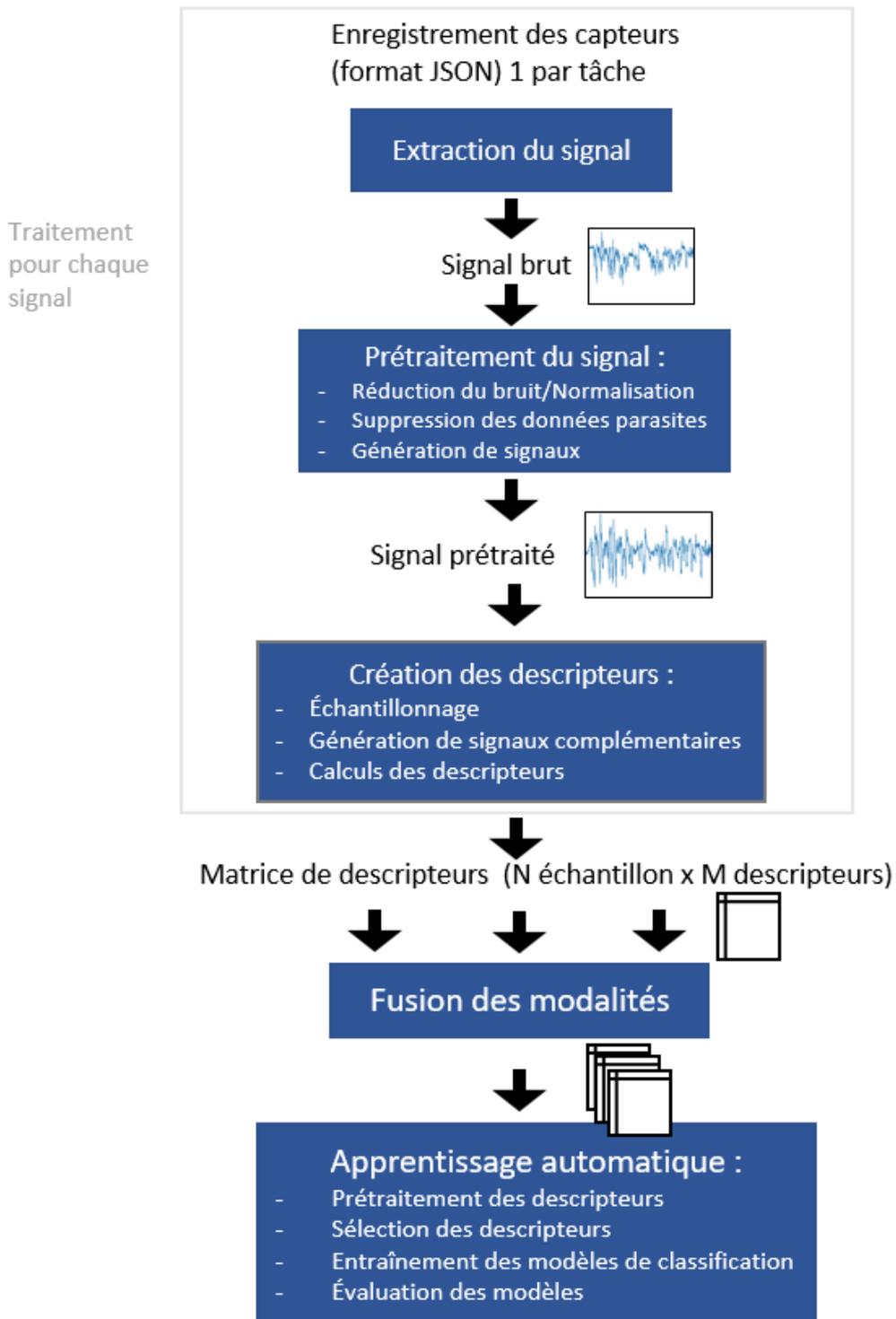


FIGURE IV.14 – Démarche générale pour le traitement des signaux et l'apprentissage automatique

Nous détaillerons chacune de ces étapes dans les sous-sections suivantes. Le schéma IV.14 résume les étapes générales pour obtenir des modèles de classification à partir des enregistrements. Il faut noter que certains signaux n'ont pas besoin de prétraitement particulier pour passer à l'étape de la génération de descripteurs. C'est notamment le cas pour les événements claviers et certaines mesures physiologiques. La démarche de traitement des données jusqu'à l'obtention de matrices de descripteurs est à appliquer pour chaque tâche de l'expérimentation. Il y a donc 480 enregistrements à traiter pour chaque descripteur (12 tâches pour chacun des 40 participants).

IV.6.3 Extraction des signaux bruts

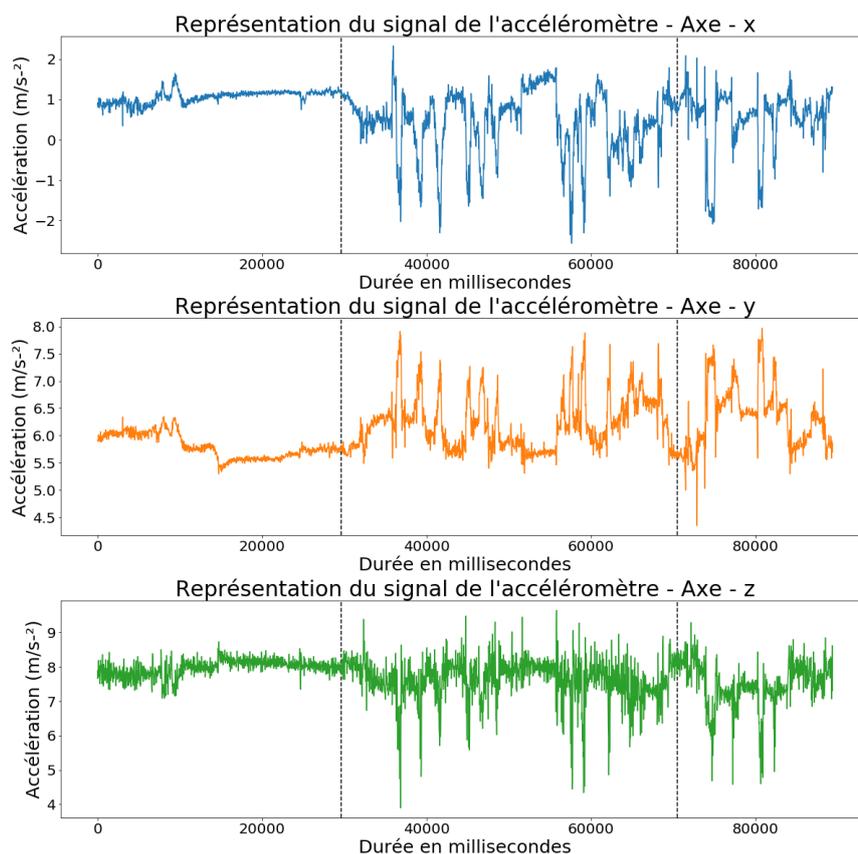


FIGURE IV.15 – Représentation d'une série temporelle de signal d'accéléromètre pour une tâche

Pour chaque tâche, les enregistrements sont extraits du format JSON afin d'être exploitables en code Python. La librairie *Pandas* permet de lire les données dans un format structuré et exploitable. À ce stade, les signaux peuvent être représentés graphiquement (McKinney, 2010). Par exemple, la figure IV.15 représente le signal d'accéléromètre brut sur les trois axes (x, y et z), pour une tâche de l'expérimentation principale, pour un participant. Elle inclut les phases de visionnage, de description et d'auto-évaluation. Visuellement, nous pouvons constater la différence de mouvements entre les différentes phases. En effet, les phases de description et d'auto-évaluation montrent des mouvements plus importants par rapport à la phase de visionnage. L'axe temporel est en millisecondes.

Des exemples de représentations des signaux sont disponibles en annexe F.I pour les capteurs de smartphone et F.II pour les capteurs de la montre Empatica E4. Pour ces derniers, une opération complémentaire est nécessaire pour faire correspondre les enregistrements issus de la montre avec les événements des tâches de l'expérimentation sur smartphone. Les horodatages des événements de l'expérimentation d'une part, et les horodatages des enregistrements des capteurs de la montre Empatica E4 d'autre part ont permis de réaliser cette opération.

IV.6.4 Prétraitement des signaux

IV.6.4.1 Enregistrements des capteurs inertiels

Ces données sont issues des enregistrements des capteurs d'accéléromètre, de gyromètre et de magnétomètre du smartphone, ainsi que de l'accéléromètre de la montre Empatica E4. Ces signaux suivent une procédure similaire pour obtenir des descripteurs utiles à la reconnaissance d'émotions. Ils présentent chacun la caractéristique d'être composés de trois axes dans l'espace x, y et z. Ainsi, chaque axe doit être traité séparément. Dans un premier temps, afin de supprimer certaines informations inutiles, des filtres sont appliqués (Anguita *et al.*, 2013). Pour rappel, les capteurs inertiels ont enregistré les mouvements appliqués au smartphone (voir sous-section III.4.3.2). Pour les capteurs inertiels, trois filtres en particulier peuvent être appliqués :

- Premièrement, pour réduire le bruit du signal, un filtre médian ou moyen afin de diminuer les extrémités (Kołakowska *et al.*, 2020). Nous appliquons un filtre médian de taille trois sur les signaux.
- Deuxièmement, l'application d'un filtre Butterworth passe-haut à 0,3 Hz d'ordre trois permet de supprimer la force de gravité enregistrée par les capteurs.
- Enfin, pour réduire le bruit, un filtre Butterworth passe-bas à 20 Hz d'ordre trois est appliqué. En effet, selon Anguita *et al.* (2013), 99% de l'information sur les mouvements humains est contenue en dessous de 15 Hz, pour garder une marge, nous appliquons donc un filtre à 20 Hz.

IV.6. MÉTHODE DE RECONNAISSANCE AUTOMATIQUE D'ÉMOTIONS

Pour appliquer ces filtres, des fonctions de la bibliothèque *Scipy* ont été utilisées (Virtanen *et al.*, 2020). La figure IV.16 permet de visualiser le signal d'une tâche entière après application des filtres (représenté en format brut dans la figure IV.15). Nous pouvons observer des différences notables par rapport au signal brut. De fait, les accélérations sont désormais toutes centrées en zéro pour chaque axe, indiquant la suppression de la force de gravité enregistrée.

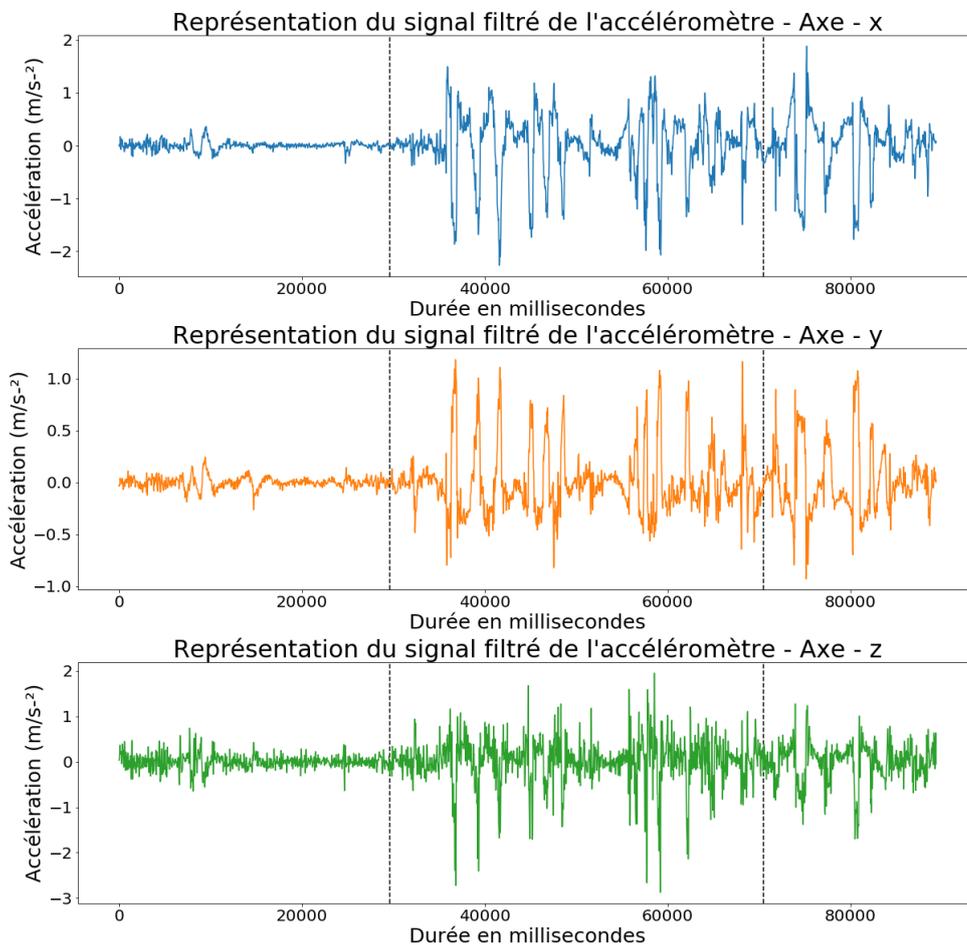


FIGURE IV.16 – Représentation d'une série temporelle filtrée de signal d'accéléromètre pour une tâche

IV.6.4.2 Mesures physiologiques

Les mesures physiologiques sont issues des enregistrements des capteurs physiologiques intégrés à la montre Empatica E4. L'ensemble des signaux enregistrés a été présenté dans la sous-section [IV.4.2.3](#). Contrairement aux signaux des capteurs inertiels, les signaux des capteurs physiologiques n'ont pas tous la même fréquence d'échantillonnage et les mêmes caractéristiques, il est donc nécessaire d'appliquer un prétraitement adapté à chaque modalité. Les prétraitements sur ces enregistrements ont été réalisés en prenant comme références les articles de Bobo Zhao *et al.* (2018) et Philip Schmidt *et al.* (2018), qui ont traité des enregistrements de la montre Empatica E4 dans leur étude. Deux modalités en particulier ont nécessité un prétraitement :

- La conductance électrodermale : ce capteur a une fréquence d'échantillonnage de 4Hz avec la montre *Empatica E4*, soit une mesure toutes les 250 millisecondes. L'enregistrement a été normalisé entre 0 et 1, afin de mettre sur la même échelle de valeurs les données de chaque participant.
- La variation de fréquence cardiaque (VFC) : ce signal n'est pas directement enregistré par les capteurs de la montre. Il peut être calculé à partir de la mesure de la pulsation sanguine. La bibliothèque Python *Heartpy* a été utilisée pour obtenir la VFC (van Gent *et al.*, 2018). La VFC contient des informations essentielles sur le système nerveux autonome (SNA). Le signal obtenu correspond à une donnée toutes les secondes, soit une fréquence d'échantillonnage de 1Hz.

À la suite de ces prétraitements, les enregistrements peuvent être exploités pour générer des matrices de données.

IV.6.5 Génération de descripteurs

IV.6.5.1 Échantillonnage

L'échantillonnage, ou segmentation, consiste à découper les signaux sur des intervalles définis et réguliers. L'objectif est de pouvoir générer pour chaque échantillon ou segment un ensemble de descripteurs pour résumer une portion de signal. Yonggang Lu *et al.* (2017) préconisent d'échantillonner des signaux provenant des capteurs inertiels avec une superposition de 75% afin d'optimiser la quantité de données utilisée. La durée des échantillons est de 10 secondes, afin de conserver une durée suffisante pour capter des patterns significatifs. Nous appliquons donc une fenêtre glissante d'une durée de 10 secondes, toutes les 2,5 secondes, pour découper les échantillons. Ce processus est schématisé en haut de la figure [IV.19](#). Ce traitement est appliqué à tous les signaux, avec les mêmes paramètres. En effet, cela permet de faciliter la fusion des différentes modalités à la fin de l'extraction des descripteurs, en faisant correspondre des échantillons de même taille entre eux.

IV.6.5.2 Descripteurs des signaux inertiels

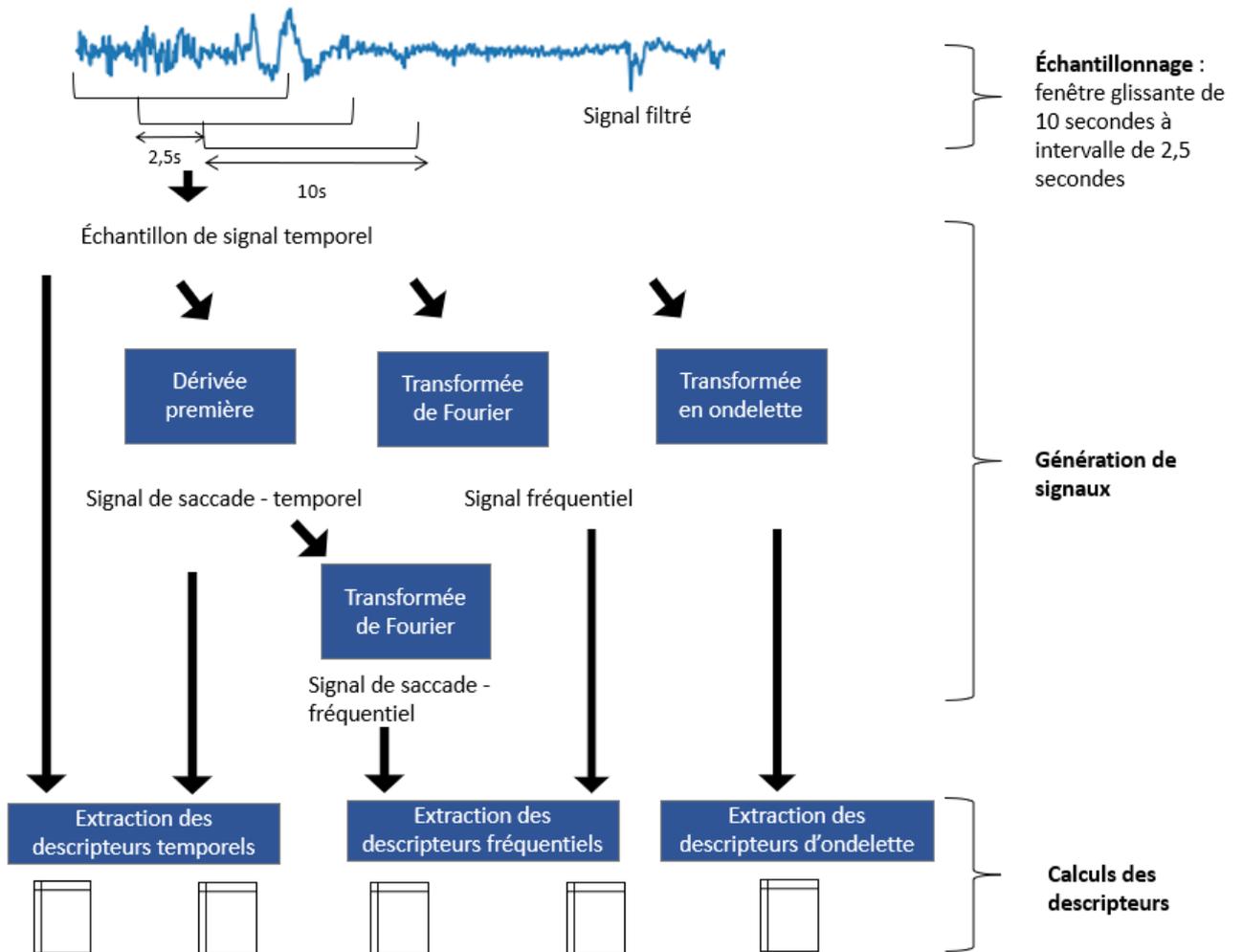


FIGURE IV.17 – Représentation des processus pour l'échantillonnage et l'extraction des descripteurs

L'objectif de cette étape est de synthétiser chaque échantillon de 10 secondes, en calculant un ensemble de descripteurs. Au préalable, à partir de chaque échantillon, nous pouvons calculer des signaux complémentaires pour augmenter la quantité de données (Anguita *et al.*, 2013) :

- Le signal fréquentiel : une transformée de Fourier permet de décomposer un signal temporel en le considérant comme une somme de sinusoides de différentes fréquences. Les résultats de la transformée, nous permet alors de visualiser les différentes fréquences présentes dans ce signal sur un axe non plus temporel,

mais fréquentiel. La figure IV.18 représente la transformation d'un échantillon d'un signal d'accéléromètre sur le domaine temporel vers sa représentation dans le domaine fréquentiel. L'axe de l'amplitude représente l'importance relative de chaque fréquence présente dans le signal. L'amplitude est calculée sur une bande fréquentielle entre 0,3 et 20 Hz, correspondant aux filtres appliqués au préalable.

- Le signal de saccade : la saccade (*jerk* en anglais) représente les variations brusques d'une accélération. Elle est calculée à partir de la dérivée temporelle du vecteur d'accélération en fonction du temps, telle que présentée dans la formule suivante : $\vec{j} = \frac{d\vec{a}}{dt}$. La figure IV.19 représente ce signal à partir d'un échantillon d'accéléromètre de 10 secondes.
- La transformée fréquentielle du signal de saccade : le signal de saccade peut être adapté dans le domaine fréquentiel à partir d'une transformée de Fourier, de la même manière que pour le signal d'accélération initial.
- La transformée en ondelettes : il s'agit d'une fonction similaire à la transformée de Fourier, permettant de décomposer un signal. À la différence d'une transformée de Fourier, une transformée en ondelettes utilise d'autres fonctions que des sinusoides. Selon Edmond Mitchell *et al.* (2013), la transformée en ondelettes est particulièrement appropriée pour extraire des descripteurs informatifs à partir de signaux de capteurs inertiels, car elle s'adapte à des séries temporelles discontinues.

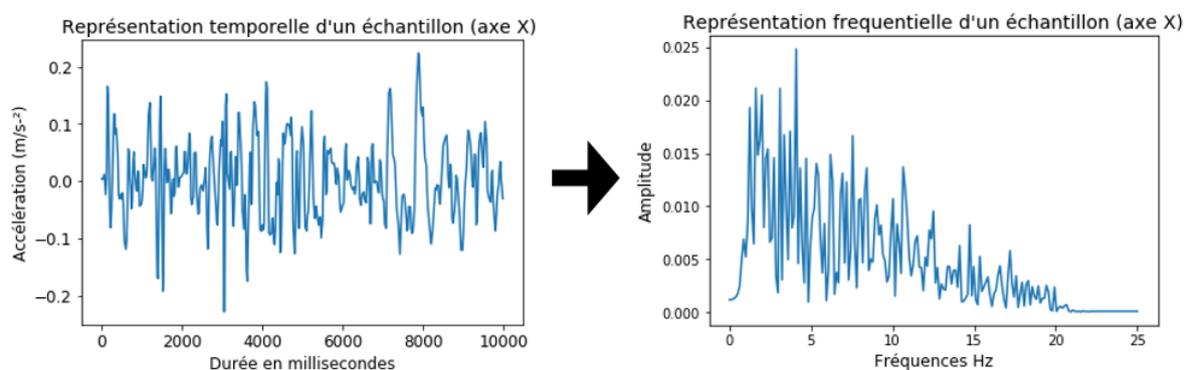


FIGURE IV.18 – Représentation des processus pour l'échantillonnage et l'extraction des descripteurs

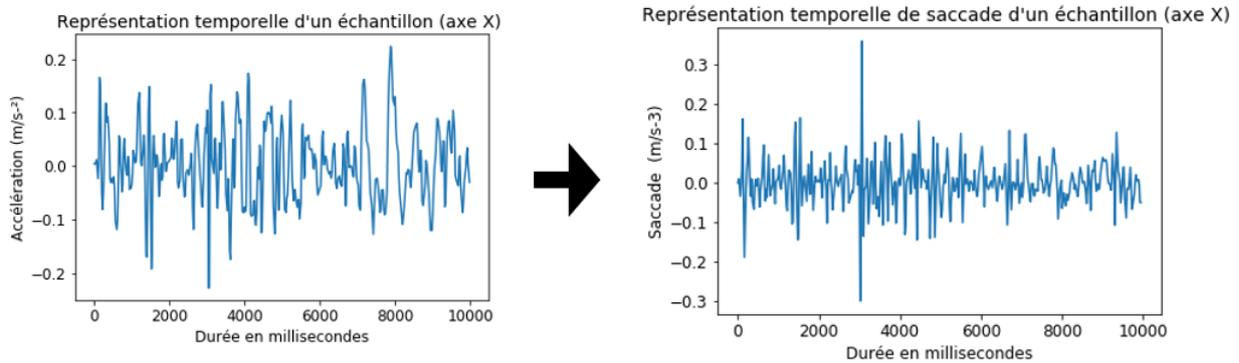


FIGURE IV.19 – Représentation des processus pour l'échantillonnage et l'extraction des descripteurs

Ainsi, nous pouvons générer des descripteurs à partir de ces nouveaux signaux. Il peut s'agir, pour les descripteurs les plus simples, de calculs de statistiques descriptives telles que des moyennes, variances ou écart-type. Nous pouvons diviser les descripteurs en trois catégories : les descripteurs issus des signaux temporels, les descripteurs issus des signaux fréquentiels et les descripteurs issus des transformées en ondelettes. Les descripteurs temporels ont été calculés par l'intermédiaire des bibliothèques *Numpy* et *Scipy* (van der Walt *et al.*, 2011; Virtanen *et al.*, 2020). La liste des descripteurs est détaillée sur le tableau IV.6, le détail des calculs est disponible en annexe G.. Ces calculs s'appliquent aux signaux des capteurs d'accéléromètre, de gyromètre et de magnétomètre. Ces descripteurs sont utilisés à la fois dans les recherches sur la reconnaissance de l'activité physique à partir d'enregistrements de capteurs inertiels de smartphones (Anguita *et al.*, 2013) ou plus récemment dans la reconnaissance d'émotions (Cui *et al.*, 2015; Kołakowska *et al.*, 2020; Mottelson & Hornbæk, 2016; Olsen & Torresen, 2016; Ruensuk *et al.*, 2020).

Les descripteurs des transformées en ondelettes ont été calculés par l'intermédiaire de la bibliothèque *PyWavelets* (Lee *et al.*, 2019). Selon la méthodologie proposée par Yufei Chen et Shen (2017), nous avons appliqué une transformée d'ondelettes avec une forme d'onde Daubechie 4 sur cinq bandes de fréquence : 0,3-1,25 Hz, 1,25-2,5 Hz, 2,5-5 Hz, 5-10 Hz et 10-20 Hz. Pour chaque bande de fréquence, l'énergie fréquentielle est calculée.

TABLE IV.6 – Descripteurs issus des signaux des capteurs inertiels

Catégorie de descripteurs	Liste des descripteurs
Domaine temporel (Signal temporel d'origine et saccade)	<ul style="list-style-type: none"> • Statistiques descriptives : moyenne, écart-type, variance, médiane, médiane absolue des écarts, maximum, minimum, écart Inter-quartile • Autre : énergie temporelle, coefficient de corrélation de Pearson (entre chaque axe)
Domaine fréquentiel (Signal fréquentiel d'origine et saccade en fréquentiel)	<ul style="list-style-type: none"> • Statistique descriptive : moyenne, écart-type, variance, médiane, médiane absolue des écarts, maximum, minimum, écart Inter-quartile, Kurtosis, asymétrie • Autres : énergie fréquentielle, entropie, coefficient de corrélation de Pearson (entre chaque axe)
Décomposition en ondelettes	<p>Énergie fréquentielle, sur 5 bandes de fréquence :</p> <ul style="list-style-type: none"> • bande 1 : 10-20 Hz • bande 2 : 5-10 Hz • bande 3 : 2,5-5Hz • bande 4 : 1,25-2,5 Hz • bande 5 : 0,3-1,25 Hz

IV.6.5.3 Descripteurs des évènements clavier

Des descripteurs particuliers sont calculés à partir des évènements claviers. Chaque évènement clavier est horodaté à la milliseconde près. Surjya Ghosh *et al.* (2017a) définissent l'*Inter-Tap Distance* comme étant l'intervalle de temps entre deux évènements clavier consécutifs. Cette mesure permet de calculer un ensemble de descripteurs, permettant de résumer chaque échantillon de 10 secondes sur la phase de description. Les descripteurs suivants sont calculés pour cette modalité :

- Le nombre de caractères dans l'échantillon.
- La moyenne de l'*Inter-Tap*.
- L'écart-type de l'*Inter-Tap*.
- Le maximum de l'*Inter-Tap*.
- Le minimum de l'*Inter-Tap*.

IV.6.5.4 Descripteurs des capteurs physiologiques

De la même manière, les données des capteurs physiologiques sont synthétisées en un ensemble de descripteurs pour des échantillons de 10 secondes. Chaque enregistrement de capteur nécessite un traitement particulier des données pertinentes, au vu des différentes caractéristiques de chaque signal. Comme pour le prétraitement de signaux, les études de Zhao *et al.* (2018) et Schmidt *et al.* (2018) ont servi de base pour générer les descripteurs appropriés. Nous décrirons donc dans cette sous-section les traitements pour chaque signal.

Conductance électrodermale :

Au préalable, le signal de conductance électrodermale est divisé en deux par l'intermédiaire d'un filtre Butterworth du 3^e degré à la fréquence 0,5Hz (Schmidt *et al.*, 2018) :

- En dessous de 0,5Hz est extrait le niveau de conductance de la peau (ou *Skin Conductance Level*). Il s'agit du niveau basique de réponse physiologique du participant enregistré avec ce capteur.
- Au-dessus de 0,5Hz est extraite la réponse de la conductance de la peau (ou *Skin Conductance Response*). Ce signal est considéré comme sensible aux stimuli externes.

Des descripteurs sur le domaine temporel sont calculés pour chacun des deux signaux.

Pulsation sanguine :

Le signal de pulsation sanguine représente l'activité cardiaque, il est enregistré à une fréquence de 64 Hz, soit une mesure toutes les 15,6 millisecondes. Il s'agit d'un signal quasi périodique. À partir des enregistrements bruts, trois signaux peuvent être calculés sur trois bandes fréquentielles distinctes. Cela permet de multiplier les sources de données pertinentes et générer plus de descripteurs finaux :

- Bande de fréquences basses : entre 0 et 0,7 Hz.
- Bande de fréquences moyennes : entre 0,7 et 2 Hz.
- Bande de fréquences hautes : entre 2 et 10 Hz.

Ainsi, trois bandes de fréquences sont extraites pour chaque échantillon. Des descripteurs sur le domaine fréquentiel sont calculés suite à une application d'une transformée de Fourier pour chaque bande. La liste des descripteurs est disponible sur le tableau IV.7.

Variation de fréquence cardiaque (VFC) :

Quatre signaux sont extraits à partir des enregistrements, répartis sur quatre bandes fréquentielles :

- Bande de fréquences très basses : entre 0,003 et 0,04 Hz.
- Bande de fréquences basses : entre 0,04 et 0,15 Hz.
- Bande de fréquences hautes : entre 0,15 et 0,4 Hz.
- Bande de fréquences très hautes : entre 0,4 et 1 Hz.

Des descripteurs temporels et fréquentiels sont calculés pour chaque signal.

L'ensemble des descripteurs de chaque signal physiologique est décrit dans le tableau IV.7. De plus, l'annexe G. liste les formules mathématiques des descripteurs.

TABLE IV.7 – Descripteurs issus des mesures physiologiques

Signal	Liste des descripteurs
Conductance électrodermale : pour chaque bande fréquentielle	Moyenne, écart-type, maximum, minimum
Pulsation Sanguine : pour chaque bande fréquentielle	Sur le domaine fréquentiel : moyenne de la puissance, minimum de la puissance, fréquence de la valeur maximum, entropie spectrale
Variation de fréquence cardiaque : pour chaque bande fréquentielle	<ul style="list-style-type: none"> • Sur le domaine temporel : moyenne, écart-type, maximum, minimum • Sur le domaine fréquentiel : moyenne de la puissance, minimum de la puissance, fréquence de la valeur maximum, entropie spectrale

IV.6.5.5 Fusion de données multimodale

L'ensemble des matrices de données générées peut être fusionné, afin d'élargir le nombre de descripteurs que les modèles d'apprentissage automatique pourront prendre en compte. Différentes approches ont été présentées dans la sous-section III.3.7. Pour notre cas de figure, nous avons implémenté une fusion au niveau des descripteurs. Les matrices de descripteurs de chaque source de données peuvent donc être assemblées pour former une matrice plus large. Chaque ligne correspond à un échantillon de 10 secondes caractérisé par un identifiant et correspondant au même horodatage, afin que les descripteurs correspondent à la même période temporelle.

Les matrices suivantes ont été générées :

- Une matrice avec les données exclusives des capteurs inertiels du smartphone : accéléromètre, gyromètre et magnétomètre.

- Une matrice avec les données exclusives des capteurs inertiels du smartphone et des événements claviers : accéléromètre, gyromètre, magnétomètre.
- Une matrice avec les données exclusives de la montre Empatica E4 : accéléromètre de la montre, CE, PS et VCF.
- Une matrice avec toutes les modalités.

Ces matrices sont associées à des labels issus de la *Geneva Emotion Wheel*, afin d'entraîner les modèles d'apprentissage automatique à classifier les émotions exprimées par les participants en fonction des descripteurs.

IV.6.6 Mise en œuvre de l'apprentissage automatique

Nous pouvons proposer cette définition générale décrivant ce qu'est l'apprentissage automatique (ou *machine learning* en anglais) :

« Machine Learning is the field of study that gives computers the ability to learn without being explicitly programmed » (Samuel, 1959)

De manière plus opérationnelle, il s'agit d'utiliser des modèles mathématiques et statistiques pour apprendre à distinguer des règles ou patterns, à partir d'un certain nombre d'exemples (Géron, 2019). La spécificité étant que les modèles doivent extraire par eux-mêmes les règles pertinentes, elles ne sont pas fournies en amont. En ce sens, l'apprentissage est donc automatique. Cette approche est particulièrement appropriée lorsque la quantité ou le format des données est trop complexe pour être étudié manuellement. En ce qui concerne les tâches, il peut s'agir de prédire l'appartenance d'une instance à un groupe ou un autre, à partir d'une liste de descripteurs. Par exemple, distinguer si une fleur appartient à une espèce ou à une autre en fonction de la taille de ses pétales. Ces modèles prennent comme référence une performance à améliorer. Ils doivent donc être en mesure d'apprendre par itération en trouvant les paramètres optimaux pour maximiser leur performance.

Plusieurs tâches existent en *machine learning*, nous pouvons en citer deux principales, de manière non exhaustive (Géron, 2019) :

- La classification supervisée : l'objectif est de déterminer si une instance appartient à une certaine catégorie, en prenant comme données en entrée des descripteurs. La classification est dite supervisée, car on donne explicitement le label de chaque instance, pour aider les modèles à les distinguer.
- La classification non supervisée : dans cette situation, les instances n'ont pas de labels explicites. Les modèles ont donc pour objectif de créer des groupes d'instances (ou *cluster*) pour rassembler les instances qui ont le plus de similitudes en fonction de leurs descripteurs.

Des modèles spécifiques doivent être utilisés pour chacune de ces tâches. Dans notre contexte, nous utilisons de la classification supervisée. En effet, la base de données assemblée dans le cadre de l'expérimentation est composée de données labellisées par l'intermédiaire de la GEW ou des conditions expérimentales. Les descripteurs sont les données qui vont permettre d'entraîner les modèles. En apprentissage automatique :

- Les descripteurs sont nommés : **variables prédictives** ou variables indépendantes.
- Les labels correspondent aux classes ou catégories que nous voulons retrouver automatiquement, nous les nommons : **variables à prédire** ou variables dépendantes.

L'objectif pour les modèles est donc d'apprendre à déterminer à partir des variables prédictives et pour chaque instance, quelle est la variable dépendante associée. Dans notre situation, les instances sont les échantillons extraits des enregistrements.

L'ensemble de la démarche pour réaliser de la classification supervisée suit les étapes suivantes, détaillées dans la figure [IV.20](#) :

- Le nettoyage et le prétraitement des descripteurs.
- La sélection des descripteurs les plus informatifs.
- L'apprentissage des modèles de classification.
- L'évaluation des modèles.

Globalement, pour entraîner et évaluer la performance des modèles, il est nécessaire de séparer les descripteurs en deux jeux de données distincts : un jeu de données d'apprentissage et un jeu de données d'évaluation. L'objectif est de ne pas se servir des mêmes données pour apprendre et pour tester, afin de ne pas biaiser les résultats des performances de classification. En effet, l'objectif est de créer des modèles en mesure de généraliser des prédictions sur de nouvelles données. Il n'est donc pas envisageable d'évaluer un modèle à partir du jeu de données d'apprentissage, car cela peut provoquer un sur-apprentissage, à savoir une trop forte adaptation à une collection de données limitée et donner une évaluation erronée de la robustesse du modèle.

Pour répondre à nos hypothèses, deux types de modèles sont réalisés :

- Des modèles généralistes ou inter-individuels.
- Des modèles personnalisés ou intra-individuels.

Pour chaque type de modèle, nous devons donc concevoir des jeux de données d'apprentissage et d'évaluation adaptés. L'approche la plus robuste pour des données extraites à partir de séries temporelles est d'utiliser la méthode *Leave One Out* (LOO) ou validation croisée un contre tous (Dehghani *et al.*, 2019). En effet, les échantillons

issus de signaux temporels ne peuvent pas être considérés comme intégralement indépendants. Des enregistrements proches dans le temps sont susceptibles d'être corrélés et donc d'augmenter artificiellement les performances d'un modèle de classification si les échantillons sont seulement aléatoirement répartis entre les jeux de données. La méthode LOO consiste à diviser les données en k blocs, afin de réaliser l'apprentissage sur $k-1$ blocs et de tester sur le bloc restant. Dans le contexte d'un modèle inter-individuel, les données de tous les participants sauf un, sont utilisées pour l'entraînement du modèle. Les données du participant restant servent à tester la robustesse du modèle. Ainsi, l'évaluation est réalisée sur des données totalement nouvelles et indépendantes pour éviter les effets de surapprentissage. L'apprentissage et l'évaluation sont réalisés de manière à ce que les échantillons de chaque participant soient inclus dans le jeu de données d'évaluation. Il y a donc k itérations. Pour chaque itération, il y a un jeu de données d'apprentissage et un jeu de données d'évaluation. Les performances des modèles sont calculées pour chaque itération et sont moyennées pour obtenir une performance globale.

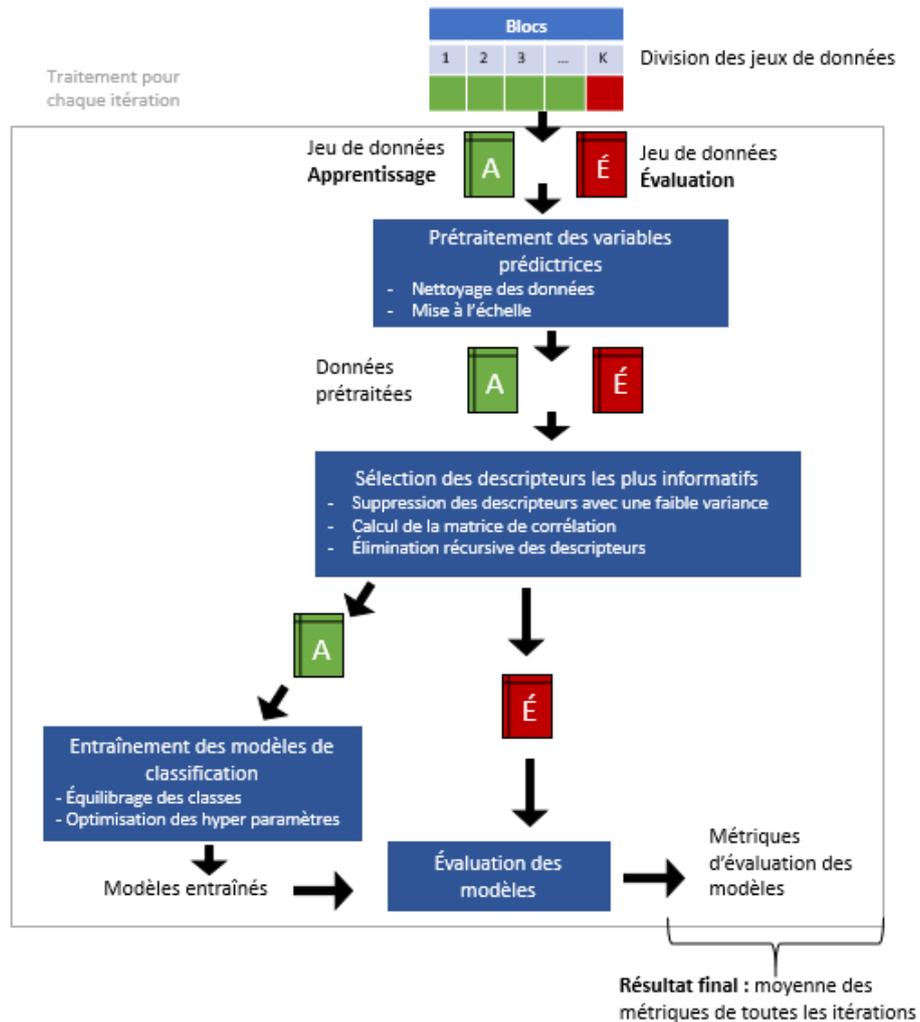


FIGURE IV.20 – Démarche générale pour l'apprentissage des modèles de classification

Dans le cas des modèles intra-individuels, une méthode de validation croisée similaire est réalisée. Cependant, les k blocs correspondent aux tâches d'un même participant. Une partie des tâches est utilisée en apprentissage et une autre pour l'évaluation, afin d'éviter d'utiliser des données dépendantes entre elles. Ainsi, les échantillons d'une même tâche ne pourront pas servir à la fois pour l'entraînement et pour l'évaluation. Plusieurs itérations sont réalisées, afin d'utiliser un maximum de tâches pour tester le modèle et extraire des métriques moyennes sur les performances des modèles.

L'ensemble du code développé a été réalisé en Python, avec l'aide notamment de la bibliothèque *Scikit-learn*, une bibliothèque en langage Python dédiée au prétraitement des données et à la mise en œuvre d'apprentissages automatiques (Pedregosa *et al.*, 2011). Les scripts développés permettent de réaliser automatiquement les étapes

représentées sur la figure IV.20.

IV.6.6.1 Prétraitement des descripteurs

L'ensemble des traitements présentés dans les sous-sections précédentes ont montré comment générer des matrices de descripteurs à partir des enregistrements bruts, des capteurs du smartphone ou de la montre Empatica E4. Chaque ligne correspond à un échantillon de 10 secondes et chaque colonne à des descripteurs ou à des labels d'émotions associés, qui permettront de procéder à la classification supervisée. Pour chaque itération d'un apprentissage automatique, la matrice de descripteurs est divisée en un jeu de données d'apprentissage et un jeu de données d'évaluation. Cette étape est réalisée en amont du prétraitement, afin que les échantillons prévus pour l'évaluation n'aient pas d'effet sur le paramétrage.

Nettoyage des données : en l'état, ces descripteurs peuvent être incomplets et peuvent être sur des échelles numériques très variables. Les descripteurs des deux jeux de données doivent ainsi être nettoyés pour permettre l'apprentissage. Les traitements suivants sont ainsi appliqués (Géron, 2019) :

- Si toutes les lignes d'un descripteur sont composées de valeurs nulles, la colonne est supprimée.
- Dans le cas où un descripteur présente quelques lignes avec des valeurs nulles ou extrêmes (comme une valeur infinie), ces valeurs sont remplacées par la médiane statistique des données restantes.

Il est nécessaire que les jeux de données d'apprentissage et d'évaluation aient un nombre de descripteurs correspondants et les mêmes médianes attribuées. Les prétraitements sont donc, dans un premier temps réalisés sur le jeu de données d'apprentissage, qui est de plus grande taille, puis appliqués avec les mêmes paramètres au jeu de données d'évaluation.

Mise à l'échelle : les valeurs numériques des descripteurs ne sont pas à la même échelle numérique, ce qui n'est pas adapté pour l'entraînement des jeux de données. En effet, un descripteur avec des valeurs entre 0 et 100 pourrait avoir un poids différent d'un descripteur entre 0 et 1, selon les modèles statistiques ou mathématiques employés. La solution est donc de redimensionner l'échelle des données pour les mettre entre 0 et 1. La normalisation peut être réalisée avec la formule suivante :

$$z_i = \frac{x_i - \max(x_i)}{\max(x_i) - \min(x_i)} \times (\max(x_i) - \min(x_i)) + \min(x_i)$$

Les données nettoyées et normalisées sont désormais utilisables pour des modèles d'apprentissage automatique.

IV.6.6.2 Sélection des descripteurs

Selon les modalités sélectionnées, le nombre de descripteurs peut atteindre une valeur très élevée. En effet, pour le seul signal d'accéléromètre la somme des descripteurs temporels, fréquentiels et d'ondelettes calculés pour chacun des trois axes et signaux générés amène à plus de 100 descripteurs. Cependant, il n'est pas nécessaire de conserver autant de descripteurs. D'une part, certaines variables peuvent être suffisamment corrélées avec d'autres et ne pas apporter d'informations supplémentaires. D'autre part, des variables peuvent n'apporter aucune information pertinente pour les modèles et donc représenter du bruit. Réduire le nombre de descripteurs, pour ne garder que les plus informatifs, est donc pertinent pour simplifier l'apprentissage et éviter une trop forte complexité des modèles. Nous appliquons les opérations suivantes :

- Selon la méthodologie proposée par Ruensuk *et al.* (2020), nous calculons dans un premier temps la variance de chaque descripteur, afin de supprimer les variables qui présentent une variance faible. L'objectif est de supprimer des variables qui n'évoluent pas suffisamment au cours du temps et qui n'apportent, a priori pas d'information, car elles n'ont pas évolué au cours de l'expérimentation. Les descripteurs avec une variance inférieure à 0,1 sont donc retirés.
- Nous générons une matrice de corrélation, en calculant les corrélations entre chaque variable prédictive d'une part et entre les variables prédictives et la variable à prédire d'autre part. Si deux variables ont une corrélation supérieure à 80%, nous considérons que ces variables sont trop similaires pour apporter de l'information. Nous ne conservons alors que la variable la plus corrélée avec la variable à prédire. La corrélation entre les descripteurs et la variable à prédire est enregistrée.
- Enfin, nous appliquons un algorithme destiné à ne conserver que les variables les plus pertinentes pour les modèles de classification : le *Recursive Features Elimination* (RFE) ou élimination récursive des descripteurs. La fonction de la bibliothèque *Scikit-learn* utilise un algorithme développé par Guyon *et al.* (2002). L'algorithme utilise un modèle de classification comme une régression logistique ou des machines à vecteurs de support, pour déterminer les descripteurs les plus utiles. Plusieurs itérations sont réalisées jusqu'à déterminer la liste des descripteurs la plus adaptée pour classer la variable à prédire. La fonction prend en paramètre un nombre ou une proportion de descripteurs. Nous appliquons une RFE avec une régression logistique, avec l'objectif de sélectionner la moitié des descripteurs encore présents. Le nombre de descripteurs sélectionnés est donc variable en fonction des modalités choisies en entrée.

Ces opérations sont appliquées sur le jeu de données d'apprentissage. Une fois la liste des descripteurs déterminée, nous sélectionons les mêmes descripteurs dans le jeu de données d'évaluation.

IV.6.6.3 Apprentissage des modèles de classification

Lors de cette étape, les modèles vont être entraînés avec le jeu de données d'apprentissage. L'objectif pour les modèles est d'apprendre à distinguer pour chaque échantillon son appartenance à l'une des classes de la variable à prédire. Pour tester notre démarche, nous avons d'abord entraîné les modèles avec une tâche plus simple : distinguer si le signal provient d'une tâche de visionnage ou d'une tâche de description, lors de l'expérimentation principale. Les classes de la variable à prédire sont donc la phase de visionnage ou de description. Nous réaliserons un modèle inter-personnel, à partir des données des 40 participants.

Il faut au préalable prendre en compte que le nombre d'échantillons est variable en fonction de la classe. En effet, la durée des tâches de description et de visionnage n'est pas la même. Les tâches de visionnage durent environ 30 secondes et celles de description peuvent durer plusieurs minutes. Il y a donc plus d'échantillons provenant des tâches de description (13761 échantillons) par rapport aux tâches de visionnage (4851 échantillons). Le déséquilibre de taille des classes est un problème récurrent en apprentissage automatique. En conséquence, les modèles de classification peuvent être biaisés par le manque d'informations d'une classe minoritaire. Il peut alors être plus difficile de prédire l'appartenance d'une instance à une classe minoritaire, ce qui limite sa capacité à généraliser ses prédictions et donc sa robustesse.

Une solution efficace consiste à utiliser une technique de suréchantillonnage, comme le SMOTE (*Synthetic Minority Oversampling Technique*) (Chawla *et al.*, 2002). Cette méthode vise à augmenter artificiellement le nombre d'échantillons dans la classe minoritaire, afin d'obtenir un jeu de données avec un nombre d'échantillons identique pour chaque classe. L'algorithme va créer de nouvelles instances en fusionnant un nombre n d'instances qui ont des valeurs proches. Le SMOTE n'est appliqué que sur le jeu de données d'apprentissage. Le jeu de données d'évaluation n'est pas modifié et conserve son déséquilibre de classe.

Le jeu de données d'apprentissage rééquilibré peut ensuite servir à entraîner des modèles de classification. Nous avons sélectionné huit modèles de classification couramment utilisés pour la classification de ce type de données (Cui *et al.*, 2015 ; Kołakowska *et al.*, 2020 ; Mottelson & Hornbæk, 2016 ; Olsen & Torresen, 2016 ; Ruensuk *et al.*, 2020). De plus, nous ajoutons un modèle de classification de base, qui ne classe que les instances d'une classe de référence donnée (ici la phase de description). Ce modèle est utile pour comparer la performance des autres modèles. La liste des modèles de classifi-

cation est la suivante, nous les avons implémentés par l'intermédiaire de la bibliothèque *Scikit-learn* :

- Adaboost (Hastie *et al.*, 2009).
- Arbres de décision (Breiman *et al.*, 1984).
- Classifieur à renforcement de gradient (Friedman, 2001).
- K plus proches voisins (Altman, 1992).
- Machine à vecteur de support (*Support Vector Machine SVM*) (Guo *et al.*, 2016).
- Réseaux de neurones : perceptron multicouche (Glorot & Bengio, 2010).
- Régression logistique (Fan *et al.*, 2008).
- Réseaux bayésiens (Rish *et al.*, 2001).

Nous avons sélectionné l'ensemble de ces modèles avec l'objectif d'obtenir un état des lieux global des performances de classification. La comparaison des résultats de classification de chaque modèle permet de déterminer lesquels sont les plus robustes sur ce type de jeux de données. La bibliothèque *Scikit-learn* se charge de trouver les meilleurs paramètres pour maximiser leur robustesse.

IV.6.6.4 Évaluation des modèles de classification

À la suite de l'apprentissage, les modèles sont testés sur le jeu de données d'évaluation. Les résultats, à savoir les classes prédites par les modèles, sont comparés avec les classes réelles pour chaque échantillon d'évaluation. Dans le cas d'une classification binaire, la démarche est de déterminer si l'instance ou l'échantillon a été prédit dans la bonne classe. Dans l'exemple de la classification des phases de description (*classe +*) ou de visionnage (*classe -*), la phase de description est notre classe de référence. Il y a donc quatre cas de figure :

- L'échantillon appartient dans la réalité à une phase de description (*classe +*) et a été prédit comme tel (*classe +*) : il s'agit donc d'un **vrai positif** (VP).
- L'échantillon appartient dans la réalité à une phase de description (*classe +*), mais n'a pas été prédit comme tel (*classe -*) : il s'agit donc d'un **faux négatif** (FN).
- L'échantillon appartient dans la réalité à une phase de visionnage (*classe -*) et a été prédit comme tel (*classe -*) : il s'agit donc d'un **vrai négatif** (VN).
- L'échantillon appartient dans la réalité à une phase de visionnage (*classe -*), mais n'a pas été prédit comme tel (*classe +*) : il s'agit donc d'un **faux positif** (FP).

TABLE IV.8 – Matrice de confusion

		Classes prédites	
		+	-
Classes réelles	+	Vrai Positif (VP)	Faux Négatif (FN)
	-	Faux Positif (FP)	Vrai Négatif (VN)

L'ensemble de ces catégories permet de constituer une matrice de confusion pour chaque modèle. Les VP et VN sont les échantillons bien classés, qui démontrent la robustesse du modèle. Le tableau IV.8 représente les différents cas de figure des résultats d'une classification binaire. Ces résultats permettent de calculer des métriques d'évaluation qui résument l'ensemble des performances du modèle. Les métriques et les formules sont détaillées sur le tableau IV.10.

TABLE IV.9 – Métriques d'évaluation

Nom des métriques	Formules
Exactitude	$Exactitude = \frac{VP + VN}{VP + FN + FP + VN}$
Précision	$Precision = \frac{VP}{VP + FP}$
Rappel	$Rappel = \frac{VP}{VP + FN}$
F1-score	$F1 = 2 \times \frac{Precision \times Rappel}{Precision + Rappel}$

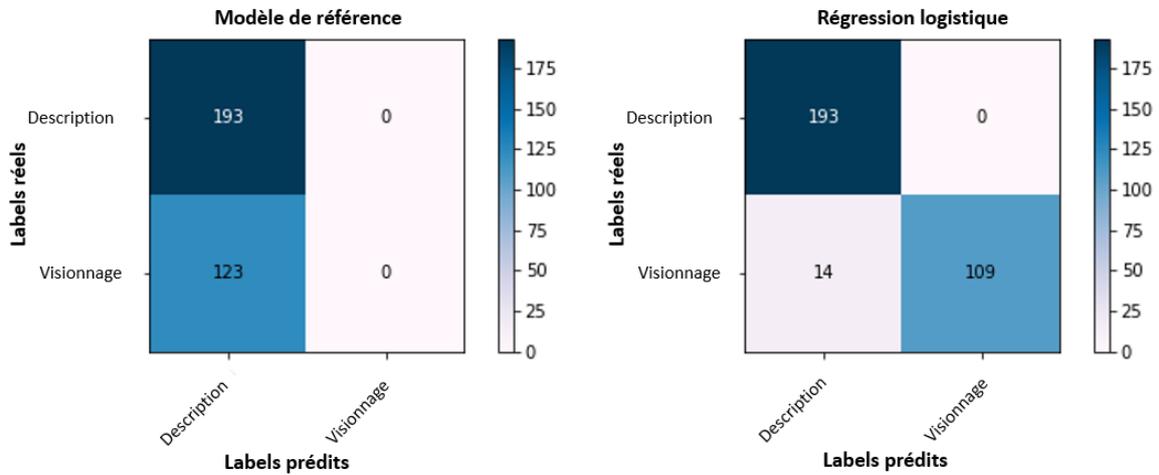


FIGURE IV.21 – Matrices de confusion sur la classification de phase (*watching* = visionnage). La matrice de gauche est pour le modèle de référence et le modèle de droite un modèle de régression logistique

La figure IV.21 représente deux matrices de confusion pour une itération d'apprentissage automatique. Les modèles ont été entraînés sur 39 participants et testés sur le participant restant. La matrice de confusion de gauche représente le modèle de référence qui ne classe les instances que dans la classe de référence, à savoir la phase de description (*classe +*). Nous pouvons observer le déséquilibre entre les classes sur le jeu de données d'évaluation, constitué de 193 échantillons de phase de description et de 123 échantillons de phase de visionnage. Il est pertinent de conserver le déséquilibre entre les classes pour le jeu de données d'évaluation pour garder ses caractéristiques d'origine. La robustesse d'un modèle sera évaluée en comparaison avec le modèle de référence. La matrice de confusion de droite représente les résultats obtenus suite à l'apprentissage d'un modèle de régression logistique. Nous pouvons observer une forte proportion de VP et VN, indiquant que la majorité des prédictions ont été correctes, malgré quelques erreurs. Nous pouvons, à partir de cette matrice, calculer des métriques d'évaluation :

L'exactitude (ou *accuracy* en anglais) est une des métriques les plus couramment utilisées en apprentissage automatique. Elle donne une idée générale de la performance de classification en donnant le taux d'instances correctement classées sur le total des instances classées. Nous appliquons la formule d'exactitude sur les deux matrices de confusion.

$$exactitude_{reference} = \frac{193 + 0}{193 + 0 + 123 + 0} \times 100 = 61,07$$

$$exactitude_{reglog} = \frac{193 + 109}{193 + 0 + 14 + 109} \times 100 = 95,55$$

Ainsi, nous obtenons le pourcentage d'exactitude du modèle de référence et du modèle de régression logistique. La régression logistique surpasse le modèle de référence de 34,48%. Ce modèle ne classe donc pas les échantillons au hasard, nous pouvons conclure que la généralisation a fonctionné pour cette itération.

La précision est le nombre d'instances bien classées de la classe de référence (*classe +*), parmi toutes les instances prédites dans cette classe (*classe +*). Elle permet de répondre à la question suivante : quelle proportion des instances prédites comme étant dans la *classe +* sont effectivement dans la *classe +*. Dans notre exemple, il s'agit de la proportion d'échantillons qui ont été correctement prédits comme étant dans la phase de description, parmi tous les échantillons qui ont été prédits comme tels.

$$precision_{reglog} = \frac{193}{193 + 14} \times 100 = 93,23$$

Ainsi, lorsque le modèle de régression logistique prédit que l'échantillon est dans une tâche de description, il est correct à 93,64%. Cela indique que le modèle a une légère tendance à considérer des échantillons de visionnage comme étant dans la classe de description.

Le rappel est le nombre d'instances bien classées de la classe de référence (*classe +*), parmi toutes les instances réellement dans cette classe (*classe +*). Il répond à la question suivante : quelle proportion des échantillons de la *classe +* dans la réalité est correctement classée ? Dans l'exemple, combien d'échantillons de description sont réellement classés comme tels.

$$rappel_{reglog} = \frac{193}{193 + 0} \times 100 = 100$$

Le modèle de régression logistique ne fait donc aucune erreur lorsqu'un échantillon fait partie de la phase de description dans la réalité. La précision et le rappel sont des métriques complémentaires à l'exactitude. Dans l'exemple donné, le rappel est meilleur que la précision, ce qui signifie que le modèle est plus performant pour bien classer des échantillons de phase de description que des échantillons de phase de visionnage.

Le score F1 est une moyenne harmonique des métriques de précision et de rappel. Cette métrique permet de rendre compte de l'équilibre global du modèle.

$$scoreF1_{reglog} = 2 \times \frac{1 \times 0,93}{1 + 0,93} \times 100 = 96,37$$

Les métriques sont calculées pour chaque itération. Pour avoir une idée plus large de la robustesse des modèles, nous pouvons calculer la moyenne pour les 40 itérations réalisées (voir tableau IV.10). Nous pouvons ainsi observer l'exactitude moyenne de chaque modèle, pour l'ensemble des itérations de la validation croisée un contre tous.

Pour chaque modèle, l'exactitude obtenue est supérieure au modèle de référence, ce qui montre que les modèles parviennent à classifier avec des performances supérieures au hasard. Les modèles utilisant des machines à vecteur de support et des réseaux neuronaux sont les plus performants avec une exactitude supérieure à plus de 26% au modèle de référence. Le modèle de réseaux bayésiens est le moins performant avec une différence d'exactitude inférieure à 20%.

TABLE IV.10 – Moyenne d'exactitude pour une classification de phase de description vs visionnage

Modèles	Exactitude moyenne	Différence avec la référence
Modèle de référence	69,72 %	0 %
AdaBoost	94,99 %	25,27 %
Arbre de décision	93,18 %	23,47 %
Classeur à renforcement de gradient	95,67 %	25,69 %
K plus proches voisins	93,19 %	23,47 %
Machines à vecteur de support	95,93 %	26,21 %
Réseaux de neurones	95,79 %	26,07 %
Régression logistique	94,27 %	24,55 %
Réseaux bayésiens	85,45 %	15,73 %

L'évaluation des modèles est la dernière étape du système de classification automatique. Les évaluations sont enregistrées pour chaque itération avec une matrice de confusion. Les paramètres de chaque modèle sont aussi enregistrés, ainsi que les descripteurs les plus corrélés à la variable à prédire.

IV.6.7 Discussion

Nous avons décrit l'ensemble des processus pour exploiter les données issues des capteurs de smartphone et des capteurs physiologiques, caractérisées par leur dimension temporelle. L'ensemble du code développé en Python pour traiter ces données de l'état brut jusqu'à l'apprentissage et l'évaluation des modèles de classification constitue un apport majeur de ce travail de recherche. Ainsi, le code a été conçu pour être flexible, afin de prendre en compte de nouvelles modalités, développer de nouveaux calculs de descripteurs ou modifier les paramètres de l'échantillonnage. L'apprentissage réalisé pour distinguer les phases de description et de visionnage montrent que l'ensemble de la démarche fonctionne et permet d'obtenir des modèles de classification robustes.

Pour tenter de maximiser les chances de trouver des éléments pour discriminer des états émotionnels, un effort particulier a été apporté pour générer un maximum de descripteurs. Cependant, le fait de générer de nombreux descripteurs rend l'étape d'échantillonnage particulièrement longue, dû moins sur une machine personnelle avec des performances intermédiaires. Ainsi, plusieurs heures sont nécessaires pour générer l'ensemble des matrices pour 40 participants. Il est souhaitable de limiter le nombre de descripteurs aux plus essentiels, à condition de pouvoir les identifier dans un premier temps. Cela est d'autant plus nécessaire dans l'optique de créer des bases de données avec plus de sujets.

Au vu de la complexité des traitements, de nombreux paramètres peuvent influencer sur les résultats finaux. La taille de la fenêtre d'échantillonnage, par exemple, constitue un paramètre essentiel. Le choix d'une fenêtre d'échantillonnage large a été réalisé, afin d'assurer la prise en compte de patterns longs, la superposition permettant de détailler au mieux les enregistrements et d'éviter de rater des patterns spécifiques. Cui *et al.* (2015) ont, par exemple, mis en évidence l'importance de la taille d'échantillonnage pour la reconnaissance d'émotions avec des données d'accéléromètre pour des activités de marche. Dans le cas d'interaction avec un smartphone, il serait alors pertinent de créer plusieurs matrices avec des fenêtres d'échantillon différentes, afin de déterminer une taille optimisée pour ce type de données. Cela nécessite en revanche des temps de calcul relativement longs.

Pour la fusion de données, nous avons opté pour une fusion au niveau des descripteurs, en raison de sa simplicité, de son économie de ressources et de la nature de signaux en série temporelle. Il est envisageable d'intégrer d'autres approches, comme une approche au niveau décisionnel ou hybride, avec des modèles de classification par modalité (voir III.3.7). Une couche supplémentaire devrait alors être intégrée pour agréger les décisions des modèles et proposer une réponse définitive. L'approche hybride serait appropriée pour associer les décisions provenant de modalités, telle que des données textuelles dans le cadre d'un échange sur une messagerie instantanée, analysé par du TAL.

Sur la démarche générale, un corpus de données regroupant une plus grande quantité de données pourrait être exploité par apprentissage profond. Les étapes de génération de descripteurs sont ainsi supprimées, car les modèles d'apprentissage profond génèrent eux-mêmes leurs descripteurs. Pour ce type de données, cette démarche est utilisée pour la reconnaissance d'activité sur smartphone, une discipline plus mature et avec des bases de données plus anciennes et étendues, que sur la reconnaissance d'émotions. De plus, les recherches en informatique affective utilisent largement l'apprentissage profond pour la reconnaissance d'émotions par des images d'expression faciale ou pour les enregistrements de voix. L'application de la démarche par apprentissage profond constitue donc une piste pertinente, notamment avec des modèles spécialisés dans les séries temporelles tels que les réseaux de neurones

récurrents.

Enfin, la labellisation des données représente un élément essentiel pour la classification. Au vu de la complexité de cette tâche, il peut être pertinent d'utiliser des modèles d'apprentissage non supervisé afin de rassembler des groupes de données similaires. La démarche peut être hybride, en générant des labels à partir d'apprentissage non supervisé puis en entraînant des modèles. Une approche testée, par exemple, par Wampfler *et al.* (2020) pour de la classification de données de frappe au clavier.

En conclusion, l'analyse de données issues des capteurs se définit par un ensemble de phases de traitement à mettre en œuvre, en amont de la classification. Chacune des étapes peut avoir un effet significatif sur le résultat final. Nous avons donc listé dans cette discussion d'autres approches envisageables pour traiter le corpus de données. L'implémentation que nous proposons répond à nos besoins en traitant les données pour aboutir à la création de modèles de classification par apprentissage automatique. L'ensemble du code développé a été conçu pour être adaptable et évolutif et est une contribution majeure de ce travail de recherche. Nous pouvons ainsi l'utiliser pour créer et évaluer des modèles de classification automatique d'émotions.

IV.7 Conclusion

Dans le contexte de notre projet de recherche, des outils spécifiques ont été développés pour répondre à notre besoin de recueil de données. Premièrement, l'application d'expérimentation et ses diverses variantes ont permis à la fois de créer un contexte d'expérimentation de messagerie instantanée et d'intégrer plusieurs types de protocoles d'induction d'émotions. Les outils se sont avérés globalement polyvalents et flexibles. L'ensemble des expérimentations a permis d'améliorer ces outils et même de rendre certains d'entre eux disponibles en *open source* comme pour l'application GEWMI. De la même manière, nous pouvons envisager de rendre l'application d'expérimentation *open source* pour faciliter le recueil de données dans le contexte de messagerie instantanée sur smartphone.

L'induction d'émotions est une tâche complexe. Les smartphones étant un contexte récent d'expérimentation pour l'informatique affective, il est nécessaire d'adapter les protocoles d'induction d'émotions existants. Cependant, les smartphones offrent aussi un large panel de possibilités pour expérimenter. D'une part, de nombreux médias et activités peuvent être intégrés et d'autre part, avec le format d'application mobile, il est envisageable de diffuser largement certaines procédures expérimentales pour toucher un plus large public. Le protocole de manipulation de la difficulté montre qu'il est possible d'innover pour proposer des méthodes engageantes pour provoquer des émotions. En approfondissant cette approche, il est concevable d'utiliser, par exemple,

des jeux mobiles pour recueillir des données d'expressions motrices labellisées sur des émotions. En revanche, il faut garder à l'esprit que les habitudes d'usage de smartphones ne sont pas égales dans toutes les catégories de population.

Au cours de ce travail de recherche, les expérimentations ont principalement porté sur des études en laboratoire. Une approche complémentaire par des études de terrain afin de recueillir des données en situation écologique est donc pertinente à réaliser. En situation d'utilisation d'application de conversation instantanée, les données peuvent être labellisées par l'intermédiaire des verbatims. La création d'un corpus est faisable, à condition de pouvoir enregistrer les données des capteurs chez des utilisateurs volontaires. Nous pouvons nous poser la question des effets de l'utilisation d'un agent conversationnel, plus neutre qu'un expérimentateur humain. En effet, l'utilisation d'agents conversationnels permet de réduire la variabilité sur la passation des conditions expérimentales et éventuellement d'autres biais. En revanche, le fait de savoir que l'agent conversationnel n'est pas humain peut limiter les émotions causées par des échanges conversationnels. La multiplication de *chatbots* sur les sites marchands ou dans les services clients montre qu'il est pertinent d'étudier les effets des *chatbots* sur les émotions des utilisateurs.

La conception et le développement d'un système de reconnaissance automatique ont été détaillés. À partir d'enregistrements bruts issus des expérimentations, le système est en mesure de traiter des données et de générer des matrices de descripteurs. Ensuite, le système entraîne et évalue des modèles de classification par apprentissage automatique. L'évaluation prend en compte l'aspect dépendant des données temporelles collectées, en appliquant une validation croisée un contre tous, afin d'éviter des biais dans l'apprentissage et évaluer le plus précisément possible la robustesse des modèles. Des métriques d'évaluation sont enregistrées, ainsi que la corrélation entre les descripteurs et les variables à prédire. En testant l'ensemble du système pour classer des phases d'activités des participants lors de l'expérimentation principale (visionnage contre description), nous obtenons des résultats robustes. Des modalités peuvent être combinées et ajoutées pour générer des matrices de données, selon une logique de fusion au niveau des descripteurs.

Suite aux expérimentations, les données nous permettent d'étudier trois composants de la théorie de l'évaluation cognitive. Premièrement, le sentiment subjectif, mesuré par l'intermédiaire des outils d'auto-évaluation SAM et GEW. Deuxièmement, l'expression motrice, par le biais du corpus constitué des interactions avec les capteurs inertiels et la frappe au clavier lors de l'expérimentation principale. Enfin, les réponses psychophysiologiques peuvent être appréhendées avec les données enregistrées sur la montre Empatica E4 utilisée au cours de l'expérimentation principale. Le chapitre V décrira l'ensemble des résultats obtenus.

Chapitre V

Résultats

Sommaire

V.1	Introduction	169
V.2	Analyse du sentiment subjectif	170
V.2.1	Méthodologie d'analyse	170
V.2.2	Pré-expérimentation	170
V.2.3	Expérimentation principale	177
V.2.4	Expérimentation complémentaire	191
V.2.5	Discussion	196
V.3	Résultats du système de reconnaissance automatique	197
V.3.1	Labellisation des descripteurs	197
V.3.2	Classification de données d'interaction avec un smartphone	198
V.3.3	Classification avec les données de la montre Empatica E4	201
V.3.4	Classification avec fusion des modalités	204
V.3.5	Discussion	206
V.4	Conclusion	209

V.1 Introduction

Le chapitre IV a permis de rendre compte des outils développés, du matériel utilisé et des protocoles expérimentaux établis. Cela a permis de réaliser des expérimentations pour recueillir des données, afin de répondre aux objectifs suivants :

- Mettre en place des protocoles expérimentaux efficaces dans un contexte d'interaction avec un smartphone et avec un format de messagerie instantanée. Ainsi que des mesures pour vérifier les effets du protocole et du matériel utilisé, par l'intermédiaire d'outils d'auto-évaluation.
- Recueillir des données des interactions avec un smartphone labellisées, afin de former un corpus pour entraîner des modèles de classification. Par extension, prendre en compte les enregistrements des capteurs physiologiques de la montre Empatica E4 pour comparer les résultats de modèles de classification.

Ce chapitre a donc pour ambition de répondre aux hypothèses énoncées dans le chapitre précédent.

Dans un premier temps, nous analysons les effets des conditions expérimentales en termes d'induction d'émotions, élément essentiel des études en laboratoire en informatique affective. Pour cela, les mesures obtenues par les auto-évaluations permettent d'estimer de manière quantitative le sentiment subjectif des participants. Des tests statistiques permettent de confirmer ou d'infirmer des hypothèses opérationnelles formulées à partir des hypothèses générales. Ces analyses nous permettront de répondre à la **question fondamentale Q_Induction** (voir section IV.1). En ayant testé plusieurs approches d'induction d'émotions, nous pouvons comparer les effets sur le sentiment subjectif déclaré. Premièrement, de manière synthétique en comparant des dimensions émotionnelles. Deuxièmement, de manière plus approfondie avec les libellés de la GEW.

Dans un second temps, l'ensemble des enregistrements des capteurs de smartphone constituent le corpus pour concevoir un système de reconnaissance automatique d'émotions. Ces enregistrements représentent l'interaction avec l'appareil que nous associons à la composante de l'expression motrice, nous décrivons donc l'entraînement et l'évaluation des modèles de classification pour ces données. Pour répondre à la problématique générale, nous nous intéressons à la mise en œuvre de modèles inter-individuels (**question fondamentale Q_Interindividu**), puis de modèles intra-individuels (**question fondamentale Q_Intraindividu**). Enfin, nous comparons les résultats avec une autre source de données, provenant principalement de capteurs physiologiques (**question Q_Equivalence_Physio**).

V.2 Analyse du sentiment subjectif

V.2.1 Méthodologie d'analyse

Dans cette section, nous expliciterons les analyses pour estimer l'effet des inductions d'émotions réalisées, au niveau de la composante du sentiment subjectif. Ces analyses sont effectuées à partir des résultats obtenus par l'intermédiaire des outils d'auto-évaluation, respectivement le SAM (*Self Assessment Manikin*) pour la pré-expérimentation et la GEW (*Geneva Emotion Wheel*) pour les deux autres expérimentations. En synthétisant les auto-évaluations proposées par les sujets, nous pouvons ainsi comparer les résultats de chaque condition expérimentale. De plus, nous analyserons aussi les verbatims des descriptions réalisées au cours de l'expérimentation principale.

V.2.2 Pré-expérimentation

La pré-expérimentation est marquée par l'utilisation du SAM pour l'auto-évaluation d'émotions. Pour chaque participant, sans compter une auto-évaluation suite à l'entraînement, cinq auto-évaluations ont été enregistrées pour la valence et pour l'activation. Soit 200 auto-évaluations pour chaque dimension. Chaque auto-évaluation se présente sous la forme d'une valeur entre 1 et 5. Les distributions des auto-évaluations des deux dimensions sont représentées en figures V.1 et V.2. Nous pouvons observer que les participants n'ont pas déclaré d'auto-évaluations de valence de niveau 1. Pour la dimension de l'activation, les déclarations ont une forme plus symétrique avec une forte proportion de déclarations de valeur 3.

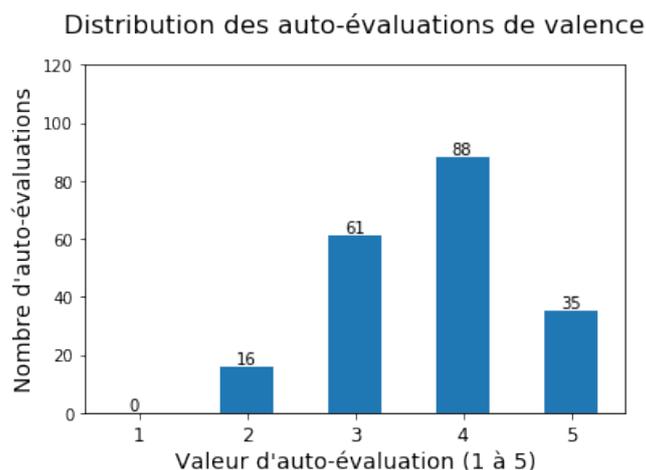


FIGURE V.1 – Distribution des auto-évaluations de valence (1 = valence négative, 5 = valence positive)

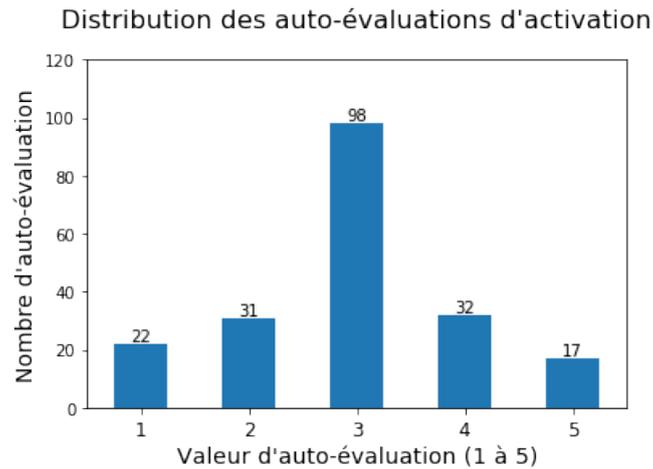


FIGURE V.2 – Distribution des auto-évaluations d'activation (1 = activation faible, 5 = activation forte)

V.2.2.1 Analyse de la valence des auto-évaluations en fonction des extraits vidéo

Nous voulons déterminer, à partir des auto-évaluations de l'outil SAM, si les extraits vidéo ont eu un effet sur la dimension de la valence. L'hypothèse opérationnelle est la suivante :

Hypothèse I.A : Après un bloc de vidéos avec une valence positive, les valences déclarées sur le SAM sont supérieures en moyenne par rapport à un bloc de vidéos de valence neutre. De la même manière, pour un bloc de vidéos neutres, les valences déclarées seront supérieures sur le SAM par rapport à un bloc avec des vidéos de valence négative. (Condition de valence positive > Condition de valence neutre > Condition de valence négative).

Les moyennes d'auto-évaluation par conditions expérimentales (négative, neutre ou positive) correspondant aux blocs d'extraits vidéo, sont représentées sur la figure V.3. Visuellement, les moyennes ne semblent pas être sensiblement différentes entre les conditions, avec au plus bas 3,65 pour la condition négative et 3,79 pour la condition positive.

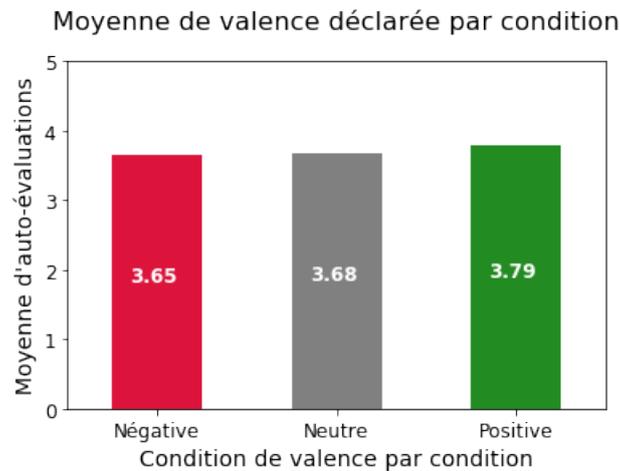


FIGURE V.3 – Moyenne de valence déclarée par condition

Test d'hypothèse : nous cherchons à déterminer si les moyennes des auto-évaluations déclarées pour chaque condition sont significativement différentes, d'un point de vue statistique. L'hypothèse nulle est donc que chaque condition expérimentale possède une moyenne similaire. Nous effectuons un test de normalité pour choisir le test adapté, les échantillons de chaque condition ne suivent pas une loi normale selon un test de Shapiro-Wilk ($p\text{-value} < 0.05$). Nous devons donc utiliser un test d'hypothèse non paramétrique. Le nombre d'échantillons étant de trois, nous utiliserons donc un test de Kruskal-Wallis (Howell *et al.*, 2008). Selon le test de Kruskal-Wallis ($W = 1,347$, $p\text{-value} = 0.509 > 0,05$), nous ne pouvons pas rejeter l'hypothèse nulle. Il n'y a pas de différence significative entre les moyennes de valence déclarée pour les trois conditions.

Nous réalisons une analyse plus précise, afin de comparer les moyennes de chaque condition deux à deux, par l'intermédiaire de tests de Mann-Whitney (voir le tableau V.1). Nous n'observons aucune différence significative pour chacune des comparaisons effectuées. En revanche, nous observons une $p\text{-value}$ plus faible entre les conditions positives et négatives, qui ont le plus d'écart en termes de moyenne. Cependant, l'**hypothèse IA** n'est pas validée, il n'y a pas de différence significative concernant la valence des auto-évaluations entre les conditions expérimentales.

TABLE V.1 – Test post-hoc de Mann-Whitney sur la dimension de la valence

Test	Statistique	P-value
Positif vs Neutre	1468,5	0,217
Négatif vs Neutre	1580	0,454
Positif vs Négatif	2897	0,135

Selon les auto-évaluations des participants, il n'y a donc pas de différence marquée

entre les différentes conditions expérimentales sur la dimension de la valence. Ainsi, il semblerait que l'induction d'émotions par l'intermédiaire d'extraits vidéo sur plusieurs modalités de valence n'ait pas eu d'effet mesurable sur le sentiment subjectif des participants.

V.2.2.2 Analyse de l'activation des auto-évaluations en fonction des extraits vidéo

De la même manière que pour l'analyse précédente, nous voulons estimer si les extraits vidéo ont eu un effet sur la dimension de l'activation. Nous formulons donc l'hypothèse opérationnelle suivante :

Hypothèse I.B : Après un bloc de vidéos avec une activation élevée, les activations déclarées sur le SAM sont en moyenne supérieures par rapport à un bloc d'activation neutre. De la même manière, pour un bloc de vidéos d'activation neutre, les activations déclarées seront supérieures sur le SAM par rapport à un bloc d'activation faible. (Condition d'activation positive > Condition d'activation neutre > Condition d'activation faible).

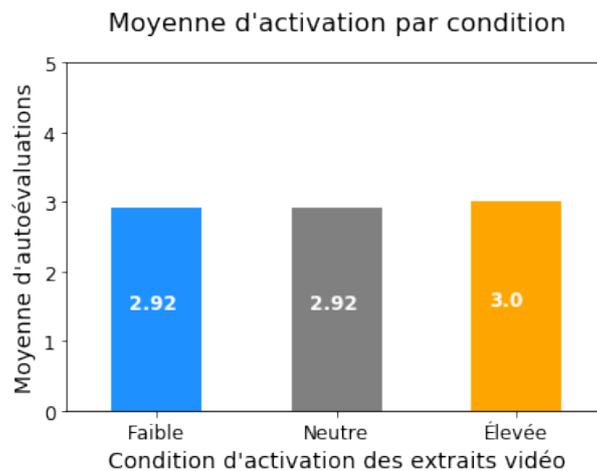


FIGURE V.4 – Moyenne d'activation déclarée par condition

Les moyennes pour chaque condition expérimentale (faible, neutre et élevée) sont représentées en figure V.4. Nous n'observons pas, visuellement, d'écart significatif entre les conditions expérimentales sur la dimension de l'activation. En effet, les conditions d'activation faible et neutre possèdent une moyenne d'activation de 2,92 déclarée sur le SAM et la condition élevée une moyenne de 3,0.

Test d'hypothèse : l'hypothèse nulle est que chaque condition expérimentale possède une moyenne d'auto-évaluation similaire. Un test de Shapiro-Wilk montre

que les échantillons ne suivent pas une loi normale, nous sélectionnons donc un test non paramétrique. Étant donné qu'il y a trois échantillons, nous appliquons le test de Kruskal-Wallis. Les résultats ne nous permettent pas de rejeter l'hypothèse nulle ($W = 0,408$, $p\text{-value} = 0.815 > 0,05$), il n'y a donc pas de différence significative entre les trois conditions.

Des analyses pour comparer les conditions deux à deux nous permettent d'obtenir le tableau V.2. Les résultats des tests statistiques de Mann-Whitney montrent qu'il n'y a aucune différence significative entre chacune des conditions. L'hypothèse I.B n'est donc pas validée, il n'y a pas de différences significatives au niveau de l'activation des auto-évaluations entre les différentes conditions expérimentales.

TABLE V.2 – Tests de Mann-Whitney sur la dimension d'activation

Test	Statistique	P-value
Faible vs Intermédiaire	1535,5	0,351
Élevé vs Intermédiaire	1572,5	0,435
Élevé vs Faible	3031,5	0,270

De la même manière que pour la dimension de la valence, il semblerait que les extraits vidéo n'aient pas eu d'effets mesurables sur la dimension de l'activation chez les sujets.

V.2.2.3 Relation entre les réponses au test et auto-évaluations

Nous voulons estimer si le nombre de réponses correctes par bloc d'extraits vidéo est significativement différent en fonction des niveaux de la valence des auto-évaluations. Les participants pouvaient obtenir entre 0 et 10 bonnes réponses pour chaque bloc de vidéos. L'hypothèse opérationnelle est formulée de la manière suivante :

Hypothèse I.C : Plus le nombre de réponses correctes au test d'attention et de mémoire est élevé et plus les niveaux de valence déclarés sont élevés.

Pour réaliser cette analyse, nous avons divisé les auto-évaluations de valence en deux catégories : les déclarations inférieures ou égales à trois (soit 77 auto-évaluations) et les déclarations strictement supérieures à trois (123 auto-évaluations). Nous pouvons observer en figure V.5, la différence de moyenne de réponses correctes entre les modalités de valence des auto-évaluations. Ainsi, nous pouvons observer que visuellement, lorsque les sujets ont déclaré une valence élevée, ils obtiennent un score plus élevé, avec un score moyen de 3,32, contre 2,64 lorsqu'une valence faible a été déclarée.

Nombre moyen de réponses correctes en fonction de la valence déclarée

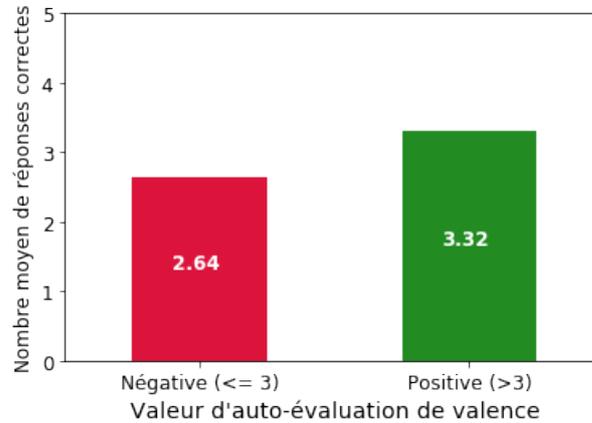


FIGURE V.5 – Nombre moyen de réponses correctes en fonction de l'auto-évaluation de la valence

Test d'hypothèse : nous voulons vérifier s'il y a une différence significative sur les nombres moyens de réponses correctes entre les catégories de valence. Selon un test de Shapiro-Wilk, les échantillons ne suivent pas une loi normale (p -value < 0.05), nous appliquons donc un test d'hypothèse non paramétrique. Suite à un test de Mann-Whitney, car il y a deux groupes d'échantillons, nous obtenons une différence significative entre les deux groupes de valence (statistique = 3726, p -value = 0.00503 < 0.001). Il y a donc significativement plus de réponses correctes lorsque les participants ont déclaré une valence positive suite à un bloc d'extraits vidéo. L'**hypothèse I-C** est donc validée.

En conclusion, nous observons un lien entre le nombre de réponses correctes dans un bloc et les niveaux de valence déclaré par les participants. Une valence déclarée élevée indique un nombre moyen de réponses correctes plus élevé.

V.2.2.4 Discussion

L'hypothèse générale formulée dans la sous-section [IV.3.3](#) n'est donc que partiellement validée. En effet, il n'y a pas d'effet significatif mesurable des extraits vidéo sur les auto-évaluations de valence et d'activation, les hypothèses opérationnelles **I-A** et **I-B** n'étant pas validées. Cependant, les émotions déclarées sur le SAM ont été significativement associées aux performances lors des tests de mémorisation et d'attention, seule l'hypothèse opérationnelle **I-C** est validée.

Le nombre important d'extraits vidéo, ainsi que leur contenu peu réaliste a pu amoindrir l'effet d'induction d'émotions. Susciter des émotions étant une tâche com-

plexe, vouloir influencer sur deux dimensions avec cinq conditions expérimentales au total semble rendre le protocole laborieux. L'effet de l'induction d'émotion par les extraits vidéo a aussi pu être diminué en raison des effets plus importants de l'activité de réponse aux questions sur le contenu. Ainsi, il est possible que les participants se soient faiblement engagés dans le contenu émotionnel des vidéos. De plus, l'interface initiale de l'application avec des vidéos de petite taille et de faible qualité a pu avoir un impact sur l'implication des participants dans le contenu des médias et provoquer de la frustration. En effet, expérimenter sur smartphone présente des contraintes sur le format des médias et des interactions possibles. C'est pourquoi il était nécessaire de pré-expérimenter et d'interroger les participants sur l'ergonomie de l'application d'expérimentation.

Le champ libre laissé à la fin de l'expérimentation a permis d'obtenir des éléments de réponses sur les difficultés d'induction d'émotions avec les vidéos, dans le cadre de cette application. Premièrement, les vidéos, incluses dans les bulles de dialogue de l'agent conversationnel, n'ont pas toutes été jugées suffisamment visibles en raison de la taille et du manque de contraste. De plus, le contenu des vidéos, issues de films en libre accès n'a pas été jugé crédible par certains participants. Il a donc été nécessaire de trouver des extraits vidéo avec des contenus plus réalistes et plus impactant pour susciter des émotions. Enfin, le nombre de vidéos a été jugé trop élevé, causant de la fatigue et de l'ennui chez les sujets de l'expérimentation.

Ces résultats permettent de justifier les modifications apportées pour l'expérimentation suivante, à savoir :

- Moins d'extraits vidéo pour éviter les effets de fatigue et de lassitude.
- Une seule dimension émotionnelle étudiée pour éviter la complexité d'induire des catégories d'émotions diversifiées.
- Une activité plus simple pour engager les sujets sans perturber l'induction par des vidéos : une tâche de description.
- L'amélioration globale de l'interface de l'application d'expérimentation pour ne pas frustrer les sujets.

Nous avons pu observer une relation entre le nombre de réponses correctes et les niveaux de valence des auto-évaluations. Cela nous permet de justifier l'intérêt de procéder à des protocoles expérimentaux fondés sur des tâches de performances, pour provoquer des états de valence variés. Ainsi, une expérience complémentaire centrée sur la manipulation de la difficulté est pertinente pour tester cette approche à part entière, sans l'utilisation de médias pour influencer sur les émotions.

Bien qu'adapté pour une évaluation globale des niveaux de valence et d'activation, l'outil SAM ne permet pas de réaliser des analyses avec un niveau de granularité

élevé. Certains participants ont exprimé la difficulté d'annoter leurs émotions avec l'outil SAM, plus particulièrement sur la dimension de l'activation. Les personnages représentés n'ont pas été jugés comme suffisamment représentatifs des émotions ressenties. Il est donc complexe de comprendre plus en détail les sentiments subjectifs des sujets de l'expérimentation. Pour pallier cela, nous avons donc intégré l'instrument GEW dans nos expérimentations suivantes afin d'obtenir des données plus complètes sur le sentiment subjectif ressenti et exprimé par les participants.

L'ensemble de cette expérimentation n'a pas permis d'obtenir des effets mesurés importants sur le sentiment subjectif. Cependant, cette expérimentation a permis d'adapter les conditions expérimentales au contexte du smartphone. Ainsi, les résultats obtenus ont été essentiels pour la suite de notre projet de recherche.

V.2.3 Expérimentation principale

Dans le cadre de cette expérimentation, chaque sujet réalise 12 auto-évaluations sur la GEW, une pour chaque extrait vidéo visualisé. Pour chaque auto-évaluation les participants peuvent proposer zéro à deux émotions. Chaque émotion est déclarée avec un niveau d'intensité entre un et six. Les émotions correspondent, soit à une valence positive (à droite du cercle voir figure IV.4), soit à une valence négative (à gauche du cercle). Ainsi, une auto-évaluation peut être mixte et contenir à la fois une émotion de valence positive et une émotion de valence négative.

Pour cette expérimentation, nous analysons les différences entre les trois conditions expérimentales, chacune avec une induction de valence différente : négative, neutre ou positive.

V.2.3.1 Nombre d'émotions déclarées

Nous voulons mesurer la différence entre le nombre d'émotions déclarées en fonction des modalités de valence des extraits vidéo. Au préalable, nous formulons deux hypothèses opérationnelles :

- **Hypothèse II.A.1** : Les vidéos avec une valence neutre ont en moyenne significativement moins d'émotions sélectionnées par rapport aux vidéos positives et négatives. (Condition neutre < Conditions positive et négative)
- **Hypothèse II.A.2** : Les vidéos avec une valence positive et négative ont en moyenne le même nombre d'émotions sélectionnées. (Condition positive = Condition négative)

Le nombre moyen d'émotions exprimées par les participants par auto-évaluation et par condition expérimentale est représenté en figure V.6. Nous pouvons observer, graphiquement, une moyenne d'émotions déclarées élevée pour les conditions négatives et positives avec respectivement 1,87 et 1,74 émotions déclarées par auto-évaluation. La condition neutre semble susciter moins d'émotions avec une moyenne de 1,13 émotions déclarées par auto-évaluation.

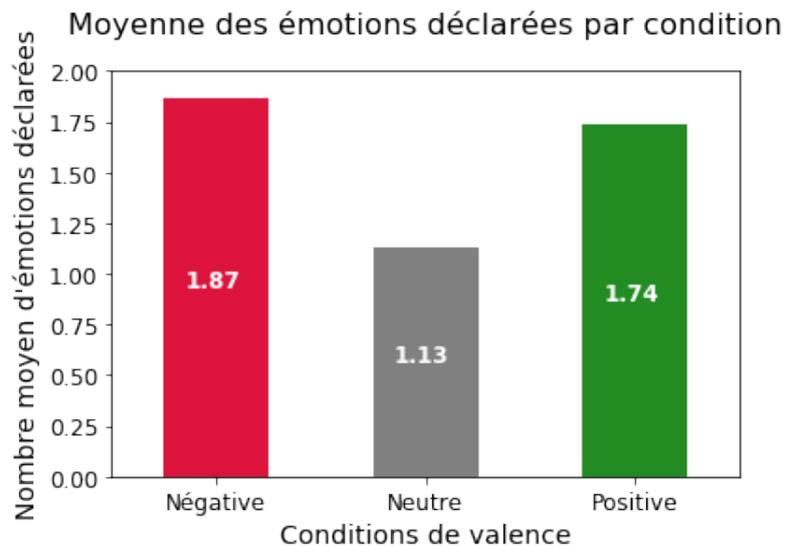


FIGURE V.6 – Nombre moyen d'émotions déclarées

Test d'hypothèse : l'hypothèse nulle correspond à des moyennes similaires pour chacune des conditions expérimentales. Les échantillons de chaque condition ne suivent pas une loi normale selon un test de Shapiro-Wilk ($p\text{-value} < 0.05$), nous choisissons donc un test non paramétrique. Selon le test de Kruskal-Wallis (statistique = 110,237, $p\text{-value} = 1.154 \times 10^{-24} < 0.001$), nous pouvons rejeter l'hypothèse nulle. Il y a donc des différences significatives de nombre d'émotions déclarées entre les trois conditions expérimentales.

Pour valider les hypothèses, nous appliquons un test post-hoc pour comparer les moyennes des conditions deux à deux. Les résultats sont visibles sur le tableau V.3. Il y a donc bien des différences significatives entre les conditions positive et neutre d'une part, et entre les conditions négative et neutre d'autre part. Nous pouvons donc valider l'hypothèse II.A.1 la condition neutre a significativement moins d'émotions déclarées par auto-évaluation par rapport aux autres conditions. En revanche, nous avons postulé qu'il n'y aurait pas de différence significative entre les conditions positive et négative, mais le test d'hypothèse montre qu'il y a une différence significative entre les deux. Il y a donc significativement plus d'émotions déclarées sous la condition négative que sous

la condition positive. L'hypothèse II.A.2 n'est donc pas validée.

TABLE V.3 – Tests de Mann-Whitney sur le nombre d'émotions déclarées

Test	Statistique	P-value
Négatif vs Neutre	6060	1.229×10^{-21}
Positif vs Neutre	7330,5	5.509×10^{-14}
Positif vs Négatif	11130,5	0,0015

Le nombre d'émotions exprimées nous permet de comprendre globalement l'effet provoqué par les extraits vidéo sous les différentes conditions expérimentales sur le sentiment subjectif. Sans prendre en compte la valence ou l'intensité des émotions exprimées, nous voyons que la condition neutre suscite moins d'expression d'émotions chez les participants par rapport aux conditions plus chargées émotionnellement. Nous avons envisagé que le nombre d'expressions d'émotions soit similaire entre les conditions négative et positive, cependant, il semblerait que la condition négative ait un effet plus important. Il est donc possible que les vidéos négatives aient suscité des émotions plus fortes par rapport aux vidéos positives.

V.2.3.2 Intensité des émotions déclarées

Nous explorons ensuite les différences de niveaux d'intensité en fonction des conditions expérimentales. Nous formulons deux hypothèses opérationnelles pour diriger notre analyse :

- **Hypothèse II.A.3** : Les vidéos neutres ont une moyenne d'intensité significativement inférieure par rapport aux vidéos positives et négatives. (Condition neutre < Conditions positive et négative).
- **Hypothèse II.A.4** : Les vidéos avec une valence positive et négative ont des moyennes intensités déclarées similaires. (Condition positive = Condition négative).

Les moyennes d'intensité par condition sont représentées sur la figure V.7. Les moyennes d'intensité des déclarations sont visuellement plus élevées pour les conditions négative et positive avec respectivement, 4,28 et 4,05. La moyenne est plus faible pour la condition neutre avec 3,47.

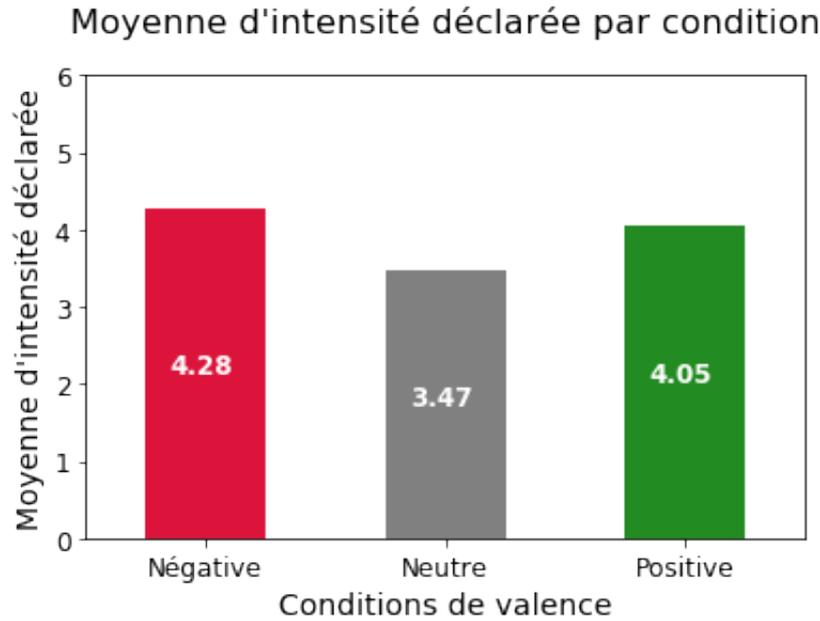


FIGURE V.7 – Moyenne d'intensité des émotions déclarées

Test d'hypothèse : l'hypothèse nulle correspond à des moyennes similaires pour chacune des conditions expérimentales. Un test de Shapiro-Wilk (p -value < 0.05) permet d'attester du caractère non paramétrique des échantillons. Nous appliquons un test de Kruskal-Wallis (statistique = 29,398, p -value = $4.122 \times 10^{-7} < 0.001$), nous pouvons rejeter l'hypothèse nulle. Il y a donc des différences significatives sur l'intensité des émotions déclarées entre les trois conditions expérimentales.

Le tableau V.4 représente les tests d'hypothèses deux à deux, entre chaque condition expérimentale. Nous observons une différence significative entre la moyenne d'intensité dans la condition neutre par rapport aux autres conditions, selon les tests de Mann-Whitney. Nous pouvons donc valider l'**hypothèse II.A.3**. De plus, il n'y a pas de différence significative entre les conditions positive et négative. Cependant, la p -value de ce test est proche de 0,05. Nous pouvons donc valider aussi l'**hypothèse II.A.4**.

TABLE V.4 – Tests de Mann-Whitney sur les moyennes d'intensité

Test	Statistique	P-value
Positif vs Neutre	9745,5	1.013×10^{-3}
Négatif vs Neutre	8469,5	7.069×10^{-8}
Positif vs Négatif	11501,5	0,0567

En conclusion, les extraits de vidéos neutres ont suscité des intensités plus faibles par rapport aux extraits vidéo positive et négative. De plus, les conditions positive et

négative possèdent une intensité déclarée similaire, bien que légèrement plus élevée pour la condition négative et très proche d'une différence significative.

V.2.3.3 Valence des émotions déclarées

Nous cherchons à évaluer les différentes valences exprimées par l'intermédiaire de la GEW, en fonction des conditions expérimentales : extraits vidéo négatifs, neutres ou positifs. La figure V.8 décrit les nombres moyens d'émotions déclarées par type de valence et pour chaque type d'extraits vidéo. Les colonnes colorées en vert représentent les nombres moyens d'émotions déclarées avec une valence positive, les colonnes en rouge représentent les nombres moyens d'émotions déclarées avec une valence négative. Visuellement, nous pouvons observer un nombre moyen élevé d'émotions de valence négative pour la condition avec des extraits vidéo négatifs avec 1,04 émotions déclarées. De plus, nous pouvons observer un nombre moyen élevé d'émotions de valence positive pour la condition avec des vidéos positives à 1,56. Cependant, nous observons aussi une moyenne d'émotions de valence positive déclarée relativement élevée sous la condition négative.

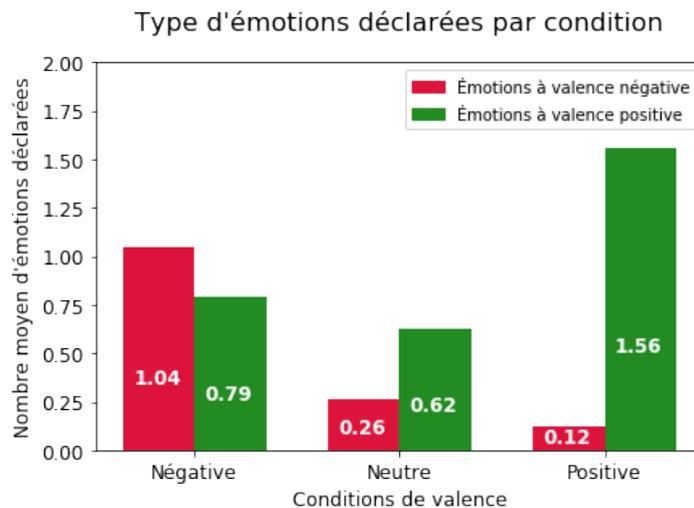


FIGURE V.8 – Nombre moyen d'émotions par valence et par condition

Émotions de valence positive : nous étudions d'abord les différences du nombre moyen d'émotions avec une valence positive. Pour cela, nous formulons les hypothèses opérationnelles suivantes :

- **Hypothèse II.A.5** : Les vidéos positives ont un nombre d'auto-évaluations positives significativement supérieur aux vidéos neutres ou négatives. (Condition positive > Conditions neutre et négative).

- **Hypothèse II.A.6** : Les vidéos neutres ont un nombre d'auto-évaluations positives significativement similaire aux vidéos négatives. (Condition neutre = Condition négative).

L'objectif est de comparer les nombres moyens d'émotions de valence positive entre les conditions expérimentales. D'une part, nous voulons vérifier s'il y a plus d'émotions de valence positive dans la condition positive par rapport aux autres conditions. D'autre part, nous voulons vérifier si les conditions négatives et neutres ont des moyennes similaires.

Test d'hypothèse : selon l'hypothèse nulle, les moyennes sont similaires pour chacune des conditions expérimentales. Nous voulons déterminer dans un premier temps s'il existe des différences significatives entre les trois échantillons. Un test de Shapiro-Wilk montre que les échantillons ne suivent pas une loi normale (p -value < 0,05), il est donc nécessaire d'utiliser un test non paramétrique. Un test de Kruskal-Wallis permet de rejeter l'hypothèse nulle, les moyennes sont donc statistiquement différentes (statistique = 123,964, p -value = $1.206 \times 10^{-27} < 0.001$).

Des analyses complémentaires ont été réalisées pour comparer les conditions deux à deux, visibles dans le tableau V.5. La condition positive est significativement différente par rapport aux conditions négative et neutre, **l'hypothèse II.A.5** est donc validée. En revanche, nous notons aussi une différence significative entre la condition négative et la condition neutre, avec une moyenne d'émotions avec une valence positive supérieure dans la condition négative. **L'hypothèse II.A.6** n'est donc pas validée.

TABLE V.5 – Tests de Mann-Whitney sur la valence positive

Test	Statistique	P-value
Positif vs Neutre	4844,5	8.057×10^{-25}
Négatif vs Neutre	11133,5	0,0141
Positif vs Négatif	5952,5	4.070×10^{-19}

Globalement, nous obtenons des résultats cohérents avec l'induction d'émotions recherchée pour les différentes conditions. En revanche, le nombre moyen d'émotions de valence positive est curieusement élevé pour la condition négative par rapport à nos attentes. Une analyse plus détaillée des libellés est donc nécessaire pour comprendre les émotions de valence positives proposées dans la condition négative.

Émotions de valence négative : de la même manière que pour les émotions de valence positive, nous explorons les nombres moyens d'émotions de valence négative entre les conditions en prenant comme référence les hypothèses opérationnelles suivantes :

- **Hypothèse II.A.7** : Les vidéos négatives ont un nombre d'auto-évaluations négatives déclarées significativement supérieur aux vidéos neutres ou positives. (Condition négative > Conditions neutre et positive).
- **Hypothèse II.A.8** : Les vidéos positives et négatives ont un nombre d'auto-évaluations négatives similaires. (Condition positive = Condition neutre).

Nous pouvons observer sur la figure V.8 les différences de moyenne par condition. Seule la condition négative possède une moyenne élevée avec 1,04 émotion négative déclarée en moyenne par auto-évaluation. Les conditions neutre et positive présentent un nombre moyen d'émotions négatives déclarées beaucoup plus faible avec respectivement 0,26 et 0,12.

Nous cherchons à déterminer si la condition négative a statistiquement plus d'émotions de valence négative par rapport aux autres conditions. De plus, nous voulons vérifier si les conditions positive et neutre ont des moyennes de valence négative similaires.

Test d'hypothèse : l'hypothèse nulle correspond à une moyenne similaire entre toutes les conditions expérimentales. Un test de Shapiro-Wilk a permis de déterminer que les échantillons ne suivent pas une loi normale (p -value < 0,5), nous utiliserons donc un test non paramétrique. Un test de Kruskal-Wallis met en évidence les différences significatives entre les moyennes des échantillons (statistique = 268,30, p -value = $2.839 \times 10^{-37} < 0.001$).

Une analyse des conditions deux à deux est disponible sur le tableau V.6. L'hypothèse II.A.7 est validée, en effet le nombre moyen d'émotions de valence négative sous la condition négative est significativement supérieur aux conditions neutres et positives. Cependant, l'hypothèse II.A.8 n'est pas validée. En effet, on observe une moyenne significativement plus élevée d'émotions négatives pour la condition neutre que pour la condition positive.

TABLE V.6 – Tests de Mann-Whitney sur la valence négative

Test	Statistique	P-value
Positif vs Neutre	11324	0,00304
Négatif vs Neutre	5653	1.0669×10^{-21}
Positif vs Négatif	4347,5	6.489×10^{-31}

Les résultats d'induction d'émotions sont globalement en correspondance avec ce qui était attendu. Nous observons que la majorité des déclarations d'émotions de valence négative sont donc concentrées dans la condition négative. De plus, en moyenne, les auto-évaluations de valence négative sont bien plus élevées sous la condition négative que sous les autres conditions. Nous constatons cependant une

différence entre la condition neutre et positive. Certains extraits vidéo de la condition neutre ont donc suscité des réactions négatives chez les sujets.

Analyse des libellés :

Les figures V.9, V.10 et V.11 représentent les distributions des libellées pour les conditions positives, négatives et neutres. Pour la condition positive, nous pouvons observer que certaines émotions sont fortement représentées, c’est le cas notamment de « l’amusement » (136 occurrences), de « la joie » (44 occurrences) et « du plaisir » (25 occurrences). Ces libellés sont cohérents avec le registre des extraits vidéo. Certains items n’ont pas ou peu été utilisés, comme pour « l’admiration », « le soulagement » ou « la fierté », avec moins de cinq déclarations. Les émotions de valence négative ne dépassent pas les cinq occurrences au maximum.

Distribution des émotions déclarées sur les vidéos positives

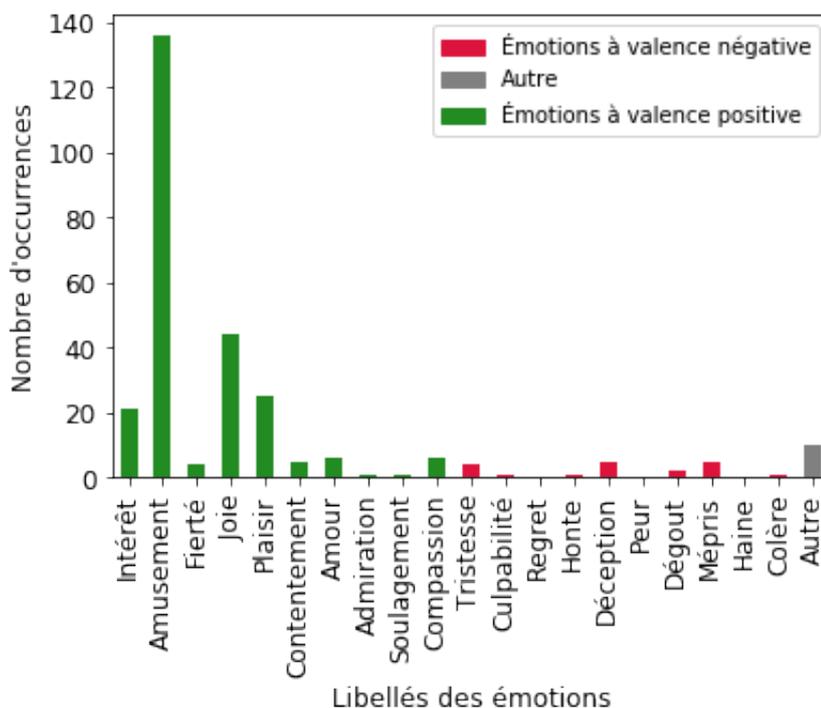


FIGURE V.9 – Distribution des libellés sous la condition positive

Pour la condition négative, nous pouvons constater une plus grande diversité d’émotions déclarées. Premièrement, certaines émotions de valence positive ont été proposées, principalement « la compassion » qui rassemble 77 occurrences. Quatre items représentent la majorité des libellés de valence négative déclarés, à savoir « la peur » (44 occurrences) « le dégoût » (34 occurrences), « le mépris » (24 occurrences) et

« la tristesse » (29 occurrences). Peu d'émotions autres ont été proposées.

Distribution des émotions déclarées sur les vidéos négatives

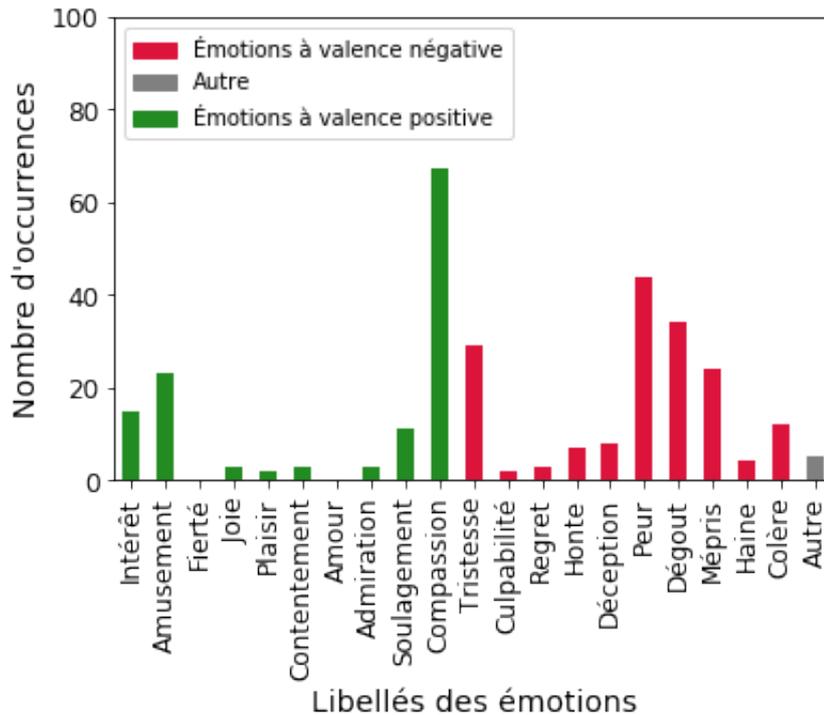


FIGURE V.10 – Distribution des libellés sous la condition négative

La condition neutre est marquée par quelques émotions très représentées, telles que « l'intérêt » (37 occurrences), « la compassion » (28 occurrences) et « la déception » (18 occurrences). De manière plus détaillée, un certain extrait vidéo a suscité un nombre de réactions négatives particulièrement important, probablement causé par une expression faciale d'un des protagonistes de la vidéo. De plus, de nombreux autres libellés ont été proposés par les participants (39 occurrences). Parmi ces propositions figurent « l'indifférence » (9 occurrences), « le désintérêt » (9 occurrences) et « la curiosité » (5 occurrences).

Distribution des émotions déclarées sur les vidéos neutres

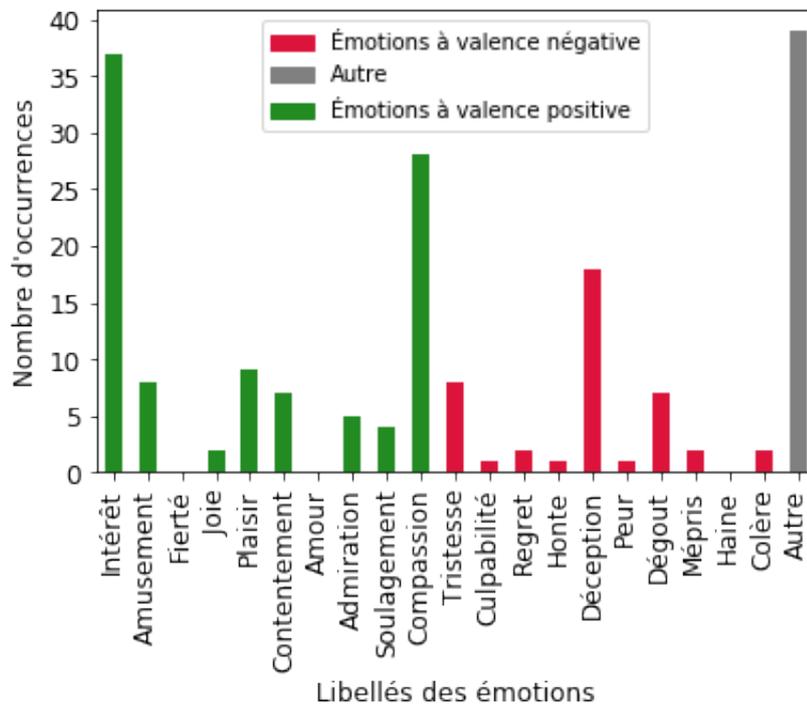


FIGURE V.11 – Distribution des libellés en condition neutre

L'analyse détaillée des libellés permet d'en apprendre plus sur les auto-évaluations proposées par les participants et met en lumière certains éléments. Pour le cas de la condition positive, peu de libellés représentent une majorité des déclarations. À l'inverse, dans la condition négative, les émotions sont plus réparties, mais nous pouvons noter une forte utilisation du libellé « compassion », pourtant classé comme une émotion avec une valence positive dans la GEW.

V.2.3.4 Analyse des verbatims

Afin de compléter les analyses des données obtenues avec la GEW, les descriptions rédigées par chaque participant ont été analysées par traitement automatique du langage, afin d'en extraire une polarité. Nous avons utilisé la bibliothèque Python *Text-Blob-Fr* pour calculer la polarité de chaque description (Loria, 2018). Cette bibliothèque Python fonctionne avec une approche lexicale, telle qu'expliquée dans la sous-section III.3.4. Elle se fonde sur une base de données prévue pour la langue française. Chacune des descriptions a permis d'obtenir une notation entre -1 et 1, -1 indiquant une polarité faible et donc une valence négative associée au contenu et 1 indiquant une polarité élevée et donc une valence positive. Nous formulons les hypothèses opérationnelles sui-

vantes :

- **Hypothèse II.B.1** : La polarité des verbatims des vidéos de valence positive est significativement supérieure à la polarité des verbatims des vidéos neutres et négatives.
- **Hypothèse II.B.2** : La polarité des verbatims des vidéos de valence négative est significativement inférieure à la polarité des verbatims des vidéos neutres et positives.

La figure V.12 représente le résultat moyen par condition expérimentale. Les moyennes montrent une adéquation entre les conditions expérimentales et les polarités obtenues dans le texte.

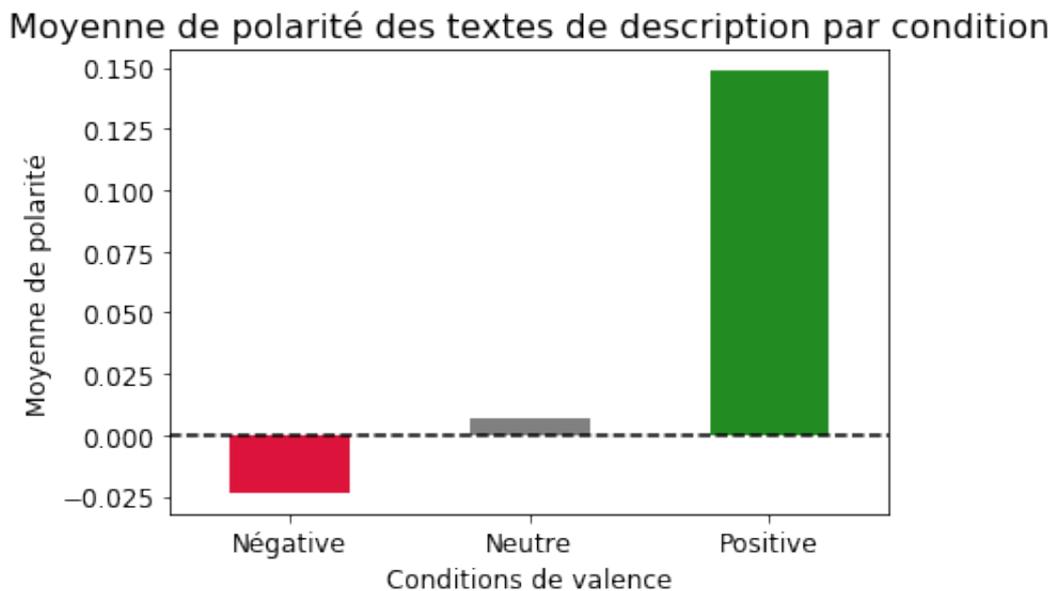


FIGURE V.12 – Moyenne de polarité des émotions déclarées

Des tests statistiques ont été réalisés pour répondre aux hypothèses II.B.1 et II.B.2. Nous cherchons à identifier si sous la condition positive, la polarité est en moyenne plus élevée que sous la condition neutre, et si cette dernière a une polarité moyenne plus élevée que la condition négative. Selon un test Shapiro-Wilk, les échantillons suivent une loi normale ($p\text{-value} > 0.05$). En revanche, ils ne sont pas homoscedastiques (les échantillons n'ont pas de variances similaires), selon un test de Bartlett ($p\text{-value} < 0.05$), nous avons donc utilisé un test non paramétrique. Selon un test de Kruskal-Wallis, les différences entre les polarités des conditions expérimentales sont significatives.

Des tests d'hypothèses deux à deux permettent de vérifier les différences entre chaque condition (voir le tableau V.7). La différence est significative entre les conditions positive et neutre, ce qui valide l'hypothèse II.B.1. De plus, la différence est significative entre les conditions négative et neutre, ce qui valide l'hypothèse II.B.2. En revanche, ce dernier test présente une p-value plus élevée, indiquant une plus faible différence entre les conditions neutre et négative, visible au niveau du graphique V.12.

TABLE V.7 – Tests de Mann-Whitney sur la polarité des descriptions

Test	Statistique	P-value
Positif vs Neutre	6925,5	4.618×10^{-15}
Négatif vs Neutre	11776,5	0,020
Positif vs Négatif	6909,5	6.051×10^{-14}

Dans l'ensemble, les résultats de polarité obtenus concordent avec les conditions expérimentales et les hypothèses formulées. Les textes rédigés par les participants contiennent donc des éléments de langage correspondant aux extraits vidéo visualisés.

V.2.3.5 Discussion

L'ensemble des données recueillies par l'instrument GEW nous a permis d'extraire de nombreuses informations pour réaliser des analyses sur le sentiment subjectif exprimé par les sujets. Dans l'ensemble, les résultats correspondent avec de nombreuses hypothèses que nous avons formulées :

- Le nombre d'émotions est plus élevé dans les conditions positive et négative par rapport à la condition neutre. Cependant, il y a en moyenne plus d'émotions exprimées sous la condition négative. (Validation de l'hypothèse II.A.1, non-validation de l'hypothèse II.A.2)
- De la même manière, les intensités sont en moyenne significativement plus fortes pour les conditions positive et négative. Il y a une différence faible et peu significative entre la condition positive et la condition négative, cette dernière étant légèrement plus élevée. (Validation de l'hypothèse II.A.3, faible validité de l'hypothèse II.A.4).
- Pour les émotions de valence positive, nous avons confirmé qu'elles étaient significativement plus élevées pour les extraits vidéo de la modalité positive. En revanche, nous avons trouvé un nombre d'auto-évaluations positives relativement élevé pour la condition négative, par rapport à nos attentes. Par conséquent, le nombre d'émotions positives étaient significativement plus élevé dans la condition négative que dans la condition neutre. (Validation de l'hypothèse II.A.5, non-validation de l'hypothèse II.A.6)

- Enfin, nous avons trouvé un nombre plus élevé d'émotions de valence négative sous la condition négative, par rapport aux autres conditions. Cependant, la condition neutre a suscité plus d'émotions de valence négative que la condition positive. (Validation de l'**hypothèse II.A.7**, non-validation de l'**hypothèse II.A.8**)

Pour répondre à l'hypothèse générale formulée dans la sous-section [IV.4.3](#), les conditions positive et négative présentent des effets significatifs par rapport à la condition neutre, avec la validation des **hypothèses II.A.1, II.A.3, II.A.5 et II.A.7**. Ainsi, ces conditions expérimentales sont liées à la déclaration d'un plus grand nombre d'émotions, avec une plus forte intensité. De plus, les conditions négative et positive sont associées respectivement aux déclarations d'émotions de valence négative et positive. Ce dernier effet est renforcé par la validation des **hypothèses II.B.1 et II.B.2** sur la polarité des verbatims. Nous pouvons donc valider le fonctionnement du protocole expérimental pour induire des émotions avec une forte intensité et correspondant aux conditions expérimentales à partir d'extraits vidéo. Ainsi, le contenu et la qualité des extraits vidéo et leur intégration dans une application mobile expérimentale ont permis d'engager les participants et d'obtenir des résultats mesurables et significatifs en matière d'induction d'émotions.

Cependant, la non-validation des **hypothèses II.A.2 et II.A.4** montre un déséquilibre entre la condition négative et positive, avec des effets plus forts de la condition négative sur le nombre d'émotions et l'intensité déclarée. Ce résultat semble cohérent avec la méta-analyse réalisée par Westermann *et al.* (1996), qui met en évidence que les extraits vidéo avec une valence négative ont généralement un effet plus élevé que ceux avec une valence positive. Par rapport à la valence, les **hypothèses II.A.6 et II.A.8** montrent des effets inattendus. Nous notons, par exemple, un nombre d'émotions avec une valence positive relativement élevée dans la condition négative. Ces déséquilibres laissent penser qu'il est possible d'ajuster le protocole expérimental pour mieux cibler les émotions à induire.

Plus spécifiquement, l'étude approfondie des libellés montre une faible variété dans le choix des libellés pour les extraits vidéo avec une valence positive. Cela peut être causé par le contenu des extraits vidéo qui sont principalement composés de situations impliquant des passages avec des enfants ou des animaux réalisant des actions drôles. Ces vidéos ont été choisies, car elles sont celles qui ont été annotées le plus positivement dans le corpus proposé par Samson *et al.* (2016). Nous pouvons donc envisager d'inclure plus de diversité de contenu pour les extraits vidéo positifs, afin de susciter des émotions comme le plaisir ou le soulagement. Il peut aussi y avoir une tendance chez les participants d'être plus affecté par les situations négatives et donc d'être plus expressif sur cette condition. Nous pouvons conclure sur un certain déséquilibre entre les conditions positive et négative sur l'effet perçu par les participants.

Certains libellés n'ont pratiquement jamais été utilisés comme « l'amour », « la fierté » ou « la honte », il s'agit d'émotions que nous pouvons classer comme des émotions liées à l'implication personnelle des sujets. Ainsi, dans le modèle OCC d'Ortony *et al.* (1990), « la fierté » et « la honte » sont attribuées à des contextes dans lesquels le sujet est directement impliqué (voir figure II.3). Dans le cas d'un protocole, dans lequel les participants ont un rôle passif face aux événements présentés, ces émotions sont donc moins pertinentes. À partir du champ libre, les ressentis des sujets sur l'application ont pu être recueillis. Bien que l'instrument GEW ait été bien compris, certains participants ont suggéré qu'il manquait certaines émotions, telles que la surprise. Paradoxalement, certains sujets ont estimé le nombre d'émotions sur la GEW trop important et d'autres ont estimé qu'il n'y en avait pas assez. Cela montre la variabilité inter-individuelle des effets d'un même stimulus. Par conséquent, il est difficile de concevoir des outils de mesures subjectives génériques pour obtenir le sentiment subjectif des participants.

Nous pouvons aussi nous poser la question du choix de valence pour certaines émotions. En effet, la forte utilisation du label « compassion » sous la condition négative, bien qu'attribuée dans la GEW à une valence positive, soulève une problématique. Nous pouvons estimer qu'il s'agit ici de l'expression de l'empathie de la part des participants, en rapport aux événements désagréables vécus par les personnages dans les extraits vidéo. Dans ce contexte, associer la compassion à une valence positive n'est pas nécessairement un choix évident. L'empathie étant liée à une dimension sociale, il n'est pas nécessairement pertinent de l'associer à une valence positive ou négative. Cette dimension sociale n'est pas présente explicitement dans l'instrument GEW et donc complexe à évaluer. En effet, l'objet de l'émotion peut être tourné vers autrui plutôt que vers soi-même. Il pourrait alors être pertinent d'ajouter des solutions pour permettre aux utilisateurs de spécifier si l'émotion est ressentie dans une situation sociale. Une autre approche pourrait être de créer une roue spécifique avec des libellés adaptés pour des situations où les émotions sont portées vers un objet externe, tel qu'un média ou un autre agent.

Une analyse lexicale par traitement automatique du langage a permis d'obtenir des informations supplémentaires sur le ressenti des participants. Il peut être complexe d'inférer sur le sentiment subjectif à partir de ces données, car les textes sont censés retranscrire les événements qui surviennent dans les extraits vidéo. Néanmoins, cela nous permet d'une part de tester la pertinence de l'analyse pour extraire la polarité émotionnelle et d'autre part de vérifier que les participants ont retranscrit le contenu des médias de manière cohérente. Ce dernier point peut indiquer, entre autres, l'engagement des participants pour les médias présentés. Les deux hypothèses proposées ont été validées. La précision de l'analyse lexicale reste cependant limitée, car elle se limite à l'extraction des termes un à un pour établir une moyenne de polarité sur un corpus. Le texte n'est donc pas analysé dans son ensemble, mais mot à mot. Il nous semble cependant que la modalité textuelle est essentielle à prendre en compte dans le cadre de discussion sur messagerie instantanée, car elle est en mesure d'apporter des

informations pertinentes sur le ressenti des utilisateurs.

Globalement, ce protocole expérimental met en évidence l'intérêt d'utiliser la GEW et plus particulièrement l'outil GEWMI pour analyser le sentiment subjectif des participants. Les résultats montrent la faisabilité et la pertinence de réaliser des expérimentations sur les émotions sur application mobile par l'intermédiaire d'un agent conversationnel. L'usage d'extraits vidéo, en nombre plus limité, de meilleure qualité et avec un contenu plus réaliste, semble avoir engagé les participants et facilité l'induction d'états émotionnels. Avec 40 participants, la taille de la base de données réalisée reste cependant limitée, il serait donc souhaitable d'augmenter les effectifs de cette expérimentation pour collecter plus de données. Des modifications peuvent sans doute être apportées pour améliorer le protocole expérimental. En l'état, cette version de l'application d'expérimentation et ce protocole expérimental constituent des apports importants de cette thèse.

V.2.4 Expérimentation complémentaire

L'expérimentation complémentaire a permis de tester une méthodologie d'induction d'émotions orientée vers les performances des sujets. La particularité a été de proposer deux conditions différentes avec des niveaux de difficulté variable, afin de provoquer des situations d'échec ou de réussite, par rapport à un objectif à atteindre.

Deux conditions expérimentales ont été proposées : une condition facile et une condition difficile. L'objectif de l'ensemble de l'expérimentation était d'obtenir plus de six items justes. Parmi les sujets, nous obtenons les résultats suivants :

- Pour la condition facile : 12 ont réussi l'objectif et 8 ont échoué.
- Pour la condition difficile : 3 ont réussi et 17 ont échoué.

Le score moyen obtenu sous la condition facile est de 7,5 contre 5,1 sous la condition difficile. Un test de Mann-Whitney permet de montrer qu'il existe une différence significative entre ces deux résultats (statistique = 113, $p < 0,01$). Nous étudions les données des émotions déclarées à la fin de l'expérimentation sur l'outil GEWMI, les participants pouvaient déclarer jusqu'à trois émotions. 40 auto-évaluations ont donc été analysées. L'objectif est de comparer les émotions déclarées entre les 15 participants qui ont réussi le test et les 25 qui ont échoué.

V.2.4.1 Valence des émotions déclarées

La figure V.13 représente les moyennes d'émotions déclarées par type de valence et en fonction du succès ou de l'échec des participants. Nous observons un nombre

similaire d'émotions de valence positive et négative dans un cas d'échec avec respectivement 1,19 et 1,04 émotions en moyenne. À l'inverse, la majorité des émotions déclarées sous la condition de succès sont des émotions de valence positive avec 2,13 émotions de valence positive déclarées en moyenne contre 0,4 émotions de valence négative en moyenne.

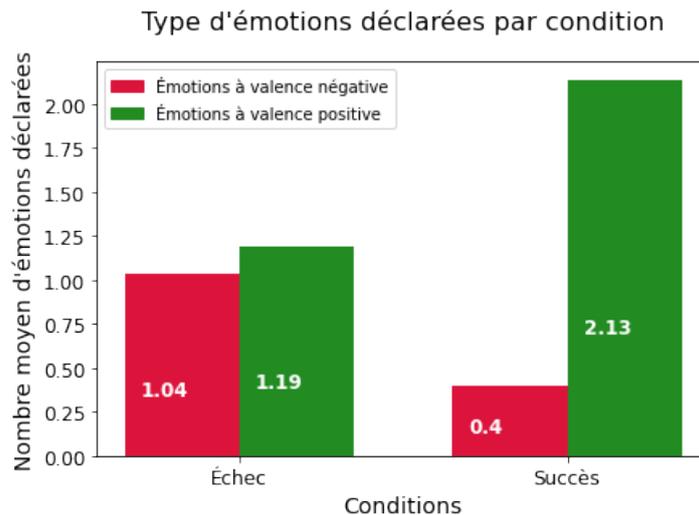


FIGURE V.13 – Nombre moyenne d'émotions déclarés par valence et par condition

D'une part, nous voulons vérifier si en condition de succès le nombre d'émotions de valence positive déclarées est supérieur par rapport à la condition d'échec au test. De même, nous voulons vérifier si en situation d'échec, il y a plus d'émotions de valence négative déclarées, par rapport à la condition de succès. Nous formulons donc les hypothèses opérationnelles suivantes :

- **Hypothèse III.1** : Les sujets qui ont réussi le test déclarent un nombre d'auto-évaluations positives significativement supérieur aux sujets qui ont échoué au test.
- **Hypothèse III.2** : Les sujets qui ont échoué le test déclarent un nombre d'auto-évaluations négatives significativement supérieur aux sujets qui ont réussi le test.

Émotions de valence positive : les échantillons ne suivent pas une loi normale selon un test Shapiro-Wilk ($p\text{-value} < 0.05$), nous appliquons donc un test non paramétrique. Selon un test de Mann-Whitney (statistique = 113, $p\text{-value} = 0,00223 < 0,001$), les émotions de valence positive sont donc significativement plus élevées en cas de succès qu'en cas d'échec. Nous pouvons donc valider l'hypothèse **III.1**.

Émotions de valence négative : un test de Shapiro-Wilk ($p\text{-value} < 0.05$) montre qu'il est nécessaire d'appliquer un test non paramétrique. Un test de Mann-Whitney

(statistique = 268,30, p-value = 0,015 < 0,05). Ainsi, les émotions avec une valence négative sont significativement plus élevées en cas d'échec. L'hypothèse III.2 est donc aussi validée.

Les validations des hypothèses opérationnelles confirment qu'il y a un effet de la réussite ou de l'échec au test sur la valence exprimée.

V.2.4.2 Détails des libellés

Nous pouvons étudier plus en détail les libellés sélectionnés pour les conditions de succès ou d'échec par l'intermédiaire des figures V.14 et V.15. En cas de succès, les participants déclarent principalement de « l'intérêt » (10 occurrences), de « la fierté » et du « contentement ». Certains participants déclarent aussi ressentir de la déception, malgré la situation de succès.

Distribution des émotions déclarées en condition de succès

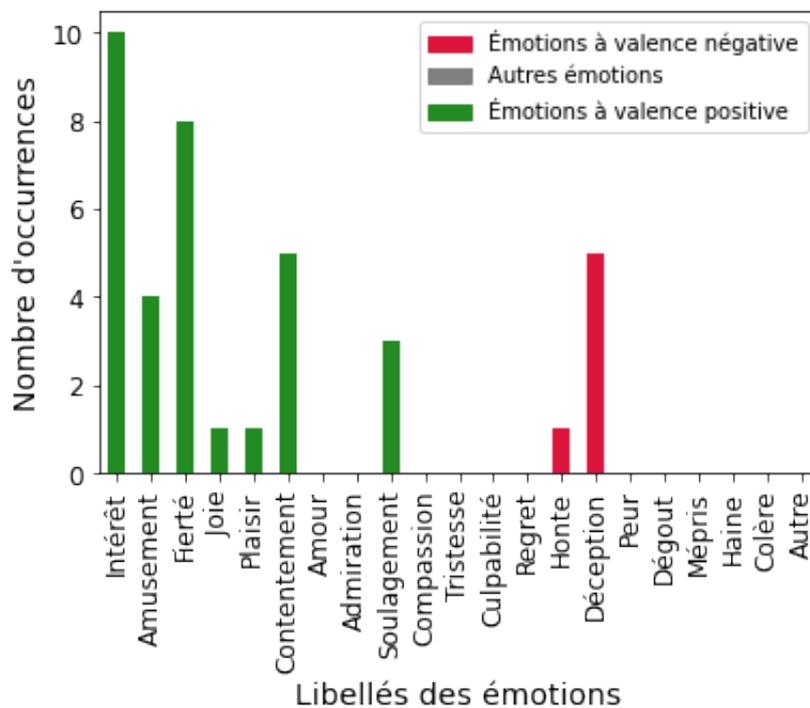


FIGURE V.14 – Distribution des libellés en condition de succès

La condition d'échec est caractérisée par une plus grande variété de valence dans les émotions déclarées. En valence positive, certains participants ont exprimé de « l'amusement » (14 occurrences) et de « l'intérêt » (9 occurrences). En valence négative, « la

déception » a particulièrement été sélectionnée par les sujets avec 13 occurrences, en deuxième position « la honte » (5 occurrences).

Distribution des émotions déclarées en condition d'échec

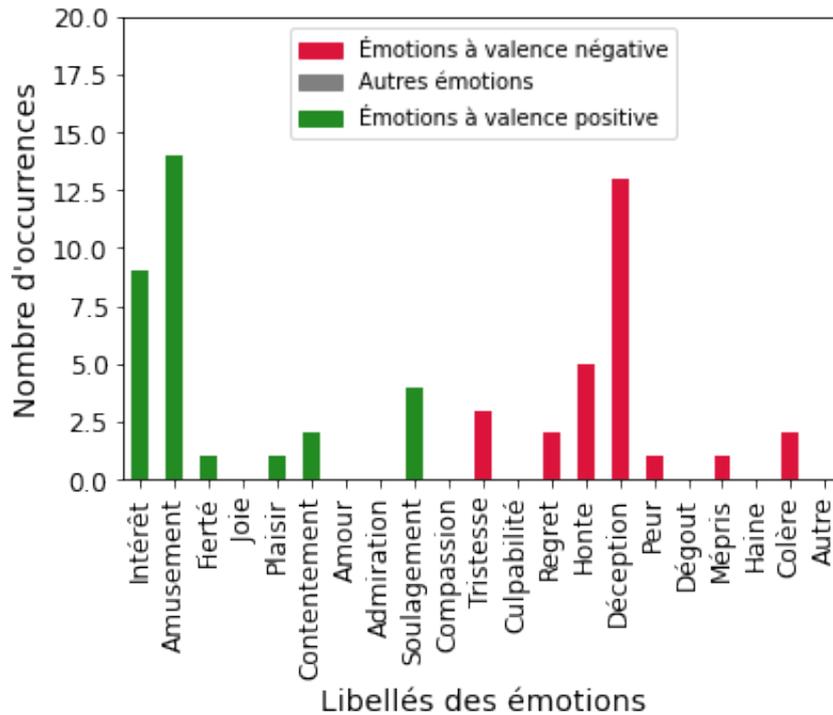


FIGURE V.15 – Distribution des libellés en condition d'échec

V.2.4.3 Discussion

L'ensemble des hypothèses sur l'induction d'émotions est validé pour cette expérimentation complémentaire, par l'étude des mesures de la GEW. Ainsi, nous pouvons donc confirmer l'hypothèse générale décrite dans la sous-section IV.5.4. La réussite ou l'échec d'un objectif a un effet significatif sur les valences exprimées dans le cadre de ce protocole sur smartphone.

Par rapport à l'équilibrage du test, même si les scores sont significativement différents en fonction des conditions facile et difficile, le nombre de sujets ayant échoué sous la condition facile est plus élevé que prévu. Le nombre de sujets ayant échoué et le nombre de sujets ayant réussi sont donc déséquilibrés. Ainsi, il est nécessaire de concevoir un test plus équilibré, afin de rendre la condition facile plus à même de produire des réussites. De plus, un panel de participants plus large pourrait permettre d'assurer, d'avoir des tailles d'échantillons suffisamment grandes. Néanmoins, nous

pouvons observer des effets mesurables sur les émotions déclarées sur la GEW à l'issue du test.

L'analyse des libellés permet de mettre en avant l'utilisation d'émotions portées sur l'implication personnelle du sujet selon le modèle OCC, telles que « la fierté » en valence positive ou « la honte » en valence négative (Ortony *et al.*, 1990), (voir figure II.3). De même, nous pouvons noter des émotions liées aux attentes des participants comme « le soulagement », indiquant un événement anticipé, mais non confirmé. Cette émotion étant surtout utilisée en situation de succès, nous pouvons estimer que l'évènement évité est l'échec au test ou un score trop faible. En outre, « la déception » indique la réalisation d'un événement non désiré, à savoir l'échec au test ou un score trop faible par rapport aux attentes du sujet en cas de succès. De la même manière que pour l'expérimentation principale, nous pouvons observer qu'une minorité de libellés est utilisée par les participants. Certains sujets ont estimé que les émotions présentes pouvaient être affinées. En effet, pour une tâche de performance, des émotions comme la « frustration » pourraient être intégrées. Nous pouvons imaginer concevoir une version de GEWMI orientée pour ce type d'activité.

Par rapport aux autres expérimentations, la population est beaucoup plus homogène. En effet, nous avons porté notre attention sur des étudiants, la plupart ont moins de 25 ans et sont habitués à l'usage de smartphones. Ce type de protocole semble fonctionner avec cette population, habituée à réaliser des tests et à être évaluée. Il serait donc pertinent d'appliquer cette méthodologie à une population plus diversifiée, pour en étudier les effets. De plus, la mise en œuvre de l'expérimentation à distance soulève des difficultés sur la maîtrise des conditions expérimentales. En effet, les participants ont passé le test à leur domicile à cause de la situation de confinement. Bien que les expérimentateurs se soient connectés avec les participants par visio-conférence, de nombreuses conditions ne peuvent être contrôlées : telles que le bruit, la luminosité, les interruptions dans l'environnement du participant. Cependant, il est possible que le déroulement d'expérimentations à domicile, dans un contexte plus éloigné des laboratoires, permette une expression plus naturelle et écologique des émotions.

L'application d'un protocole de manipulation de difficulté dans une application mobile pour smartphone constitue une contribution notable de cette thèse et une nouveauté dans ce domaine. Les protocoles de manipulation de difficulté présentent l'avantage d'engager fortement les participants dans une activité et pour induire des émotions mesurables par le sentiment subjectif. Cette approche semble notamment s'adapter au format des smartphones et constitue donc une piste pertinente pour expérimenter en laboratoire par ce média. Comme nous l'avons précisé dans le chapitre méthodologique, la population ayant réalisé l'expérimentation était principalement constituée d'étudiants, qui ont pu se sentir engagés et se créer des attentes fortes sur ce type de protocole. Il serait donc pertinent de le tester sur une population plus

diversifiée et moins habituée à être évaluée de cette manière.

V.2.5 Discussion

Avec ces analyses, nous avons obtenu des éléments de réponses pour la question fondamentale **Question Q_Induction : la mise en œuvre de méthodes d'induction d'émotions sur smartphone nous permet-elle d'obtenir des résultats mesurables à partir d'auto-évaluation?** L'induction d'émotions est une procédure complexe qui ne garantit pas des résultats fiables. Une attention particulière doit être apportée pour équilibrer les protocoles afin de s'adapter au contexte et aux sujets. Le choix du matériel pour l'induction d'émotions d'une part et des outils pour l'expression du sentiment subjectif d'autre part, sont des éléments essentiels pour engager les participants et obtenir des résultats exploitables.

La pré-expérimentation, malgré des résultats peu concluants, a permis de guider notre réflexion pour concevoir des protocoles expérimentaux en mesure d'enregistrer des données plus pertinentes pour la création d'un corpus. Ainsi, le choix du corpus de vidéos et de l'activité à réaliser ont été déterminants pour améliorer l'implication des participants et pour faciliter l'induction d'émotions. De plus, le choix de l'instrument GEW par rapport au SAM, a permis d'entreprendre des analyses plus détaillées et plus poussées pour comprendre le ressenti des participants.

En effet, la GEW s'avère être un outil flexible, permettant des analyses avec plusieurs niveaux de granularité. Ainsi, le sentiment subjectif exprimé par les participants peut être synthétisé sur des dimensions, telles que la valence, ce qui permet de simplifier la compréhension de l'ensemble des états émotionnels vécus. En effet, comme indiqué dans l'état de l'art dans la sous-section II.4.5, les modèles dimensionnels représentent une approche pragmatique pour simplifier la caractérisation d'épisodes émotionnels vécus dans des grands. Bien que moins précis, ils sont largement utilisés en informatique affective, pour labelliser des jeux de données et pour la création de systèmes de reconnaissance automatique. L'expérimentation principale et l'expérimentation complémentaire montrent des effets mesurables sur la valence. De surcroît, obtenir des labels discrets offre une grille d'analyse complémentaire. En effet, nous avons pu mieux comprendre les ressentis des participants. Des contextes qui intègrent une anticipation des événements à venir (réussite ou échec d'un objectif par exemple) ou une dimension sociale avec de l'empathie ont un effet sur la distribution des libellés. Ces notions se retrouvent notamment dans le modèle OCC et peuvent être rapprochées à des situations de communication dans le cas de service client. L'utilisation de la GEWMI associée à la méthodologie appropriée (Scherer *et al.*, 2013) doit permettre de concevoir une nouvelle version de la GEW mieux adaptée aux expériences émotionnelles ressenties par un sujet communiquant avec un téléconseiller ou un agent conversationnel.

En outre, la variabilité individuelle doit être prise en compte. Ainsi, bien que des effets globaux aient pu être mis en évidence, les inductions d'émotions ne sont pas égales entre tous les participants. De plus, un sentiment subjectif mesurable n'indique pas forcément que les participants aient ressenti des émotions fortes et mesurables, sur les composantes physiologique, comportementale ou cognitive. En effet, selon le constat de R. A. Calvo et D'Mello (2010), explicité dans la sous-section III.3.7, l'expression n'est pas nécessairement indicatrice de l'expérience vécue. D'autant plus que l'expression du sentiment subjectif est qu'une partie émergée de l'ensemble de ce composant, il ne rend donc pas compte avec exactitude de l'expérience vécue. Également, selon R. A. Calvo et D'Mello (2010), même si le sentiment subjectif retransmet correctement l'expérience vécue, la synchronisation avec les autres composants n'est pas assurée, si l'intensité de l'émotion n'est pas suffisamment élevée. Le sentiment subjectif reste néanmoins l'une des approches les plus efficaces pour labelliser des jeux de données, que ce soit en situation contrôlée ou en expérimentation sur le terrain.

V.3 Résultats du système de reconnaissance automatique

V.3.1 Labellisation des descripteurs

Nous avons présenté le développement de l'ensemble du système de traitement de données et d'entraînement des modèles de classification dans la section IV.7. Nous pouvons appliquer ce système sur des données labellisées en fonction du sentiment subjectif déclaré par les participants. L'utilisation de la GEW a permis d'associer aux descripteurs générés de nombreux labels potentiels. Tel que le nombre d'émotions déclarées, les intensités ou la valence des émotions. Nous nous intéressons à la tâche de description qui suit l'induction d'émotion par les vidéos. Nous ne sélectionnons donc que des descripteurs calculés à partir de cette phase. De plus, la phase de description inclut le plus de modalités, avec les données de frappe au clavier en plus des données des capteurs inertiels, pour les interactions sur smartphone.

Dans le cadre de l'apprentissage automatique, nous avons fait le choix de labelliser les jeux de données en fonction du niveau de valence. En proposant des vidéos avec plusieurs niveaux de valence, nous nous sommes particulièrement intéressés pour classer cette dimension. Le graphique V.8 sert de référence pour nous aider à correctement labelliser les descripteurs. En effet, nous pouvons observer que de nombreux extraits vidéo de la condition négative ont reçu des déclarations avec des libellés de valence positive. Ce résultat s'est expliqué par une forte présence du libellé « compassion » (voir figure V.10). Nous ne pouvons donc pas sélectionner une classe avec des déclarations

exclusivement négatives, car une majorité des déclarations dans la condition négative sont mixtes. De plus, nous pouvons voir que certaines vidéos de la condition neutre ont aussi suscité des émotions de valence négative. Nous cherchons donc à classer les interactions enregistrées entre une valence négative contre une valence non négative, afin de simplifier le problème. Cette classification permet d'obtenir des classes avec un nombre d'échantillons similaires.

V.3.2 Classification de données d'interaction avec un smartphone

Dans un premier temps, il est nécessaire d'associer ces labels avec les jeux de données issues des interactions avec le smartphone. Cela comprend les données issues des capteurs inertiels, ainsi que les données issues des événements de frappe au clavier.

Nous procédons à la classification en fonction de la valence déclarée sur la GEW. Nous réalisons un modèle général ou inter-personnel dans un premier temps. Les modèles inter-personnels utilisent les données de tous les sujets, sauf un, pour entraîner les modèles de classification. Les échantillons sont divisés en deux classes, tels que présentés dans la sous-section précédente : valence négative et valence non négative. L'objectif est de classer des échantillons avec au moins un libellé de valence négative déclaré contre des échantillons de tâches sans libellés négatifs déclarés. Cela correspond à 5502 échantillons de la classe négative contre 6018 échantillons de la classe non négative. Le tableau V.8 représente les exactitudes moyennes pour l'ensemble des 40 itérations réalisées, une pour chaque participant. Le modèle de référence s'établit 51,46 %, correspondant à la classification de tous les échantillons dans la classe dominante (à savoir la classe non négative). L'arbre de décision atteint une moyenne d'exactitude de 5,18% au-dessus du modèle de référence. Les autres modèles parviennent à obtenir une exactitude supérieure d'au moins 1,22 % de l'exactitude du modèle de référence.

Une forte variabilité est présente entre les participants. Pour 19 participants, nous obtenons un modèle avec une exactitude supérieure à 10% au modèle de référence de l'itération. Ainsi, les modèles inter-individuels ne semblent pas être en mesure de se généraliser sur l'ensemble de la population, à partir de ces données.

Nous testons ensuite la classification de valence avec des modèles intra-individuels. Nous utilisons donc les échantillons de tâches de description d'un même individu, labellisées par des libellés de valence négative contre des déclarations sans libellés de valence négative. Les modèles sont ensuite testés sur d'autres tâches, selon plusieurs itérations pour obtenir un résultat moyen par participant. Nous obtenons donc une exactitude moyenne pour chaque participant. Certains participants ne sont pas sélectionnés à cause du manque du nombre de tâches minimum associé à chaque classe pour réaliser l'apprentissage. Nous conservons ainsi 38 participants. Nous calculons

TABLE V.8 – Modèles inter-personnels sur les données d’interaction : moyenne d’exactitude pour la classification de la valence (négative vs non négative)

Modèles	Exactitude moyenne	Différence avec la référence
Modèle de référence	51,46 %	0 %
AdaBoost	53,56 %	2,10 %
Arbre de décision	56,64 %	5,18 %
Classifieur à renforcement de gradient	53,93 %	2,47 %
K plus proche voisin	56,73 %	5,27 %
Machines à vecteur de support	53,93 %	2,48 %
Réseaux de neurones	54,88 %	3,43 %
Régression logistique	53,52 %	2,06 %
Réseaux bayésien	52,69 %	1,22 %

ensuite la moyenne d’exactitude pour tous les participants et pour chaque modèle. Le tableau V.9 représente les résultats obtenus. La moyenne de l’exactitude du modèle de référence est de 56,07%. En moyenne, trois modèles surpassent l’exactitude du modèle de référence de plus de 10% : le classifieur à renforcement de gradient avec 69,89 % d’exactitude, l’adaboost avec 69,35 % et l’arbre de décision avec 67,73 %. Les résultats des modèles intra-individuels sont donc supérieurs par rapport aux modèles de classification inter-individuels.

À l’échelle individuelle, 19 participants ont des modèles de classification en mesure de dépasser l’exactitude de leur modèle de référence de plus de 10%, sur les 38 participants testés. Comme pour les modèles inter-individuels, il existe donc une forte variabilité individuelle pour la performance des modèles de classification. La figure V.16 représente un exemple de matrices de confusion pour un des 19 sujets cités précédemment. Il s’agit donc d’un participant pour lequel l’entraînement a permis d’obtenir des modèles avec des performances supérieures au modèle de référence. Le modèle de référence (à gauche de la figure V.16), classant tous les échantillons dans la classe dominante (labels non négatifs), possède une exactitude de 63,13 %. Nous pouvons donc noter un déséquilibre dans les classes, avec une majorité d’échantillons labellisés comme ayant une valence non négative. Cela indique que ce participant a déclaré moins de tâches réalisées avec une valence négative. Le modèle des machines à vecteur de support (à droite de la figure V.16) atteint une exactitude de 76,26 %, soit 13,13 % de plus que le modèle de référence. Visuellement, nous observons un déséquilibre dans la classification. Avec un rappel à 63,26 % et une précision à 84,93 % (voir sous-section IV.6.6.4), ce modèle est moins fiable pour prédire correctement les échantillons labellisés comme non négatifs pour cet individu.

TABLE V.9 – Modèles intra-personnels sur les données d’interaction : moyenne d’exactitude pour la classification de la valence (négative vs non négative)

Modèles	Exactitude moyenne	Différence avec la référence
Modèle de référence	56,04 %	0 %
AdaBoost	69,35 %	13,31 %
Arbre de décision	67,73 %	11,69 %
Classifieur à renforcement de gradient	69,89 %	13,85 %
K plus proche voisin	61,40 %	5,36 %
Machines à vecteur de support	65,31 %	8,82 %
Réseaux de neurones	65,02 %	9,27 %
Régression logistique	65,89 %	9,85 %
Réseaux bayésien	63,36 %	7,79 %

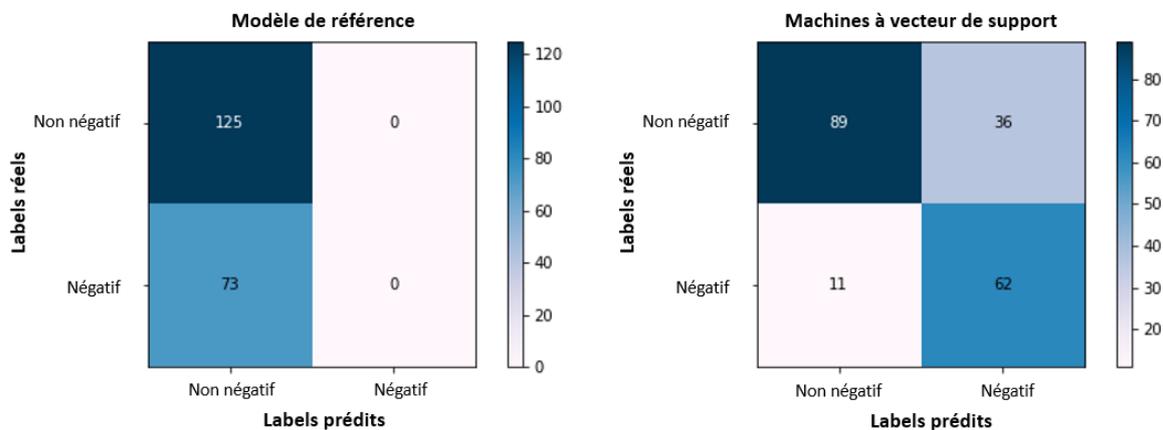


FIGURE V.16 – Matrices de confusion pour un sujet sélectionné sur les données d’interaction avec un smartphone

De plus, les descripteurs les plus corrélés à la labellisation de la valence sont aussi enregistrés pour chaque participant. Nous avons représenté les dix descripteurs les plus corrélés avec la valence pour le participant sélectionné précédemment (voir le tableau V.10). Nous observons une forte présence de descripteurs issus des signaux de saccades et d’ondelettes, particulièrement pour les enregistrements du gyromètre et de l’accéléromètre. Cependant, aucun descripteur issu de la frappe au clavier n’est présent dans les 100 descripteurs les plus corrélés pour ce sujet.

TABLE V.10 – Liste des dix descripteurs les plus corrélés à la variable dépendante

Nom du descripteurs	Capteur / Signal	Détails	Coefficient de corrélation (Pearson)
<i>Gyr_jerk_t_xy_corr</i>	Gyromètre / Saccade (temporelle)	Corrélation des axes x et y	0,611
<i>Gyr_dwt_y_energy_4</i>	Gyromètre / Ondelettes	Énergie fréquentielle de la bande 4 de l'axe y	0,594
<i>Gyr_jerk_f_yz_corr</i>	Gyromètre / Saccade (fréquentielle)	Corrélation des axes y et z	0,587
<i>Acc_dwt_y_energy_3</i>	Accéléromètre / Ondelettes	Énergie fréquentielle de la bande 4 de l'axe y	0,517
<i>Acc_f_yz_corr</i>	Accéléromètre / Signal fréquentiel	Corrélation des axes y et z	0,508
<i>Acc_dwt_y_energy_2</i>	Accéléromètre / Ondelettes	Énergie fréquentielle de la bande 2 de l'axe y	0,505
<i>Acc_jerk_f_skewness</i>	Accéléromètre / Saccade	Asymétrie de l'axe z	0,504
<i>Gyr_jerk_t_y_mad</i>	Gyromètre / Saccade (temporelle)	Médiane absolue des écarts de l'axe y	0,470
<i>Mag_jerk_f_z_entropy</i>	Magnétomètre / Saccade (fréquentielle)	Entropie de l'axe z	0,427
<i>Gyr_jerk_t_xz_corr</i>	Gyromètre / Saccade (temporelle)	Corrélation des axes x et z	0,420

Les matrices de confusion de chaque modèle et les descripteurs les plus informatifs sont enregistrés pour chaque participant et permettent de mieux comprendre le poids de chaque descripteur pour les modèles de classification.

V.3.3 Classification avec les données de la montre Empatica E4

Afin d'approfondir les résultats, nous réalisons des classifications de la valence à partir des données issues de la montre Empatica E4. Nous nous intéressons particulièrement aux modèles intra-individuels, car ils présentent de meilleures performances sur les données d'interaction avec un smartphone. À cause d'enregistrements incomplets pour certains participants, nous n'entraînons les modèles que pour 34 participants. Nous réalisons des modèles intra-individuels pour comparer les exactitudes des modèles de classification entraînés à partir des données de la montre Empatica E4. Ces données incluent les enregistrements de capteurs physiologiques et d'un capteur d'accéléromètre. Les résultats d'exactitude sont représentés dans le tableau V.11. Le modèle de

référence possède une exactitude de 55,77%. Tous les modèles obtiennent en moyenne un score supérieur de 10% au modèle de référence, sauf l'arbre de décision. Le modèle le plus performant est la régression logistique avec une exactitude moyenne de 72,89 %. En moyenne, les performances des classifications réalisées à partir des données de la montre Empatica E4 sont supérieures par rapport aux classifications réalisées à partir des données de smartphone.

TABLE V.11 – Modèles intra-personnels sur les données physiologiques : moyenne d'exactitude pour la classification de la valence (négative vs non négative)

Modèles	Exactitude moyenne	Différence avec la référence
Modèle de référence	55,77 %	0 %
AdaBoost	70,71 %	14,94 %
Arbre de décision	64,15 %	8,38 %
Classifieur à renforcement de gradient	68,51 %	12,74 %
K plus proche voisin	68,88 %	13,11 %
Machines à vecteur de support	70,81 %	15,04 %
Réseaux de neurones	70,45 %	14,67 %
Régression logistique	72,89 %	17,12 %
Réseaux bayésien	67,94 %	12,16 %

Nous observons aussi une variabilité entre les sujets, pour les données issues de la montre Empatica E4. En effet, 26 participants sur les 34 ont au moins un modèle de classification avec une exactitude supérieure de 10% au modèle de référence. En comparant avec les résultats obtenus avec les données d'interaction sur le smartphone, nous avons identifié les profils suivants :

- 13 participants ont des modèles de classification de valence avec une exactitude supérieure de 10% à leur modèle de référence, à la fois pour les données d'interaction smartphone et les données issues de la montre Empatica E4.
- 13 participants ont des modèles de classification de valence avec une exactitude supérieure de 10% à leur modèle de référence, seulement pour les données issues de la montre Empatica E4.
- 6 participants ont des modèles de classification de valence avec une exactitude supérieure de 10% à leur modèle de référence, seulement pour les données d'interaction de smartphone.

La figure V.17 représente les matrices de confusion pour le même participant sélectionné dans la sous-section précédente, pour comparer les résultats sur les données de la montre Empatica E4, avec les mêmes modèles de classification. Pour ce participant et pour ce modèle, la classification a été légèrement moins performante avec les données de la montre. En effet, le modèle des machines à vecteur de support obtient une exactitude de 67,62 % (à droite de la figure V.17) contre 62,85 % pour le modèle de référence (à gauche de la figure V.17). À l’image de la classification à partir des données de smartphone, nous observons un rappel plus élevé par rapport à la précision, indiquant une plus faible fiabilité pour prédire correctement la classe des échantillons de valence non négative. Il est à noter que d’autres modèles de classification ont atteint des performances plus élevées pour ce participant et pour ces données (78,77 % pour le modèle par arbre de décision).

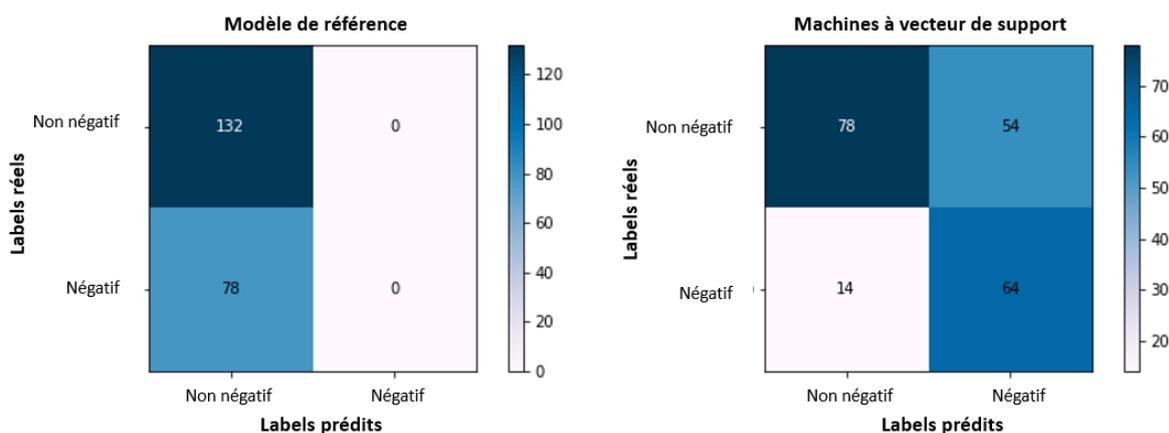


FIGURE V.17 – Matrices de confusion pour un sujet sur les données d’Empatica E4

Le tableau V.12 représente les descripteurs de la montre Empatica E4, les plus corrélés avec la labellisation pour le participant sélectionné. Nous voyons une forte corrélation pour les signaux de conductance électrodermale et de la variation de la fréquence cardiaque (le calcul des descripteurs est disponible dans la sous-section IV.6.5.4). Des descripteurs issus du signal de saccade d’accélération figurent aussi parmi les plus corrélés avec la valence. Le tableau V.12 représente les dix descripteurs les plus corrélés à la variable dépendante pour les enregistrements de la montre Empatica E4. Les descripteurs issus du capteur de la conductance électrodermale sont les plus corrélés, ainsi que la variation de la fréquence cardiaque, calculée à partir de la pulsation cardiaque. L’accéléromètre de la montre fournit aussi plusieurs descripteurs parmi les plus corrélés avec la valence pour ce participant.

TABLE V.12 – Liste des dix descripteurs à partir des données fusionnées les plus corrélés à la variable dépendante

Nom du descripteurs	Capteur / Signal	Détails	Coefficient de corrélation (Pearson)
<i>Eda_scl_eda_max</i>	Conductance électrodermale	Maximum du niveau de conductance	0,672
<i>Eda_scr_eda_std</i>	Conductance électrodermale	Écart-type de la réponse de conductance	0,417
<i>Hrv_vlf_max_power</i>	Variation de la fréquence cardiaque	Maximum de la bande de fréquence très basses	0,358
<i>Acc_jerk_y_f_x_max</i>	Accéléromètre / Saccade (fréquentielle)	Maximum de l'axe x	0,344
<i>Acc_jerk_t_z_iqr</i>	Accéléromètre / Saccade (temporelle)	Écart inter-quartile de l'axe z	0,299
<i>Acc_f_xz_corr</i>	Accéléromètre / Signal fréquentiel	Corrélation des axes x et z	0,290
<i>Acc_jerk_t_y_iqr</i>	Accéléromètre / Saccade (temporelle)	Écart inter-quartile de l'axe y	0,288
<i>Acc_t_x_max</i>	Accéléromètre / Signal temporel	Asymétrie de l'axe z	0,273
<i>Eda_scl_std</i>	Conductance électrodermale	Écart-type du niveau de conductance	0,253
<i>Hrv_vlf_min</i>	Variation de la fréquence cardiaque	Minimum de la bande de fréquence très basses	0,251

V.3.4 Classification avec fusion des modalités

Afin de tester la capacité du système de reconnaissance d'émotions pour classer les modalités de valence à partir de sources de données distinctes, nous réalisons des classifications de la valence (négative contre non négative) en combinant les données issues des interactions avec le smartphone et de la montre Empatica E4. Les modèles sont entraînés pour 34 participants. Les résultats sont visibles sur le tableau V.13. En moyenne, le modèle de référence possède une exactitude de 55,96 %. Nous pouvons observer une légère amélioration de l'exactitude pour certains modèles, par rapport aux seules données physiologiques. Ainsi, le modèle AdaBoost avec une exactitude de 73,62 % dépasse le modèle de référence de 17,66%. Sept des huit modèles dépassent de plus de 10% d'exactitude le modèle de référence.

TABLE V.13 – Modèles intra-personnels sur les données fusionnées : moyenne d’exactitude pour la classification de la valence (négative vs non-négative)

Modèles	Exactitude moyenne	Différence avec la référence
Modèle de référence	55,96 %	0 %
AdaBoost	73,62 %	17,66 %
Arbre de décision	71,83 %	15,87 %
Classifieur à renforcement de gradient	73,02 %	17,06 %
K plus proche voisin	63,83 %	7,87 %
Machines à vecteur de support	68,82 %	12,86 %
Réseaux de neurones	67,80 %	11,85 %
Régression logistique	69,96 %	14,01 %
Réseaux bayésien	67,35 %	11,39 %

Afin de comparer avec les classifications précédentes, nous représentons les matrices de confusion du participant sélectionné précédemment. Les matrices de confusion de référence et de machines à vecteur de support sont représentées en figure V.18. Le modèle de référence présente une exactitude de 63,13 % (à gauche de la figure V.18) et le modèle de machines à vecteur de support possède une exactitude de 78,28 % (à droite de la figure V.18). Nous obtenons une exactitude supérieure de 15,15 %, légèrement supérieure à l’exactitude obtenue seulement avec les données de smartphone.

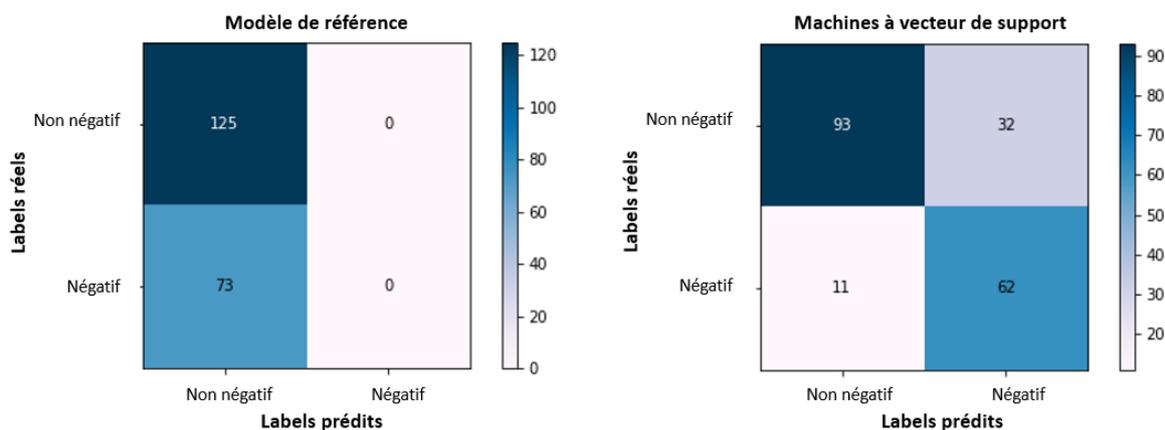


FIGURE V.18 – Matrices de confusion pour un sujet à partir des données fusionnées

Le tableau V.14 représente les descripteurs les plus corrélés avec la valence. Pour ce

participant, un descripteur de conductance électrodermale figure comme le descripteur le plus corrélé, suivi par des descripteurs uniquement calculés à partir des capteurs inertiels du smartphone.

TABLE V.14 – Liste des dix descripteurs de la montre Empatica E4 les plus corrélés à la variable dépendante

Nom du descripteurs	Capteur / Signal	Détails	Coefficient de corrélation (Pearson)
<i>Eda_scl_eda_max</i>	Conductance électrodermale	Maximum du niveau de conductance	0,689
<i>Gyr_jerk_t_xy_corr</i>	Gyroscope / Saccade (temporelle)	Corrélation des axes x et y	0,588
<i>Gyr_dwt_y_energy_4</i>	Gyromètre / Ondelettes	Énergie fréquentielle de la bande 4 de l'axe y	0,585
<i>Gyr_jerk_f_yz_corr</i>	Gyromètre / Saccade (fréquentielle)	Corrélation des axes y et z	0,579
<i>Acc_dwt_y_energy_3</i>	Accéléromètre (smartphone) / Ondelettes	Énergie fréquentielle de la bande 3 de l'axe y	0,564
<i>Acc_dwt_y_energy_2</i>	Accéléromètre (smartphone) / Ondelettes	Énergie fréquentielle de la bande 2 de l'axe y	0,514
<i>Acc_jerk_f_z_skewness</i>	Accéléromètre (smartphone) / Saccade (fréquentielle)	Asymétrie de l'axe z	0,489
<i>Acc_f_yz_corr</i>	Accéléromètre (smartphone) / Signal fréquentiel	Corrélation des axes y et z	0,479
<i>Gyr_jerk_t_y_med</i>	Gyroscope / Saccade (temporelle)	Médiane absolue des écarts de l'axe y	0,444
<i>Mag_jerk_f_z_entropy</i>	Magnétomètre / Saccade (fréquentielle)	Entropie de l'axe z	0,437

V.3.5 Discussion

Ces analyses avaient pour objectif de répondre à la problématique générale de cette thèse **les signaux des capteurs inertiels et de la frappe au clavier de smartphone permettent-ils d'obtenir des modèles de classification robustes pour discriminer des dimensions émotionnelles ?**

Premièrement, nous avons exploré la **Q_Interindividu** : les variations inter-individuelles des signaux enregistrés sont-elles assez fiables pour généraliser le processus de détection des émotions à un grand nombre de personnes? Nous avons ainsi formulé une hypothèse générale associée, dans la sous-section IV.4.3. D'une manière générale, avec la labellisation proposée, les modèles généraux ne présentent pas une forte robustesse pour la classification de la valence. Ils ne sont donc pas en mesure de généraliser leurs prédictions à l'ensemble de la population testée. Il existe en revanche une variabilité entre les participants, certains ayant des résultats de classification nettement supérieurs aux autres. En l'état, avec une exactitude moyenne jusqu'à 5,27% supérieure au modèle de référence, nous ne pouvons pas valider l'hypothèse générale associée à la question fondamentale **Q_Interindividu**.

Deuxièmement, nous avons traité la **Q_Intraindividu** : les variations intra-individuelles des signaux enregistrés sont-elles assez fiables pour généraliser le processus de détection des émotions au niveau d'une seule et même personne? De la même manière, nous avons formulé une hypothèse générale associée, dans la sous-section IV.4.3. Comme attendu dans cette hypothèse, les modèles personnalisés pour la classification de la valence se distinguent avec des performances supérieures par rapport aux modèles inter-individuels. La supériorité des modèles individuels par rapport aux modèles généraux est cohérente avec la méta-analyse proposée par Kołakowska *et al.* (2020). Ainsi, les modèles intra-personnels s'adaptent mieux aux caractéristiques individuelles et parviennent donc mieux à distinguer des descripteurs discriminants pour chaque participant. Cependant, ils nécessitent une plus grande quantité de données par participant pour pouvoir être mis en place. Avec une exactitude jusqu'à 13,31% supérieure au modèle de référence en moyenne, l'hypothèse générale ne peut être complètement validée, mais apporte une perspective de réflexion essentielle.

Enfin, afin de tester le système de classification et de comparer les résultats avec une autre source de données, nous avons étudié la **question fondamentale Q_Equivalence_Physio** : Les signaux issus de capteurs physiologiques permettent-ils d'obtenir des robustesses de classification similaires aux modèles de classification issus des capteurs de smartphone? Les résultats montrent des performances légèrement supérieures pour des modèles personnalisés entraînés avec des données de la montre Empatica E4, principalement composée de capteurs physiologiques. Nous pouvons en partie valider l'hypothèse proposée en sous-section IV.4.3, avec des résultats de classification similaire en moyenne entre les données issues des interactions avec un smartphone de la montre Empatica E4. Plus spécifiquement, un certain nombre de participants obtiennent des résultats de classification supérieurs au hasard pour les deux modalités.

Dans tous les types de classification, nous notons une forte variabilité entre les individus pour la performance obtenue. Plusieurs éléments de réponse peuvent nous

aider à comprendre ce résultat. Ainsi, l'induction d'émotions n'a pas nécessairement fonctionné avec la même efficacité pour chaque individu. Selon Gil (2009) et R. A. Calvo et D'Mello (2010), les méthodes d'induction d'émotions en laboratoire souffrent du manque d'écologie, bien que les extraits vidéo soient considérés comme une méthode puissante. Ainsi, l'expression d'émotion a pu rester limitée et être peu détectable au travers des interactions avec le smartphone ou par l'intermédiaire des capteurs physiologiques. C'est pourquoi il nous paraît essentiel de tester plusieurs approches pour induire des émotions en laboratoire, tel que des protocoles de manipulation de la difficulté pour impliquer plus directement les sujets.

Une problématique essentielle porte sur l'évaluation de l'efficacité de l'induction d'émotion pour chaque sujet. Nous avons ainsi inclus des mesures physiologiques, en plus des mesures subjectives pour augmenter le nombre de composantes prises en compte. Il serait envisageable d'inclure d'autres modalités pour améliorer la prise en compte des composantes des émotions exprimées. Les expressions faciales ou les postures des participants pourraient être analysées à partir des enregistrements vidéo, pour mieux identifier la composante de l'expression motrice. En revanche, l'intégration de nombreuses modalités représente une forte complexité méthodologique et technique pour traiter des sources de données diverses. D'une part, l'enregistrement des données ne doit pas gêner les sujets et permettre d'obtenir des données de bonne qualité et d'autre part, les données doivent ensuite être assemblées afin de construire une vue d'ensemble de l'expression émotionnelle des participants. Selon les problématiques proposées par R. A. Calvo et D'Mello (2010), explicitées en sous-section III.3.7, il n'est pas évident que la mesure de l'évolution d'une composante implique des évolutions mesurables et synchronisées sur les autres composantes. Ainsi, nous avons pu observer pour certains individus que certaines sources de données semblaient plus efficaces que d'autres. La fusion des données issues du smartphone et de la montre Empatica E4 a permis une légère augmentation des résultats. Cependant, les composantes émotionnelles n'évoluent donc pas nécessairement de manière synchronisée et les mesures actuelles et leur traitement ne sont pas forcément adaptés.

Le choix des labels représente aussi une difficulté. Dans notre situation, l'analyse des mesures obtenues par la GEW a guidé notre labellisation afin d'obtenir des classes de données pertinentes et équilibrées, pour la dimension de la valence. Nous pouvons imaginer établir des classifications avec d'autres caractéristiques de la GEW telles que le niveau d'intensité ou le nombre d'émotions déclarées, voire des libellés plus spécifiques. Cependant, cette labellisation des données n'est pas forcément adaptée pour identifier avec détail les expressions d'émotions des individus. Nous pouvons donc nous poser la question suivante : comment identifier des expressions d'émotions particulières sans utiliser de mesures subjectives ? Nous pouvons envisager de détecter et rassembler des patterns d'expression similaire par l'intermédiaire de classification non supervisée, afin de labelliser des échantillons d'enregistrements.

Sur l'aspect technique, plusieurs choix de traitements des données (décrit dans la section IV.6.2) ont pu influencer sur les performances des modèles de classification. La taille de la fenêtre d'échantillonnage et le pourcentage de superposition peuvent être considérés comme des paramètres significatifs. Le choix des descripteurs est aussi déterminant, surtout au regard de la grande variété de descripteurs générés. En enregistrant les descripteurs les plus corrélés aux variables dépendantes (les labels), nous gardons une trace des descripteurs les plus utiles aux modèles, afin de sélectionner les plus pertinents pour entraîner des modèles dans le futur. La solution d'employer des modèles d'apprentissage profond est aussi envisageable pour générer automatiquement des descripteurs sur ce type d'enregistrements. De plus, en ce qui concerne la fusion des données, l'utilisation d'une approche de fusion au niveau décisionnel (voir sous-section III.3.7) pourrait améliorer la performance de classification multimodale, avec des modèles de classification adaptés à chaque modalité. Cela implique le développement d'une nouvelle couche décisionnelle pour le système de reconnaissance automatique.

En synthèse, des problématiques se dessinent sur le profil des sujets et la variabilité des résultats. En effet, existe-t-il des profils d'utilisateurs plus sensibles à l'induction d'émotions d'une part et existe-t-il des profils d'utilisateurs exprimant des épisodes émotionnels avec plus d'intensité d'autre part? Si cela est le cas, serait-il envisageable de concevoir un système pour distinguer les profils les plus « émotionnels »? Nous avons commencé à explorer ces problématiques en intégrant un questionnaire d'impulsivité émotionnelle. Cependant, il serait envisageable de concevoir des expérimentations avec des questionnaires plus précis sur la personnalité des sujets, permettant de détecter des profils particuliers, afin d'améliorer les systèmes de reconnaissance d'émotions. Ainsi, la prise en compte du profil de l'utilisateur pourrait affiner le système de reconnaissance automatique en ajustant les modèles de classification en fonction des caractéristiques individuelles des participants.

V.4 Conclusion

Ce chapitre s'est centré sur l'analyse des données obtenues à la suite des expérimentations réalisées et décrites dans le chapitre précédent. Dans un premier temps, nous avons analysé les mesures subjectives, afin d'estimer l'effet des protocoles et des conditions expérimentales sur les sujets. Cela a mis en évidence, d'une part, le rôle du matériel utilisé, des activités proposées et de l'équilibrage et d'autre part l'influence déterminante de l'instrument de mesure. La pré-expérimentation a mis en lumière le peu d'effets mesurables du protocole expérimental proposé, sauf concernant le nombre de réponses correctes pour le test de mémoire et d'attention. L'instrument de mesure SAM n'a pas permis d'affiner notre analyse de manière plus détaillée. En revanche, la *Geneva Emotion Wheel*, au travers la version que nous avons développée sur smartphone, a permis de mettre en évidence des effets plus marqués et de réaliser

des analyses plus approfondies des expérimentations principale et complémentaire. Cet instrument de mesure du sentiment subjectif permet à la fois de résumer des effets des conditions expérimentales sur des dimensions synthétiques (valence, intensité ...), mais aussi d'étudier plus en détail les libellés déclarés. Ainsi, nous avons pu montrer l'efficacité des protocoles de l'expérimentation principale et de l'expérimentation complémentaire. De plus, par l'intermédiaire de cet outil nous avons différencié la nature des émotions déclarées entre les deux protocoles.

L'application du système de classification automatique pour catégoriser les données, selon la valence déclarée par les participants (valence négative contre non négative), montre la complexité d'établir des modèles robustes avec les données actuelles. Les modèles de classification automatique intra-individuels présentent des résultats plus performants par rapport aux modèles inter-individuels, ce qui montre l'importance de prendre en compte les caractéristiques de chaque individu. En effet, nous observons dans l'ensemble, une forte variabilité entre les sujets de l'expérimentation. De manière plus approfondie, en comparant des classifications issues des données d'interaction avec le smartphone ou de la montre Empatica E4, certains individus ont des performances supérieures sur certaines sources de données plutôt que sur d'autres. Cela démontre la complexité d'associer des données multimodales et par extension de mesurer l'évolution de plusieurs composantes d'un épisode émotionnel.

L'ensemble des analyses réalisées dans ce chapitre montrent la complexité globale en informatique affective pour développer des systèmes de reconnaissance automatique d'émotions. Les résultats nous permettent d'envisager de nombreuses perspectives et pistes d'amélioration à la fois d'un point de vue méthodologique et d'un point de vue technique.

Chapitre VI

Conclusion

Sommaire

VI.1 Synthèse et apports des recherches	211
VI.2 Réflexions et perspectives	214

VI.1 Synthèse et apports des recherches

L'ensemble de nos travaux de recherche a eu pour objectif de répondre à la problématique générale suivante : comment reconnaître automatiquement des états émotionnels exprimés par un utilisateur, lors d'une situation de communication sur smartphone ? L'objectif sur le long terme est d'intégrer ce système au contexte du service client, dans une entreprise de télécommunication. Pour répondre à cette problématique, nous avons réalisé les actions suivantes :

- L'assemblage d'un ensemble de connaissances théoriques, méthodologiques et techniques.
- Le développement d'outils pour collecter des données labellisées, par l'intermédiaire d'expérimentations en laboratoire, afin de rassembler un corpus de données.
- Le développement et l'application d'un système de reconnaissance automatique d'émotions sur les données recueillies, avec l'objectif de traiter les données et d'entraîner des modèles de classification par apprentissage automatique et de les évaluer.

D'un point de vue théorique, une revue en science des émotions nous a permis de mettre en lumière les débats et les recherches sur la nature des émotions. Nous avons fait le choix d'utiliser la théorie de l'évaluation cognitive et sa définition des émotions, qui nous paraît pertinente pour des situations de communication. Le choix du modèle

des processus-composants a apporté une approche pragmatique des processus émotionnels, intégrant cinq sous-systèmes ou composantes. Enfin, le cadre théorique de cette thèse a été complété par une revue des représentations des émotions, dimensionnelles et catégorielles.

Le recours à l'informatique affective nous a apporté des éléments pour prendre en compte les émotions d'utilisateurs dans le contexte de l'usage des TIC. Nous considérons que l'approche de l'informatique affective peut s'adapter à des situations de communication médiées par les TIC et plus particulièrement par les smartphones. Il est également pertinent d'appliquer cette discipline au contexte du service client, particulièrement concerné par les transformations numériques. En outre, le domaine du service client n'est pas à ce jour un objet d'étude courant pour l'informatique affective, malgré des situations de communication émotionnelles et un fort usage de nouvelles technologies.

Afin d'étudier des états émotionnels avec un système de reconnaissance automatique, nous avons examiné les sources de données crédibles dans la littérature. Cela a mis en lumière les avantages et inconvénients de chaque modalité (visuelle, audio, textuelle...). Les données liées aux interactions avec des interfaces homme-machine d'ordinateurs et de smartphones ont semblé pertinentes pour rendre compte des comportements des utilisateurs. Plus spécifiquement, notre choix s'est porté sur l'enregistrement de capteurs inertiels et de la frappe au clavier. En effet, ces données présentent l'avantage de pouvoir être collectées dans un usage habituel de smartphone sans l'ajout de capteurs invasifs. De plus, ces données présentent une plus faible criticité pour la vie privée et l'identification des utilisateurs, par rapport aux modalités visuelle et audio.

Dans l'optique de réaliser un corpus à partir des interactions sur smartphone, nous avons listé des approches pour collecter des données labellisées selon des états émotionnels. Nous nous sommes positionnés sur une approche fondée sur l'induction d'émotions par l'intermédiaire de médias et de tâches à réaliser. Un travail de développement particulier s'est concentré sur la création des outils appropriés pour s'adapter au contexte de smartphone. Ainsi, nous avons conçu un support d'expérimentation flexible, sous la forme d'une application Android de messagerie instantanée, avec un agent conversationnel. Le recours à un agent conversationnel constitue une nouveauté dans cette discipline et pour ce type de données et représente une manière innovante pour interagir avec des sujets. Cette application est un des apports majeurs de notre travail de recherche et a été le support des expérimentations réalisées. Elle a servi à constituer les bases de données nécessaires, en enregistrant les signaux des capteurs inertiels et de la frappe au clavier, associés à des instruments de mesures subjectives.

Nous pouvons aussi noter la création de l'application GEWMI, afin d'adapter la *Geneva Emotion Wheel* dans un format numérique. Ce développement a fait l'objet

d'intégration dans les expérimentations réalisées. Cela a aussi mené à la création d'une application indépendante, destinée à être disponible en *open source* pour l'usage d'autres expérimentateurs.

Trois protocoles de tests distincts ont été réalisés, afin de tester différentes approches d'induction d'émotions sur smartphone.

- Une pré-expérimentation : pour tester l'effet d'extraits vidéo et de tests de mémoire, selon les dimensions de la valence et de l'activation. Cette première expérience a aussi permis de confronter l'ergonomie et l'usage de l'application d'expérimentation, afin d'apporter des améliorations.
- Une expérimentation principale : conçue à partir d'une nouvelle version de l'application mobile. Cette expérimentation intègre des extraits vidéo sur la dimension de la valence, avec une activité de description pour enregistrer des données de frappe au clavier. Nous avons inclus l'instrument GEWMI dans ce scénario de test, afin de labelliser les données enregistrées pour la création d'un corpus. De plus, nous avons intégré des mesures complémentaires avec l'utilisation de la montre Empatica E4, intégrant des mesures physiologiques et motrices.
- Une expérimentation complémentaire : pour tester un protocole impliquant plus directement les performances des sujets, basé sur une procédure de manipulation de la difficulté.

Bien que les protocoles expérimentaux soient encore éloignés de situations de communication écologiques et naturelles, ils ont permis de tester des méthodes d'induction dans un format se rapprochant d'une application de messagerie instantanée.

L'analyse des données a notamment porté sur des auto-évaluations, qui constituent les mesures les plus accessibles pour quantifier l'induction d'émotions. Les analyses ont mis en avant la complexité de l'induction d'émotions dans la pré-expérimentation et le manque de granularité de l'outil de mesure SAM. Des résultats plus concrets ont été obtenus par l'intermédiaire de GEWMI, qui ont permis d'une part d'étudier les effets des autres protocoles expérimentaux sur des mesures synthétiques et d'autre part, de détailler les émotions déclarées par les libellés. GEW combine ainsi les modèles de représentation dimensionnelle et catégorielle. Les libellés sélectionnés peuvent, par exemple, être mis en correspondance avec des modèles de représentation issus des théories de l'évaluation cognitive, tels que le modèle OCC.

Enfin, un système de reconnaissance automatique des émotions a été développé. Le système peut traiter des enregistrements de séries temporelles issues de capteurs inertiels, d'évènements de frappe au clavier ou de capteurs physiologiques. Ainsi, nous pouvons étudier plusieurs composantes émotionnelles par l'intermédiaire de cet outil. Les signaux sont filtrés et traités afin de générer des matrices de descripteurs, exploitables pour des modèles de classification par apprentissage automatique. Le système a

été testé pour comparer des phases de description et des phases de visionnage. Il en résulte que l'application du système à des matrices de descripteurs labellisées à partir de la valence déclarée sur la GEW offre pour l'instant des résultats limités, pour des données d'interaction sur smartphone. Nous avons pu noter que les modèles de classification intra-personnels apparaissaient plus robustes. En effet, ils possèdent une exactitude pouvant atteindre 13,85 % de plus que le modèle de référence en moyenne, contre 5,18 % pour les modèles inter-personnels. En moyenne, nous avons obtenu des résultats comparables pour des données issues de la montre Empatica E4. La comparaison entre les deux sources de données met en évidence la difficulté d'assembler des données provenant de plusieurs composantes émotionnelles. De nombreux facteurs ont pu influencer sur les performances des modèles de classification :

- L'efficacité de l'induction d'émotions.
- La fiabilité des mesures, avec les capteurs utilisés.
- Les nombreux paramétrages envisageables pour le traitement des données et la génération de descripteurs.

L'ensemble des réalisations mettent en relief la complexité de concevoir des systèmes automatiques robustes de reconnaissance d'émotions. Cependant, nos travaux apportent des solutions concrètes sur des aspects théoriques, méthodologiques et techniques pour répondre à la problématique générale.

VI.2 Réflexions et perspectives

L'ensemble des résultats obtenus permettent d'envisager de nombreuses pistes de réflexion.

Sur le plan méthodologique, la réalisation d'expérimentations sur smartphone est une pratique encore récente. L'utilisation d'agents conversationnels pour soumettre des tests de manière automatique offre des perspectives intéressantes pour limiter des biais liés à la présence d'un expérimentateur humain. Il est envisageable de concevoir de nombreux types de protocoles expérimentaux par l'intermédiaire de ce support. Nous pouvons imaginer inclure des agents conversationnels, non scriptés, utilisant du TAL, afin d'interagir avec des utilisateurs de manière plus naturelle. De plus, l'utilisation d'applications pour smartphone permet d'atteindre un plus large public avec une plus grande facilité de diffusion. Néanmoins, il est nécessaire de prendre en compte les différences inter-individuelles en matière de familiarité avec l'usage des smartphones.

L'induction d'émotions représente un point particulièrement complexe à relever en science des émotions. La littérature a montré une certaine efficacité dans l'utilisation de médias, tels que des images ou des vidéos. Toutefois, nous pouvons noter la passivité

des sujets dans ces cas de figure. L'engagement des utilisateurs par des protocoles basés sur les performances avec la manipulation de la difficulté représente pour nous une perspective pertinente, en informatique affective. Des jeux plus élaborés peuvent être conçus, sur des smartphones ou d'autres supports. L'idée serait d'impliquer les sujets et de provoquer des situations de valence négative ou positive, en créant des conditions expérimentales adaptées.

Il nous semble aussi essentiel d'approfondir ces recherches avec des expérimentations plus écologiques et naturelles, mettant en œuvre des situations de communication proches du contexte du service client. Nous pouvons imaginer des études en laboratoire reprenant des scénarios comparables à des interactions de service client, à condition de pouvoir suffisamment engager les sujets. En outre, il nous semble important d'établir des études de terrain avec des collectes de données en situation réelle. Cela permettrait la création de bases de données pertinentes pour ce domaine de recherche. Les principales difficultés résident dans l'accès à des données diversifiées pour représenter un maximum de composantes émotionnelles et pour la labellisation des enregistrements avec des aspects émotionnels. Sur des communications par messagerie instantanée, il est envisageable d'utiliser des méthodes de TAL pour extraire le contexte et la polarité du contenu textuel. Des interactions écologiques impliquent le recueil des données avec une plus faible maîtrise des conditions, augmentant la complexité de traiter de tels enregistrements. Cependant, ce palier devra être dépassé à l'avenir, si nous voulons rendre des systèmes de reconnaissance automatique d'émotions applicables à des contextes concrets, avec une robustesse satisfaisante.

Toujours sur le plan méthodologique, nous avons pu mettre en évidence l'importance des outils de mesure du sentiment subjectif. La GEW et notre adaptation GEWMI ont montré à quel point elles pouvaient faire ressortir les différences de contexte, avec les libellés sélectionnés par les participants. En effet, nous savons que le contexte joue un rôle prépondérant dans l'expression des émotions, cela est complètement intégré à la théorie de l'évaluation cognitive au travers des composantes cognitive et motivationnelle. Ainsi, nous avons utilisé la version générique de la roue des émotions dans nos expérimentations. Cependant, cette version générique n'est pas nécessairement adaptée à des situations de communication et plus particulièrement à des contextes de relation client. Il est nécessaire de concevoir des mesures subjectives adaptées à la situation. La méthodologie de création de la GEW explicitée par ses créateurs permet, par exemple, de concevoir des roues plus adaptées aux contextes (Scherer *et al.*, 2013). L'application GEWMI autorise une mise en application aisée de nouvelles roues afin de tester des contextes diversifiés. Les équipes d'Orange Labs Service ont notamment développé une application web, afin de tester une roue des émotions en boutique.

Sur l'aspect technique, nous avons mis en évidence la plus grande robustesse des modèles personnalisés par rapport aux modèles généraux. Ces résultats sont cohérents avec ceux de la littérature. Cependant, l'approche personnalisée nécessite d'obtenir une

quantité importante de données par participant. D'un point de vue opérationnel, il peut alors être complexe d'enregistrer autant de données pour chaque utilisateur. Il semble alors pertinent de créer des modèles hybrides, pré-entraînés sur plusieurs individus, tout en étant en mesure d'apprendre à partir des spécificités d'un individu. Des modèles adaptés à des profils d'utilisateurs types pourraient aussi être développés, en fonction de la personnalité des utilisateurs, par exemple. Ainsi, un premier modèle pourrait déterminer la personnalité des utilisateurs et appliquer le modèle de classification d'émotions le plus approprié. Cela pose, bien entendu, des questions d'éthique sur l'acceptabilité de détecter de telles informations sur les utilisateurs.

La prise en compte de plusieurs composantes émotionnelles constitue un des points les plus complexes à relever en informatique affective. Nous avons pu observer que l'intégration de plusieurs modalités est souhaitable pour améliorer la reconnaissance automatique d'émotions. Cependant, bien que nous ayons exploré l'enregistrement de données associées à l'expression motrice, issues des interactions avec un smartphone, ces seules données ne sont pour l'instant pas suffisantes pour créer des modèles suffisamment robustes. La prise en compte des données physiologiques est pertinente, mais semble peu applicable dans un contexte de service client. Il nous paraît donc important d'approfondir la composante du sentiment subjectif d'une part, mais aussi des composantes cognitives et motivationnelles, essentielles dans le processus émotionnel. Les modalités textuelles constituent donc une piste pertinente à explorer pour obtenir ces informations.

Nous pouvons nous poser la question du rôle d'un système de reconnaissance automatique d'émotions, en imaginant obtenir des modèles de classification suffisamment robustes pour offrir des prédictions fiables sur des données d'interaction. En premier lieu, ils peuvent constituer un outil d'analyse, afin de confronter les états émotionnels reconnus chez les utilisateurs avec les questionnaires de satisfaction proposés habituellement. Cet outil pourrait d'ailleurs aussi servir pour évaluer les émotions des conseillers, afin de déterminer la difficulté des interactions réalisées. De manière plus approfondie, un outil de reconnaissance d'émotions pourrait constituer une source d'information supplémentaire, afin de compenser le manque d'informations non verbales en situation de communication par messagerie instantanée. Ainsi, un tel dispositif pourrait s'intégrer aux outils à disposition des conseillers, pour les aider à s'adapter aux besoins des clients. Un travail d'ergonomie serait alors nécessaire, afin de ne pas augmenter la charge mentale des conseillers clients avec un surplus d'informations. Enfin, l'augmentation de l'usage de *chatbot* pour des services clients constitue une perspective d'application potentielle. L'intégration d'un tel système pourrait aider les agents conversationnels à s'adapter plus précisément aux comportements de leurs interlocuteurs. Ce système pourrait aussi servir à détecter des situations problématiques qu'un *chatbot* ne serait dans l'incapacité de gérer, afin de basculer la communication vers un conseiller humain.

À l'avenir, nous pouvons envisager un intérêt croissant de la recherche sur la reconnaissance d'émotions sur les TIC. Les enjeux dans ce domaine sont encore plus importants depuis la pandémie du Covid-19, qui a fortement affecté notre manière de communiquer, en nous contraignant à un usage plus intensif de TIC. Cela a mis en évidence la difficulté de communiquer nos émotions à distance et l'isolement qu'il en découle, malgré les nombreux canaux à notre disposition. Cette thèse s'inscrit donc dans une évolution globale de notre société, de plus en plus connectée, mais confrontée à la nécessité de mettre l'humain au cœur de la réflexion autour de l'usage de TIC.

Bibliographie

- ALTMAN, N. S. (1992). An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*, 46(3), 175-185 (cf. p. 160).
- ANGUITA, D., GHIO, A., ONETO, L., PARRA, X. & REYES-ORTIZ, J. L. (2013). A public domain dataset for human activity recognition using smartphones. *Esann* (cf. p. 98, 131, 144, 147, 149).
- ARACENA, C., BASTERRECH, S., SNÁEL, V. & VELÁSQUEZ, J. (2015). Neural networks for emotion recognition based on eye tracking data. *2015 IEEE International Conference on Systems, Man, and Cybernetics*, 2632-2637 (cf. p. 85).
- ARNOLD, M. B. (1960). Emotion and personality. (cf. p. 37, 38, 41, 42).
- AUE, T. (2009). Psychophysiologie des émotions (cf. p. 47, 48).
- AUE SEIL, T. (2009). Motivation et tendances à l'action. *Traité de psychologie des émotions* (cf. p. 39, 49, 50).
- AVERILL, J. R. (1985). The social construction of emotion : With special reference to love. *The social construction of the person* (p. 89-109). Springer. (Cf. p. 36).
- BAKKER, I., van der VOORDT, T., VINK, P. & de BOON, J. (2014). Pleasure, arousal, dominance : Mehrabian and Russell revisited. *Current Psychology*, 33(3), 405-421 (cf. p. 63, 65).
- BALDUCCI, F., IMPEDOVO, D., MACCHIARULO, N. & PIRLO, G. (2020). Affective states recognition through touch dynamics. *Multimedia Tools and Applications*, 79(47), 35909-35926 (cf. p. 97).
- BARROS, P., CHURAMANI, N., LAKOMKIN, E., SIQUEIRA, H., SUTHERLAND, A. & WERMTER, S. (2018). The omg-emotion behavior dataset. *2018 International Joint Conference on Neural Networks (IJCNN)*, 1-7 (cf. p. 78).
- BAVEYE, Y., BETTINELLI, J.-N., DELLANDRÉA, E., CHEN, L. & CHAMARET, C. (2013). A large video database for computational models of induced emotion. *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*, 13-18 (cf. p. 105, 123, 124).
- BEAUDOUIN-LAFON, M. (1997). Ingénierie des Systèmes Interactifs. *Cours de Master. LRI-Université Paris-Sud, 11* (cf. p. 91).
- BRADLEY, M. M. & LANG, P. J. (1994). Measuring emotion : the self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry*, 25(1), 49-59 (cf. p. 107, 117, 125).

- BREEDLOVE, S. M., ROSENZWEIG, M. R. & WATSON, N. (2012). *Psychobiologie : de la biologie du neurone aux neurosciences comportementales, cognitives et cliniques*. De Boeck. (Cf. p. 47, 51).
- BREIMAN, L., FRIEDMAN, J. H., OLSHEN, R. A. & STONE, C. J. (1984). Classification and regression trees. Belmont, CA : Wadsworth. *International Group*, 432, 151-166 (cf. p. 160).
- BRÖHL, C., RASCHE, P., JABLONSKI, J., THEIS, S., WILLE, M. & MERTENS, A. (2018). Desktop PC, tablet PC, or smartphone? An analysis of use preferences in daily activities for different technology generations of a worldwide sample. *International Conference on Human Aspects of IT for the Aged Population*, 3-20 (cf. p. 91).
- BUNGENER, C., BONNET, P. & FIORI-DUHARCOURT, N. (2016). Validation de 120 images de l'IAPS auprès d'une population française âgée de 20 à 88 ans. *Gériatrie et Psychologie Neuropsychiatrie du Vieillissement*, 14(2), 201-212 (cf. p. 104).
- CALVO, R., D'AMICO, S., GRATCH, J. & KAPPAS, A. (2015). Introduction to affective computing. *The Oxford handbook of affective computing*, 1-8 (cf. p. 72).
- CALVO, R. A. & D'AMICO, S. (2010). Affect detection : An interdisciplinary review of models, methods, and their applications. *IEEE Transactions on affective computing*, 1(1), 18-37 (cf. p. 88, 197, 208).
- CAMBRIA, E., DAS, D., BANDYOPADHYAY, S. & FERACO, A. (2017). *A practical guide to sentiment analysis* (T. 5). Springer. (Cf. p. 82, 83).
- CANNON, W. B. (1927). The James-Lange theory of emotions : A critical examination and an alternative theory. *The American journal of psychology*, 39(1/4), 106-124 (cf. p. 33, 34).
- CARIDAKIS, G., CASTELLANO, G., KESSOUS, L., RAOUZAIYOU, A., MALATESTA, L., ASTERIADIS, S. & KARPOUZIS, K. (2007). Multimodal emotion recognition from expressive faces, body gestures and speech. *IFIP International Conference on Artificial Intelligence Applications and Innovations*, 375-388 (cf. p. 80).
- CHALMETA, R. (2006). Methodology for customer relationship management. *Journal of systems and software*, 79(7), 1015-1024 (cf. p. 73).
- CHAWLA, N. V., BOWYER, K. W., HALL, L. O. & KEGELMEYER, W. P. (2002). SMOTE : synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16, 321-357 (cf. p. 159).
- CHEN, Y. & SHEN, C. (2017). Performance analysis of smartphone-sensor behavior for human activity recognition. *Ieee Access*, 5, 3095-3110 (cf. p. 149).
- CHRISTOPHE, V. (1998). *Les émotions : tour d'horizon des principales théories* (T. 3). Presses Univ. Septentrion. (Cf. p. 31-37, 39).
- CNTRL. (2012). *Centre National de Ressources Textuelles et Lexicales*. Récupérée 23 mars 2021, à partir de <https://www.cnrtl.fr/>. (Cf. p. 28-30)
- CODISPOTI, M., BRADLEY, M. M. & LANG, P. J. (2001). Affective reactions to briefly presented pictures. *Psychophysiology*, 38(3), 474-478 (cf. p. 104).
- CONNEAU, A.-C. (2016). *Reconnaissance automatique de l'émotion à partir de signaux EEG* (thèse de doct.). Paris, ENST. (Cf. p. 48).

- CORBIN, A., COURTINE, J.-J. & VIGARELLO, G. (2016a). *Histoire des émotions, vol. 1 : De l'Antiquité aux Lumières*. Seuil. (Cf. p. 29, 30).
- CORBIN, A., COURTINE, J.-J. & VIGARELLO, G. (2016b). *Histoire des émotions, vol. 2 : Des Lumières à La Fin Du XIXe Siècle*. Seuil. (Cf. p. 30).
- COUTRIX, C. & MANDRAN, N. (2012). Identifying emotions expressed by mobile users through 2D surface and 3D motion gestures. *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, 311-320 (cf. p. 99, 140).
- CUI, L., LI, S., ZHANG, W., ZHANG, Z. & ZHU, T. (2015). Identifying Emotion from Natural Walking. *arXiv preprint arXiv1508.00413* (cf. p. 99, 140, 149, 159, 165).
- DAEL, N., MORTILLARO, M. & SCHERER, K. R. (2012). Emotion expression in body action and posture. *Emotion*, 12(5), 1085 (cf. p. 52).
- DAMASIO, A. R. (2003). *Spinoza avait raison : joie et tristesse, le cerveau des émotions*. Odile Jacob. (Cf. p. 52).
- DAMASIO, A. R. (2006). *L'erreur de Descartes : la raison des émotions*. Odile Jacob. (Cf. p. 52).
- DAN-GLAUSER, E. S. (2008). Le sentiment subjectif : intégration et représentation centrale consciente des composantes émotionnelles. *Traité de Psychologie de l'Emotion*, 223-258 (cf. p. 53, 54, 103, 104).
- DARWIN, C. (1877). *L'expression des émotions chez l'homme et les animaux*. (Cf. p. 31, 51).
- DE MEIJER, M. (1989). The contribution of general features of body movement to the attribution of emotions. *Journal of Nonverbal behavior*, 13(4), 247-268 (cf. p. 52).
- DEHGHANI, A., GLATARD, T. & SHIHAB, E. (2019). Subject cross validation in human activity recognition. *arXiv preprint arXiv :1904.02666* (cf. p. 154).
- DIXON, T. (2012a). "Emotion" : The history of a keyword in crisis. *Emotion Review*, 4(4), 338-344 (cf. p. 29).
- DIXON, T. (2012b). *Thomas Brown : selected philosophical writings* (T. 9). Andrews UK Limited. (Cf. p. 29).
- DZEDZICKIS, A., KAKLAUSKAS, A. & BUCINSKAS, V. (2020). Human emotion recognition : Review of sensors and methods. *Sensors*, 20(3), 592 (cf. p. 79, 80).
- EGGER, M., LEY, M. & HANKE, S. (2019). Emotion recognition from physiological signal analysis : a review. *Electronic Notes in Theoretical Computer Science*, 343, 35-55 (cf. p. 84).
- EKMAN, P. & FRIESEN, W. V. (1971). Constants across cultures in the face and emotion. *Journal of personality and social psychology*, 17(2), 124 (cf. p. 31, 51, 59, 78).
- ELLSWORTH, P. C. (2013). Appraisal theory : Old and new questions. *Emotion Review*, 5(2), 125-131 (cf. p. 61).
- ELMORSHIDY, A. (2013). Applying the technology acceptance and service quality models to live customer support chat for e-commerce websites. *Journal of Applied Business Research (JABR)*, 29(2), 589-596 (cf. p. 74).
- EMPATICA. (2021). *Empatica E4 Wristband*. Récupérée 29 avril 2021, à partir de <https://www.empatica.com/en-eu/research/e4/>. (Cf. p. 86)

- EPP, C., LIPPOLD, M. & MANDRYK, R. L. (2011). Identifying emotional states using keystroke dynamics. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 715-724 (cf. p. [92](#), [93](#), [102](#)).
- ESULI, A. & SEBASTIANI, F. (2007). SentiWordNet : a high-coverage lexical resource for opinion mining. *Evaluation*, 17(1), 26 (cf. p. [83](#)).
- FAN, R.-E., CHANG, K.-W., HSIEH, C.-J., WANG, X.-R. & LIN, C.-J. (2008). LIBLINEAR : A library for large linear classification. *the Journal of machine Learning research*, 9, 1871-1874 (cf. p. [160](#)).
- FERRER, R. A., GRENN, E. G. & TABER, J. M. (2015). Effectiveness of internet-based affect induction procedures : A systematic review and meta-analysis. *Emotion*, 15(6), 752 (cf. p. [106](#)).
- FONTAINE, J. R., SCHERER, K. R., ROESCH, E. B. & ELLSWORTH, P. C. (2007). The world of emotions is not two-dimensional. *Psychological science*, 18(12), 1050-1057 (cf. p. [65](#)).
- FOWLES, D. C., CHRISTIE, M. J., EDELBERG, R., GRINGS, W. W., LYKKEN, D. T. & VENABLES, P. H. (1981). Publication recommendations for electrodermal measurements. *Psychophysiology*, 18(3), 232-239 (cf. p. [132](#)).
- FRAZIER, T. W., STRAUSS, M. E. & STEINHAEUER, S. R. (2004). Respiratory sinus arrhythmia as an index of emotional response in young adults. *Psychophysiology*, 41(1), 75-83 (cf. p. [104](#)).
- FRIEDMAN, J. H. (2001). Greedy function approximation : a gradient boosting machine. *Annals of statistics*, 1189-1232 (cf. p. [160](#)).
- FRIJDA, N. H. *et al.* (1986). *The emotions*. Cambridge University Press. (Cf. p. [38](#), [41](#), [50](#)).
- FRIJDA, N. H. (2003). 10 Emotions and Hedonic Experience. *Well-being : Foundations of hedonic psychology*, 190 (cf. p. [50](#)).
- FRIJDA, N. H. & TCHERKASSOF, A. (1997). Facial expressions as modes of action readiness. (cf. p. [51](#)).
- GAUDUCHEAU, N. (2008). La communication des émotions dans les échanges médiatisés par ordinateur : bilan et perspectives. *Bulletin de psychologie*, (4), 389-404 (cf. p. [75](#)).
- GÉRON, A. (2019). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow : Concepts, tools, and techniques to build intelligent systems*. O'Reilly Media. (Cf. p. [153](#), [157](#)).
- GHOSH, S., GANGULY, N., MITRA, B. & DE, P. (2017a). Evaluating effectiveness of smartphone typing as an indicator of user emotion. *Affective Computing and Intelligent Interaction (ACII), 2017 Seventh International Conference on*, 146-151 (cf. p. [96](#), [150](#)).
- GHOSH, S., GANGULY, N., MITRA, B. & DE, P. (2017b). TapSense : combining self-report patterns and typing characteristics for smartphone based emotion detection. *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services*, 2 (cf. p. [96](#)).

- GHOSH, S., HIWARE, K., GANGULY, N., MITRA, B. & DE, P. (2019). Emotion detection from touch interactions during text entry on smartphones. *International Journal of Human-Computer Studies* (cf. p. 96).
- GIL, S. (2009). Comment étudier les émotions en laboratoire. *Revue électronique de psychologie sociale*, 4, 15-24 (cf. p. 103, 108, 109, 208).
- GLOROT, X. & BENGIO, Y. (2010). Understanding the difficulty of training deep feed-forward neural networks. *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, 249-256 (cf. p. 160).
- GRANDJEAN, D., SANDER, D. & SCHERER, K. R. (2008). Conscious emotional experience emerges as a function of multilevel, appraisal-driven response synchronization. *Consciousness and cognition*, 17(2), 484-495 (cf. p. 44).
- GRANDJEAN, D. M. & BAENZIGER, T. (2009). Expression vocale des émotions. *Traité de psychologie des émotions* (cf. p. 51).
- GRANDJEAN, D. M. & SCHERER, K. R. (2009). Théorie de l'évaluation cognitive et dynamique des processus émotionnels. *Traité de psychologie des émotions* (cf. p. 40-43, 46).
- GROSS, J. J., JOHN, O. & RICHARDS, J. (1995). *Berkeley expressivity questionnaire*. Edwin Mellen Press. (Cf. p. 133).
- GROSS, J. J. & LEVENSON, R. W. (1995). Emotion elicitation using films. *Cognition & emotion*, 9(1), 87-108 (cf. p. 105).
- GUO, H.-W., HUANG, Y.-S., LIN, C.-H., CHIEN, J.-C., HARAICAWA, K. & SHIEH, J.-S. (2016). Heart rate variability signal features for emotion recognition by using principal component analysis and support vectors machine. *2016 IEEE 16th International Conference on Bioinformatics and Bioengineering (BIBE)*, 274-277 (cf. p. 84, 160).
- GUYON, I., WESTON, J., BARNHILL, S. & VAPNIK, V. (2002). Gene selection for cancer classification using support vector machines. *Machine learning*, 46(1), 389-422 (cf. p. 158).
- HAAG, A., GORONZY, S., SCHAICH, P. & WILLIAMS, J. (2004). Emotion recognition using bio-sensors : First steps towards an automatic system. *Tutorial and research workshop on affective dialogue systems*, 36-48 (cf. p. 47).
- HARPER, R. & SOUTHERN, J. (2019). End-To-End Prediction of Emotion From Heartbeat Data Collected by a Consumer Fitness Tracker. *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*, 1-7 (cf. p. 86).
- HASHMI, M. A., RIAZ, Q., ZEESHAN, M., SHAHZAD, M. & FRAZ, M. M. (2020). Motion Reveal Emotions : Identifying Emotions from Human Walk Using Chest Mounted Smartphone. *IEEE Sensors Journal*, 20(22), 13511-13522 (cf. p. 99, 140).
- HASTIE, T., ROSSET, S., ZHU, J. & ZOU, H. (2009). Multi-class adaboost. *Statistics and its Interface*, 2(3), 349-360 (cf. p. 160).
- HIBBELN, M. T., JENKINS, J. L., SCHNEIDER, C., VALACICH, J. & WEINMANN, M. (2017). How is your user feeling? Inferring emotion through human-computer interaction devices. *Mis Quarterly*, 41(1), 1-21 (cf. p. 93).

- HOWELL, D., BESTGEN, Y., YZERBYT, V. & ROGIER, M. (2008). *Méthodes statistiques en sciences humaines*. (Cf. p. 172).
- JAMES, W. (1884). What is an Emotion? *Mind*, 9(34), 188-205. <http://www.jstor.org/stable/2246769> (cf. p. 32, 33, 53)
- JERRITTA, S., MURUGAPPAN, M., NAGARAJAN, R. & WAN, K. (2011). Physiological signals based human emotion recognition : a review. *2011 IEEE 7th International Colloquium on Signal Processing and its Applications*, 410-415 (cf. p. 84).
- JOYCE, R. & GUPTA, G. (1990). Identity authentication based on keystroke latencies. *Communications of the ACM*, 33(2), 168-176 (cf. p. 92).
- KAISER, S., WEHRLE, T. & SCHENKEL, K. (2009). Expression faciale des émotions. *Traité de psychologie des émotions*. Paris : Dunod, 109-157 (cf. p. 51).
- KELTNER, D. & HAIDT, J. (1999). Social functions of emotions at four levels of analysis. *Cognition & Emotion*, 13(5), 505-521 (cf. p. 23, 75).
- KHANNA, P. & SASIKUMAR, M. (2010). Recognising emotions from keyboard stroke pattern. *International journal of computer applications*, 11(9), 1-5 (cf. p. 92).
- KINECT. (2020). Récupérée 23 avril 2021, à partir de <https://developer.microsoft.com/fr-fr/windows/kinect/>. (Cf. p. 79, 80)
- KLEINGINNA, P. R. & KLEINGINNA, A. M. (1981). A categorized list of emotion definitions, with suggestions for a consensual definition. *Motivation and emotion*, 5(4), 345-379 (cf. p. 29).
- KO, B. C. (2018). A brief review of facial emotion recognition based on visual information. *sensors*, 18(2), 401 (cf. p. 78).
- KOŁAKOWSKA, A. (2013). A review of emotion recognition methods based on keystroke dynamics and mouse movements. *Human System Interaction (HSI), 2013 The 6th International Conference on*, 548-555 (cf. p. 91-93).
- KOŁAKOWSKA, A., SZWOCH, W. & SZWOCH, M. (2020). A review of emotion recognition methods based on data acquired via smartphone sensors. *Sensors*, 20(21), 6367 (cf. p. 94, 98, 100, 140, 144, 149, 159, 207).
- KURDI, B., LOZANO, S. & BANAJI, M. (2017). Introducing the Open Affective Standardized Image Set (OASIS) [cited By 16]. *Behavior Research Methods*, 49(2), 457-470. <https://doi.org/10.3758/s13428-016-0715-3> (cf. p. 104)
- LANG, P. J., BRADLEY, M. M., CUTHBERT, B. N. *et al.* (1997). International affective picture system (IAPS) : Technical manual and affective ratings. *NIMH Center for the Study of Emotion and Attention*, 1, 39-58 (cf. p. 104).
- LANGE, C. G. (1885). The mechanism of the emotions. *The classical psychologists*, 672-684 (cf. p. 33).
- LAZARUS, R. S. (1968). Emotions and adaptation : Conceptual and empirical relations. *Nebraska symposium on motivation* (cf. p. 39, 42, 45, 46).
- LAZARUS, R. S. & ALFERT, E. (1964). Short-circuiting of threat by experimentally altering cognitive appraisal. *The Journal of Abnormal and Social Psychology*, 69(2), 195 (cf. p. 48).
- LEDOUX, J. E. (2000). Emotion circuits in the brain. *Annual review of neuroscience*, 23(1), 155-184 (cf. p. 48).

- LEE, G., GOMMERS, R., WASELEWSKI, F., WOHLFAHRT, K. & O'LEARY, A. (2019). PyWavelets : A Python package for wavelet analysis. *Journal of Open Source Software*, 4(36), 1237 (cf. p. 149).
- LEVENTHAL, H. (1980). Toward a comprehensive theory of emotion. *Advances in experimental social psychology* (p. 139-207). Elsevier. (Cf. p. 36).
- LIU, B. (2017). Many facets of sentiment analysis. *A practical guide to sentiment analysis* (p. 11-39). Springer. (Cf. p. 82).
- LORIA, S. (2018). textblob Documentation. *Release 0.15*, 2 (cf. p. 83, 186).
- LOTSTRA, F. (2002). Le cerveau émotionnel ou la neuroanatomie des émotions. *Cahiers critiques de thérapie familiale et de pratiques de réseaux*, (2), 73-86 (cf. p. 48).
- LÖVHEIM, H. (2012). A new three-dimensional model for emotions and monoamine neurotransmitters. *Medical hypotheses*, 78(2), 341-348 (cf. p. 65).
- LU, Y., WEI, Y., LIU, L., ZHONG, J., SUN, L. & LIU, Y. (2017). Towards unsupervised physical activity recognition using smartphone accelerometers. *Multimedia Tools and Applications*, 76(8), 10701-10719 (cf. p. 146).
- LUGTIG, P. & TOEPOEL, V. (2016). The use of PCs, smartphones, and tablets in a probability-based panel survey : Effects on survey measurement error. *Social Science Computer Review*, 34(1), 78-94 (cf. p. 91).
- MAHBOUB, K. (2011). *Modélisation des processus émotionnel dans la prise de décision* (thèse de doct.). (Cf. p. 37, 62).
- MARAÑÓN, G. (1924). *Contribution a l'étude de l'action emotive de l'adrenaline*. *Revue Française d'Endocrinologie*. (Cf. p. 33).
- MCCARTHY, C., PRADHAN, N., REDPATH, C. & ADLER, A. (2016). Validation of the Empatica E4 wristband. *2016 IEEE EMBS International Student Conference (ISC)*, 1-4 (cf. p. 86).
- MCKEOWN, G., VALSTAR, M., COWIE, R., PANTIC, M. & SCHRODER, M. (2011). The semaine database : Annotated multimodal records of emotionally colored conversations between a person and a limited agent. *IEEE transactions on affective computing*, 3(1), 5-17 (cf. p. 88).
- MCKINNEY, W. (2010). Data Structures for Statistical Computing in Python. In S. van der WALT & J. MILLMAN (Éd.), *Proceedings of the 9th Python in Science Conference* (p. 56-61). <https://doi.org/10.25080/Majora-92bf1922-00a>. (Cf. p. 144)
- MEDHAT, W., HASSAN, A. & KORASHY, H. (2014). Sentiment analysis algorithms and applications : A survey. *Ain Shams engineering journal*, 5(4), 1093-1113 (cf. p. 83).
- MEHRABIAN, A. (1996). Pleasure-arousal-dominance : A general framework for describing and measuring individual differences in temperament. *Current Psychology*, 14(4), 261-292 (cf. p. 63, 64).
- MEHTA, D., SIDDIQUI, M. F. H. & JAVAID, A. Y. (2018). Facial emotion recognition : A survey and real-world user experiences in mixed reality. *Sensors*, 18(2), 416 (cf. p. 77).
- MIKELS, J. A., FREDRICKSON, B. L., LARKIN, G. R., LINDBERG, C. M., MAGLIO, S. J. & REUTER-LORENZ, P. A. (2005). Emotional category data on images from the In-

- ternational Affective Picture System. *Behavior research methods*, 37(4), 626-630 (cf. p. 104).
- MITCHELL, E., MONAGHAN, D. & O'CONNOR, N. E. (2013). Classification of sporting activities using smartphone accelerometers. *Sensors*, 13(4), 5317-5337 (cf. p. 148).
- MOHAMMAD, S. M. (2017). Challenges in sentiment analysis. *A practical guide to sentiment analysis* (p. 61-83). Springer. (Cf. p. 83).
- MONGODB. (2021). Récupérée 23 juin 2021, à partir de <https://www.mongodb.com/fr>. (Cf. p. 117)
- MOORS, A. (2009). Theories of emotion causation : A review. *Cognition and emotion*, 23(4), 625-662 (cf. p. 34, 35, 38).
- MOORS, A., ELLSWORTH, P. C., SCHERER, K. R. & FRIJDA, N. H. (2013). Appraisal theories of emotion : State of the art and future development. *Emotion Review*, 5(2), 119-124 (cf. p. 42).
- MORENCY, L.-P., MIHALCEA, R. & DOSHI, P. (2011). Towards multimodal sentiment analysis : Harvesting opinions from the web. *Proceedings of the 13th international conference on multimodal interfaces*, 169-176 (cf. p. 88, 103).
- MORRISON, D., WANG, R. & DE SILVA, L. (2007). Ensemble methods for spoken emotion recognition in call-centres [cited By 181]. *Speech Communication*, 49(2), 98-112. <https://doi.org/10.1016/j.specom.2006.11.004> (cf. p. 81)
- MOTTELSON, A. & HORNBAEK, K. (2016). An affect detection technique using mobile commodity sensors in the wild. *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 781-792 (cf. p. 99, 140, 149, 159).
- MURPHY, F. C., NIMMO-SMITH, I. & LAWRENCE, A. D. (2003). Functional neuroanatomy of emotions : a meta-analysis. *Cognitive, affective, & behavioral neuroscience*, 3(3), 207-233 (cf. p. 48).
- NUGIER, A. (2009). Histoire et grands courants de recherche sur les émotions. *Revue électronique de psychologie sociale*, 4(4), 8-14 (cf. p. 31, 32, 34, 36).
- NUMMENMAA, L. & NIEMI, P. (2004). Inducing affective states with success-failure manipulations : a meta-analysis. *Emotion*, 4(2), 207 (cf. p. 105).
- OLSEN, A. F. & TORRESEN, J. (2016). Smartphone accelerometer data used for detecting human emotions. *Systems and Informatics (ICSAI), 2016 3rd International Conference on*, 410-415 (cf. p. 99, 131, 140, 149, 159).
- ORTONY, A., CLORE, G. L. & COLLINS, A. (1990). *The cognitive structure of emotions*. Cambridge university press. (Cf. p. 61, 62, 66, 136, 190, 195).
- OUELLET, C., PROVENCHER, M. D. & LANGLOIS, F. (2012). Validation d'une adaptation francophone du Berkeley Expressivity Questionnaire (cf. p. 133).
- PECCHINENDA, A. (2001). The psychophysiology of appraisals. *Appraisal processes in emotion : Theory, methods, research*, 301-315 (cf. p. 47, 48).
- PEDREGOSA, F., VAROQUAUX, G., GRAMFORT, A., MICHEL, V., THIRION, B., GRISEL, O., BLONDEL, M., PRETTENHOFER, P., WEISS, R., DUBOURG, V. et al. (2011). Scikit-learn : Machine learning in Python. *the Journal of machine Learning research*, 12, 2825-2830 (cf. p. 156).

- PESTANA, M. H., WANG, W.-C. & MOUTINHO, L. A. (2018). Global affective computing research in the period 1997-2017 : a bibliometric analysis. *International Journal of Multivariate Data Analysis*, 1(4), 348-370 (cf. p. 73).
- PIANA, S., STAGLIANO, A., ODONE, F., VERRI, A. & CAMURRI, A. (2014). Real-time automatic emotion recognition from body gestures. *arXiv preprint arXiv :1402.5047* (cf. p. 80).
- PICARD, R. W. (2010). Emotion research by the people, for the people. *Emotion Review*, 2(3), 250-254 (cf. p. 72).
- PICARD, R. W. (2015). The promise of affective computing. *The Oxford handbook of affective computing*, 11-20 (cf. p. 72).
- PICARD, R. W. *et al.* (1995). Affective computing (cf. p. 71, 72).
- PLUTCHIK, R. (1984). Emotions : A general psychoevolutionary theory. *Approaches to emotion, 1984*, 197-219 (cf. p. 32, 50, 60).
- PLUTCHIK, R. (2001). The nature of emotions : Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *American scientist*, 89(4), 344-350 (cf. p. 32).
- POLITOU, E., ALEPIS, E. & PATSAKIS, C. (2017). A survey on mobile affective computing. *Computer Science Review*, 25, 79-100 (cf. p. 94, 100, 101).
- POLLICK, F. E., PATERSON, H. M., BRUDERLIN, A. & SANFORD, A. J. (2001). Perceiving affect from arm movement. *Cognition*, 82(2), B51-B61 (cf. p. 52).
- PORIA, S., CAMBRIA, E., BAJPAI, R. & HUSSAIN, A. (2017). A review of affective computing : From unimodal analysis to multimodal fusion. *Information Fusion*, 37, 98-125 (cf. p. 72, 76, 77, 81, 87, 88).
- RAVEN, J. *et al.* (2003). Raven progressive matrices. *Handbook of nonverbal assessment* (p. 223-237). Springer. (Cf. p. 106, 136).
- REACT-NATIVE. (2021). Récupérée 1 juin 2021, à partir de <https://reactnative.dev/>. (Cf. p. 116)
- REACT-NATIVE-SENSORS. (2021). Récupérée 2 juin 2021, à partir de <https://react-native-sensors.github.io/>. (Cf. p. 117)
- RISH, I. *et al.* (2001). An empirical study of the naive Bayes classifier. *IJCAI 2001 workshop on empirical methods in artificial intelligence*, 3(22), 41-46 (cf. p. 160).
- ROSEMAN, I. J. & SMITH, C. A. (2001). Appraisal theory. *Appraisal processes in emotion : Theory, methods, research*, 3-19 (cf. p. 42, 45).
- RUENSUK, M., CHEON, E., HONG, H. & OAKLEY, I. (2020). How Do You Feel Online : Exploiting Smartphone Sensors to Detect Transitory Emotions during Social Media Use. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 4(4), 1-32 (cf. p. 131, 149, 158, 159).
- RUENSUK, M., OH, H., CHEON, E., OAKLEY, I. & HONG, H. (2019). Detecting negative emotions during social media use on smartphones. *Proceedings of Asian CHI Symposium 2019 : Emerging HCI Research Collection*, 73-79 (cf. p. 100).
- RUSSELL, J. A. (1980). A circumplex model of affect. *Journal of personality and social psychology*, 39(6), 1161 (cf. p. 63, 64, 123).

- RUSSELL, J. A. (1991). Culture and the categorization of emotions. *Psychological bulletin*, 110(3), 426 (cf. p. 36, 53).
- RUSSELL, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological review*, 110(1), 145 (cf. p. 64).
- SACHARIN, V., SCHLEGEL, K. & SCHERER, K. R. (2012). Geneva emotion wheel rating study (cf. p. 108, 115).
- SAMSON, A. C., KREIBIG, S. D., SODERSTROM, B., WADE, A. A. & GROSS, J. J. (2016). Eliciting positive, negative and mixed emotional states : A film library for affective scientists. *Cognition and Emotion*, 30(5), 827-856 (cf. p. 105, 128, 129, 189, 239).
- SAMUEL, A. L. (1959). Some studies in machine learning using the game of checkers. *IBM Journal of research and development*, 3(3), 210-229 (cf. p. 153).
- SANDER, D. & SCHERER, K. R. (2009). La psychologie des émotions : survol des théories et débats essentiels. *Traité de psychologie des émotions* (cf. p. 39-41, 63).
- SANDER, D., SCHERER, K. R. et al. (2014). *Traité de psychologie des émotions*. Dunod. (Cf. p. 32-35).
- SAXENA, A., KHANNA, A. & GUPTA, D. (2020). Emotion recognition and detection methods : A comprehensive survey. *Journal of Artificial Intelligence and Systems*, 2(1), 53-79 (cf. p. 77).
- SCHACHTER, S. & SINGER, J. (1962). Cognitive, social, and physiological determinants of emotional state. *Psychological review*, 69(5), 379 (cf. p. 34-37).
- SCHAEFER, A., NILS, F., SANCHEZ, X. & PHILIPPOT, P. (2010). Assessing the effectiveness of a large database of emotion-eliciting films : A new tool for emotion researchers. *Cognition and Emotion*, 24(7), 1153-1172 (cf. p. 104, 105).
- SCHERER, K. R. et al. (2000). Psychological models of emotion. *The neuropsychology of emotion*, 137(3), 137-162 (cf. p. 55, 56).
- SCHERER, K. R. (2004). Feelings integrate the central representation of appraisal-driven response organization in emotion. *Feelings and emotions : The Amsterdam symposium*, 136-157 (cf. p. 54).
- SCHERER, K. R. (2005). What are emotions? And how can they be measured? *Social science information*, 44(4), 695-729 (cf. p. 40, 41, 43, 53, 55-57, 65).
- SCHERER, K. R., SHUMAN, V., FONTAINE, J. & SORIANO SALINAS, C. (2013). The GRID meets the Wheel : Assessing emotional feeling via self-report. *Components of emotional meaning : A sourcebook* (cf. p. 108, 115, 120, 196, 215).
- SCHMIDT, P., REISS, A., DUERICHEN, R., MARBERGER, C. & VAN LAERHOVEN, K. (2018). Introducing wesad, a multimodal dataset for wearable stress and affect detection. *Proceedings of the 20th ACM International Conference on Multimodal Interaction*, 400-408 (cf. p. 86, 146, 151).
- SCHUCH, S., ZWEERINGS, J., HIRSCH, P. & KOCH, I. (2017). Conflict adaptation in positive and negative mood : Applying a success-failure manipulation. *Acta psychologica*, 176, 11-22 (cf. p. 106, 136).

- SHARMA, K., CASTELLINI, C., BROEK, E. L., ALBU-SCHAEFFER, A. & SCHWENKER, F. (2018). A dataset of continuous affect annotations and physiological signals for emotion analysis. *arXiv preprint arXiv :1812.02782* (cf. p. 87).
- SHU, L., XIE, J., YANG, M., LI, Z., LI, Z., LIAO, D., XU, X. & YANG, X. (2018). A review of emotion recognition using physiological signals. *Sensors*, 18(7), 2074 (cf. p. 84, 87).
- SIEGERT, I., BÖCK, R., VLASENKO, B., PHILIPPOU-HÜBNER, D. & WENDEMUTH, A. (2011). Appropriate emotional labelling of non-acted speech using basic emotions, geneva emotion wheel and self assessment manikins. *2011 IEEE International Conference on Multimedia and Expo*, 1-6 (cf. p. 133).
- SPARKFUN. (2021). Récupérée 4 mai 2021, à partir de <https://www.sparkfun.com/products/12786>. (Cf. p. 95)
- STATCOUNTER. (2021). Récupérée 4 mai 2021, à partir de <https://gs.statcounter.com/platform-market-share/desktop-mobile-tablet/worldwide>. (Cf. p. 24)
- STEUNEBRINK, B. R., DASTANI, M. & MEYER, J.-J. C. (2009). The OCC model revisited. *Proc. of the 4th Workshop on Emotion and Computing* (cf. p. 62).
- STRAPPARAVA, C., VALITUTTI, A. *et al.* (2004). Wordnet affect : an affective extension of wordnet. *Lrec*, 4(1083-1086), 40 (cf. p. 83).
- SWAIN, M., ROUTRAY, A. & KABISATPATHY, P. (2018). Databases, features and classifiers for speech emotion recognition : a review. *International Journal of Speech Technology*, 21(1), 93-120 (cf. p. 80).
- TAO, J. & TAN, T. (2005). Affective computing : A review. *International Conference on Affective computing and intelligent interaction*, 981-995 (cf. p. 73).
- TCHERKASSOF, A. & FRIJDA, N. H. (2014). Les émotions : une conception relationnelle. *L'Année psychologique*, 114(3), 501-535 (cf. p. 52).
- TOBII. (2021). *Tobii Pro Glasses 2*. Récupérée 29 avril 2021, à partir de <https://www.tobii.com/fr/produits/tobii-pro-glasses-2/>. (Cf. p. 85)
- TOMKINS, S. S. (1984). Affect theory. *Approaches to emotion*, 163(163-195) (cf. p. 31, 65).
- VAN DER WALT, S., COLBERT, S. C. & VAROQUAUX, G. (2011). The NumPy Array : A Structure for Efficient Numerical Computation. *Computing in Science Engineering*, 13(2), 22-30. <https://doi.org/10.1109/MCSE.2011.37> (cf. p. 149)
- van GENT, P., FARAH, H., NES, N. & van AREM, B. (2018). Heart rate analysis for human factors : Development and validation of an open source toolkit for noisy naturalistic heart rate data. *Proceedings of the 6th HUMANIST Conference*, 173-178 (cf. p. 146).
- VELKOVSKA, J. & BEAUDOUIN, V. (2014). Parler aux machines, coproduire un service Intelligence artificielle et travail du client dans les services vocaux automatisés. (Cf. p. 74).
- VIOT, C. & BRESSOLLES, G. (2012). LES AGENTS VIRTUELS INTELLIGENTS Quels atouts pour la relation client? *Décisions marketing*, (65), 45-56 (cf. p. 74).
- VIRTANEN, P., GOMMERS, R., OLIPHANT, T. E., HABERLAND, M., REDDY, T., COURNAPEAU, D., BUROVSKI, E., PETERSON, P., WECKESSER, W., BRIGHT, J., VAN DER WALT, S. J.,

- BRETT, M., WILSON, J., JARROD MILLMAN, K., MAYOROV, N., NELSON, A. R. J., JONES, E., KERN, R., LARSON, E., ... CONTRIBUTORS, S. 1. 0. (2020). SciPy 1.0 : Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*. <https://doi.org/https://doi.org/10.1038/s41592-019-0686-2> (cf. p. 145, 149)
- WALTHER, J. B. (1996). Computer-mediated communication : Impersonal, interpersonal, and hyperpersonal interaction. *Communication research*, 23(1), 3-43 (cf. p. 75).
- WAMPFLER, R., KLINGLER, S., SOLENTHALER, B., SCHINAZI, V. R. & GROSS, M. (2020). Affective State Prediction Based on Semi-Supervised Learning from Smartphone Touch Data. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1-13 (cf. p. 166).
- WATSON, D., CLARK, L. A. & TELLEGEN, A. (1988). Development and validation of brief measures of positive and negative affect : the PANAS scales. *Journal of personality and social psychology*, 54(6), 1063 (cf. p. 107, 117).
- WEINER, B. (2001). Intrapersonal and interpersonal theories of motivation from an attribution perspective. *Student motivation* (p. 17-30). Springer. (Cf. p. 49).
- WEISER, M. (1994). Ubiquitous computing. *ACM Conference on Computer Science*, 418(10.1145), 197530-197680 (cf. p. 94).
- WEN, W., LIU, G., CHENG, N., WEI, J., SHANGGUAN, P. & HUANG, W. (2014). Emotion recognition based on multi-variant correlation of physiological signals. *IEEE Transactions on Affective Computing*, 5(2), 126-140 (cf. p. 84).
- WESTERMANN, R., SPIES, K., STAHL, G. & HESSE, F. W. (1996). Relative effectiveness and validity of mood induction procedures : A meta-analysis. *European Journal of social psychology*, 26(4), 557-580 (cf. p. 106, 189).
- WIERZBICKA, A. (1986). Human emotions : universal or culture-specific? *American anthropologist*, 88(3), 584-594 (cf. p. 36, 53).
- WÖLLMER, M., WENINGER, F., KNAUP, T., SCHULLER, B., SUN, C., SAGAE, K. & MORENCY, L.-P. (2013). Youtube movie reviews : Sentiment analysis in an audio-visual context. *IEEE Intelligent Systems*, 28(3), 46-53 (cf. p. 88, 103).
- WU, S., XU, X., SHU, L. & HU, B. (2017). Estimation of valence of emotion using two frontal EEG channels. *2017 IEEE international conference on bioinformatics and biomedicine (BIBM)*, 1127-1130 (cf. p. 84).
- WUNDT, W. M. (1874). *Grundzüge der physiologischen Psychologie* (T. 1). W. Engelman. (Cf. p. 63).
- ZHAO, B., WANG, Z., YU, Z. & GUO, B. (2018). EmotionSense : Emotion recognition based on wearable wristband. *2018 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*, 346-355 (cf. p. 86, 146, 151).
- ZHENG, W., LU, C., ZONG, Y., ZHANG, X., CUI, Z. *et al.* (2019). MPED : A Multi-Modal Physiological Emotion Database for Discrete Emotion Recognition. *IEEE Access* (cf. p. 86).

Productions scientifiques

Publications et communications

- **ERGO'IA, Octobre 2021** - Communication orale - Nicolas Simonazzi, Jean-Marc Salotti, Marie Morelle, Caroline Dubois, Philippe Le Goff. The Geneva Emotion Wheel Mobile Interface an Instrument to Report Emotions on Android Devices. (Statut : accepté)
- **International Conference on Human Interaction and Emerging Technologies (IHiet 2020)** - Communication orale, chapitre d'ouvrage - Nicolas Simonazzi, Jean-Marc Salotti, Caroline Dubois, Dominique Seminel. Emotion Detection Based on Smartphone Using User Memory Tasks and Videos. Tareq Ahram ; Redha Tair ; Karine Langlois ; Arnaud Choplin. *Advances in Intelligent Systems and Computing 1253, Human Interaction, Emerging Technologies and Future Applications III, proceedings of the 3rd International Conference on Human Interaction and Emerging Technologies : Future Applications (IHiet 2020)*, August 27–29, 2020, Paris, France, 1253, Springer, pp.244-249, 2021, AISC. (Statut : publié)
- **Société d'Ergonomie de Langue Française (SELF 2019)** - Communication Orale, publication dans les actes - Nicolas Simonazzi, Jean-Marc Salotti, Caroline Dubois, Dominique Séminel. Détection des émotions sur un smartphone à partir des interactions des utilisateurs. Proposition de protocole expérimental. *54ème Congrès International de la Société d'Ergonomie de Langue Française, Sep 2019, Tours, France*. (Statut : publié)
- **Hermès, La Revue** - Co-auteur d'article de revue - Caroline Dubois, Dominique Séminel, Jean-Marc Salotti, Nicolas Simonazzi. Le chatbot : un outil de la relation aux clients. *Hermès, La Revue- Cognition, communication, politique, CNRS-Editions, 2019, Les Incommunications, 2 (84)* (Statut : publié)

Autres contributions

- **Salon de la Recherche Orange, 7 et 9 Avril 2021** - Communication Orale - Laetitia Gros, Nicolas Simonazzi, Nolwenn Le Godinec. *Détecter et caractériser les états cognitifs et affectifs*.
- **Contribution au blog de la recherche d'Orange Labs Service, Décembre 2020** - Article de vulgarisation - Nicolas Simonazzi, Caroline Dubois, Philippe Le Goff. *De meilleures interactions avec la reconnaissance automatique des émotions*. <https://hellofuture.orange.com/fr/de-meilleures-interactions-avec-la-reconnaissance-automatique-des-emotions/>.
- **Journée de l'école doctorale en Science et Physique de l'ingénieur, 12 Mars 2020, Talence** - Présentation de poster - Induction et détection d'émotions sur application smartphone à partir de tâches cognitives et de vidéos.

Encadrements et enseignements

- Encadrement de projets de fin d'étude, École Nationale Supérieure de Cognitique :
 - Développement d'une application mobile d'expérimentation avec un agent conversationnel (Septembre 2018 - Janvier 2019).
 - Évolution d'un protocole expérimental pour la détection d'émotions avec un agent conversationnel (Septembre 2019 - Janvier 2020).
 - Mise en place d'une expérimentation d'induction d'états affectifs sur smartphone (Septembre 2020 - Janvier 2021).
- Encadrement stages :
 - Stage de deuxième année d'ingénieur cognicien : Développement d'une application mobile d'expérimentation (Avril - Juillet 2019).
 - Stage de fin d'étude d'ingénieur cognicien : Développement d'une application mobile de recueil d'état d'émotionnel (Février - Juillet 2021).



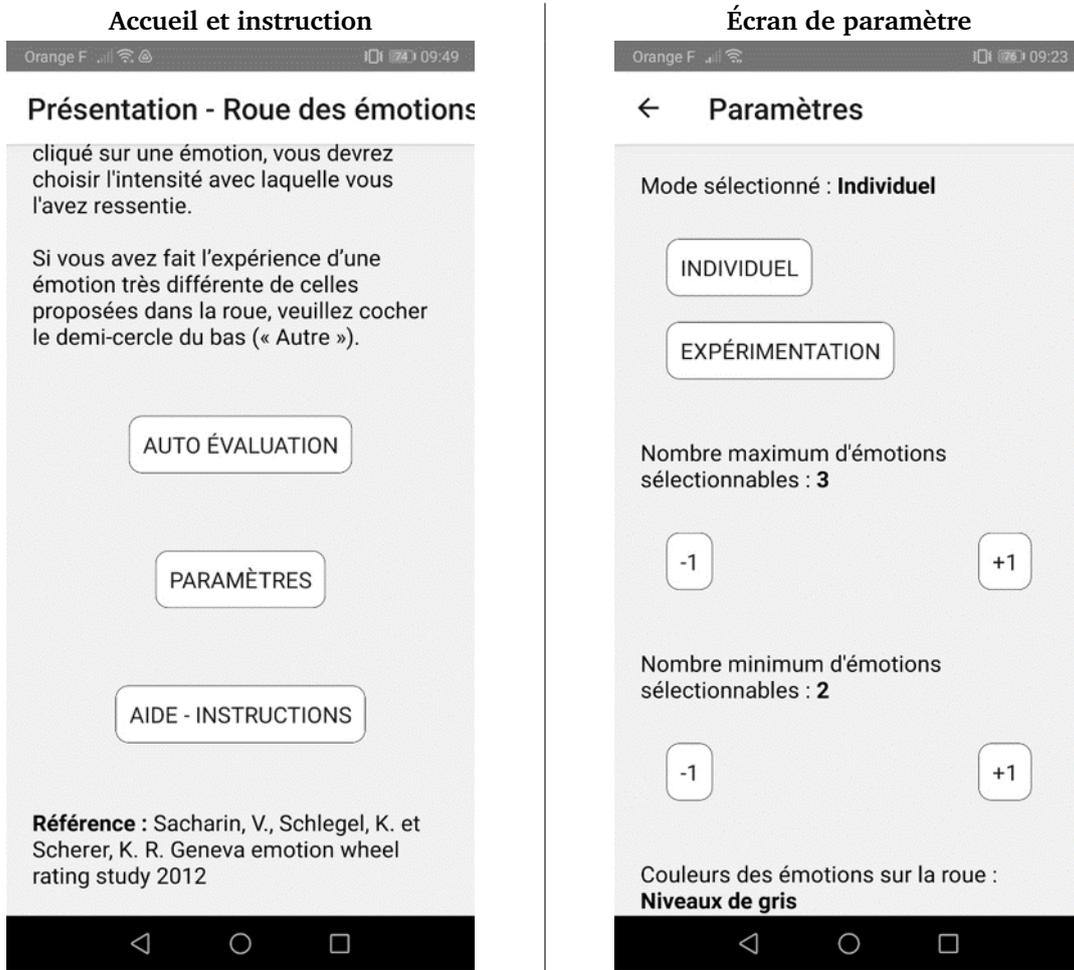
Annexes

Sommaire

A.	Captures d'écran de l'application GEWMI	234
B.	Formulaire de consentement	234
C.	Questionnaires	237
C.I	Caractéristiques personnelles	237
C.II	Habitudes et sentiment de compétence pour l'utilisation de smartphone	237
C.III	Évaluation du système d'auto-évaluation des émotions	238
C.IV	Évaluation de l'impulsivité émotionnelle	238
D.	Détails des vidéos de l'expérimentation principale	239
E.	Réponses aux questionnaires de l'expérimentation principale	240
E.I	Caractéristiques personnelles	240
E.II	Habitudes et sentiment de compétence pour l'utilisation de smartphone	241
E.III	Évaluation du système d'auto-évaluation des émotions	242
E.IV	Évaluation de l'impulsivité émotionnelle	244
F.	Représentation des signaux bruts pour une tâche de l'expérimentation principale	248
F.I	Capteurs du smartphone	248
F.II	Capteurs de la montre Empatica E4	251
G.	Détails des calculs sur les séries temporelles	255

A. Captures d'écran de l'application GEWMI

Représentations des écrans de l'application GEWMI



B. Formulaire de consentement

FORMULAIRE DE CONSENTEMENT LIBRE ET ECLAIRE POUR EXPERIENCE

Expérience de psychologie cognitive

ORANGE SA, Multipôle des Graves 23 Rue Thomas Edison, 33600 Pessac

Responsables du projet : Nicolas Simonazzi (Doctorant TGI/OLS/SRC, IMS Cognitive nicolas1.simonazzi@orange.com), Dominique Seminel (Responsable de programme TGI/OLS/SRC), Caroline Dubois (Chef de Projet TGI/OLS/BIZZ), Jean-Marc Salotti (Directeur de thèse IMS Cognitive).

Je certifie avoir donné mon accord pour participer à une étude comportementale de psychologie cognitive. J'accepte volontairement de participer à cette étude et je comprends que ma participation n'est pas obligatoire et que je peux stopper ma participation à tout moment sans avoir à me justifier ni encourir aucune responsabilité. Mon consentement ne décharge pas les organisateurs de la recherche de leurs responsabilités et je conserve tous mes droits garantis par la loi. De plus, conformément à la loi «informatique et libertés» du 6 janvier 1978, je bénéficie d'un droit d'accès et de rectification aux informations me concernant, que je peux exercer en m'adressant à Nicolas Simonazzi (nicolas1.simonazzi@orange.com).

Au cours de cette expérience, j'accepte que soient recueillies des données des capteurs du smartphone sur mes réponses. Je comprends que les informations recueillies seront traitées de façon anonyme et resteront strictement confidentielles. Elles seront à usage exclusif des investigateurs concernés. Je prends connaissance que le contenu des vidéos présentes dans cette expérience peut heurter la sensibilité d'un public non averti.

J'ai été informé que les résultats obtenus à l'issu du traitement de ce questionnaire pourront faire l'objet de publications scientifiques, mais aucun renseignement pouvant révéler mon identité ne sera dévoilé. J'accepte que les données enregistrées à l'occasion de cette étude puissent être conservées dans une base de données et faire l'objet d'un traitement informatisé non nominatif par ORANGE SA et le laboratoire IMS Bordeaux. J'ai bien noté que le droit d'accès prévu par la loi « informatique et libertés » s'exerce à tout moment auprès d'ORANGE SA et du laboratoire IMS Bordeaux.

Annexe Covid 19 :

J'ai pris connaissance :

- qu'en dehors du matériel spécifique à l'expérimentation, aucun matériel ne sera prêté (stylos...);
- des consignes de sécurité relatives à l'expérimentation à laquelle je vais prendre part ;
- des consignes pour le lavage des mains et le port du masque ;
- des gestes barrières à respecter.

Je déclare :

- ne pas relever de la liste des personnes dites « vulnérables » selon le décret n° 2020-521 du 5 mai 2020 et ne pas vivre avec une personne relevant de cette liste ;
- ne pas présenter de symptômes compatibles Covid-19 aujourd'hui ;
- ne pas avoir présenté de symptômes compatibles Covid-19 ou été diagnostiqué(e) positif(ve) dans les 21 derniers jours ;

Je soussigné(e)

- reconnais avoir lu et compris le présent formulaire de consentement libre et éclairé ;
- déclare que les informations ci-dessus sont exactes ;
- m'engage à respecter toutes les consignes.

Fait en deux exemplaires dont un sera remis au participant et un sera conservé par le responsable du projet.

Pour le participant :

Date et signature

Pour le responsable du projet :

Date et signature :

C. Questionnaires

C.I Caractéristiques personnelles

Questionnaire sur les caractéristiques sociodémographiques des participants des expérimentations.

Questions	Réponses proposées
Quel est votre genre ?	Homme / Femme / Autre
Dans quelle catégorie d'âge vous situez-vous ?	18-24 / 25-34 / 35-49 / 50-64 / 65+
Quelle est votre main dominante ?	Gauche / Droite / Pas de main dominante

C.II Habitudes et sentiment de compétence pour l'utilisation de smartphone

Questionnaire sur les habitudes et sentiment de compétence sur l'utilisation de smartphone des participants des expérimentations.

Questions	Réponses proposées
Combien de temps utilisez-vous votre smartphone par jour ?	Moins de 30 min / Entre 30 min et 1 heure / Entre 1 et 2 heures / Plus de 2 heures
Combien de temps utilisez-vous des applications de chat et de texting par jour ? (sms, whatsapp, Facebook messenger ...)	Moins de 30 min / Entre 30 min et 1 heure / Entre 1 et 2 heures / Plus de 2 heures
Quelle est, d'après vous, votre maîtrise de l'usage des fonctionnalités d'un smartphone ?	Très faible / Faible / Moyen / Elevé / Très élevé
Quelle est, d'après vous, votre maîtrise de l'usage des applications de chat/texting ? (sms, whatsapp, Facebook messenger ...)	Très faible / Faible / Moyen / Elevé / Très élevé

C.III Évaluation du système d'auto-évaluation des émotions

Questionnaire sur l'évaluation du système l'autoévaluation des émotions.

Questions	Réponses proposées
Il a été facile pour moi d'identifier les émotions ressenties.	Échelle de Likert de 7 points (1 à 7)
Les catégories d'émotions proposées étaient suffisantes pour décrire les émotions ressenties.	Échelle de Likert de 7 points (1 à 7)
Il était compréhensible pour moi de savoir quelle catégorie représente quelle émotion.	Échelle de Likert de 7 points (1 à 7)
J'avais une idée claire des émotions observées avant de passer le test.	Échelle de Likert de 7 points (1 à 7)

C.IV Évaluation de l'impulsivité émotionnelle

Questionnaire d'impulsivité émotionnelle.

Questions	Réponses proposées
Je pleure parfois aux films tristes.	Échelle de Likert de 7 points (1 à 7)
Mon corps réagit très fortement à des situations émotionnelles.	Échelle de Likert de 7 points (1 à 7)
Je ressens de fortes émotions.	Échelle de Likert de 7 points (1 à 7)
Je suis parfois incapable de cacher mes sentiments, même si j'aimerais pouvoir le faire.	Échelle de Likert de 7 points (1 à 7)
J'ai vécu des périodes dans ma vie où je n'arrivais pas à arrêter de pleurer même si j'essayais.	Échelle de Likert de 7 points (1 à 7)
Je ressens des émotions avec beaucoup d'intensité.	Échelle de Likert de 7 points (1 à 7)

D. Détails des vidéos de l'expérimentation principale

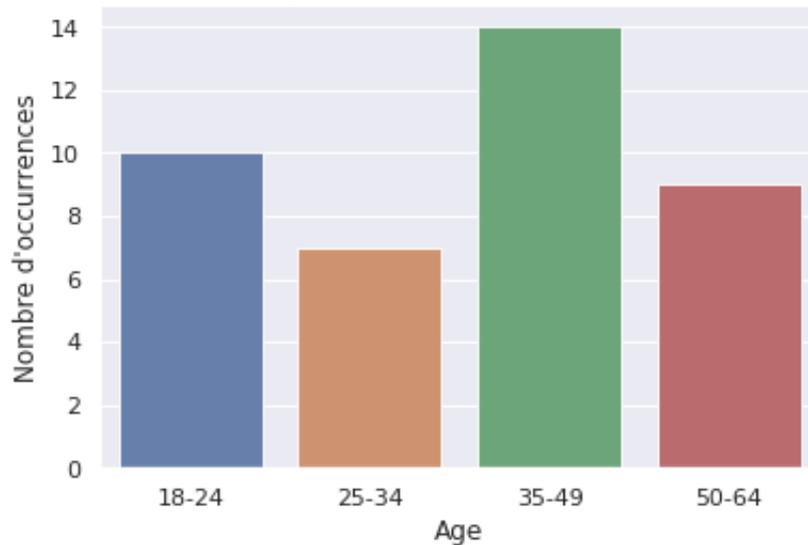
Description des extraits Samson et al. Samson *et al.*, 2016

Courte description de la vidéo	Condition	Niveau de valence (1 à 7)	Durée (secondes)
Un garçon est percuté par une voiture	Négative	1.60	20
Un crocodile mords le bras d'un homme	Négative	1.77	30
Un homme se fait frapper par un cheval	Négative	2.35	28
Un skater se casse le bras	Négative	1.68	27
Un bébé mets les bras en l'air devant une foule	Positive	5.32	20
Un bébé fait des grimaces	Positive	5.01	21
Un chat sort la langue	Positive	4.85	30
Un panda dans un zoo éternue	Positive	4.68	30
Un ouvrier trie des cartons	Neutre	3.04	26
Des gens sont assis en terrasse	Neutre	3.29	22
Une personne prépare un repas	Neutre	3.37	26
Des gens visitent un musée	Neutre	3.28	24
Un train entre dans une station	Entraînement	3.06	23

E. Réponses aux questionnaires de l'expérimentation principale

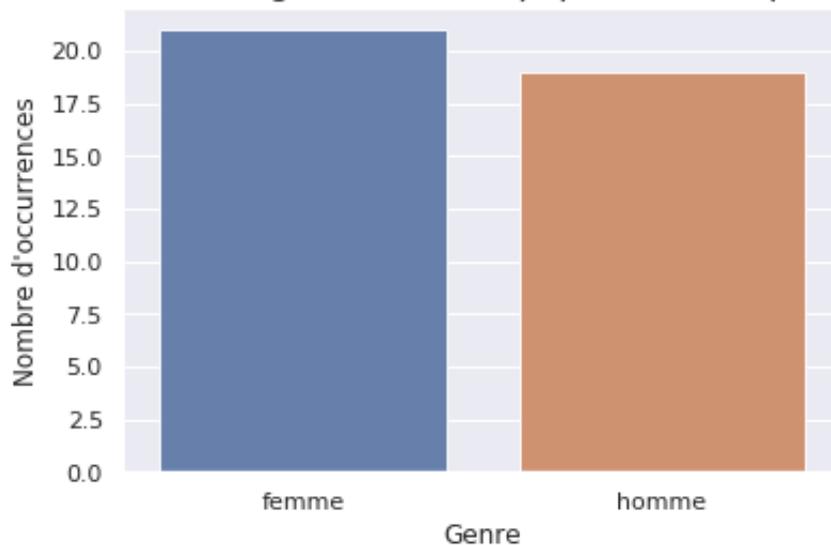
E.I Caractéristiques personnelles

Distribution des âges dans la population des participants

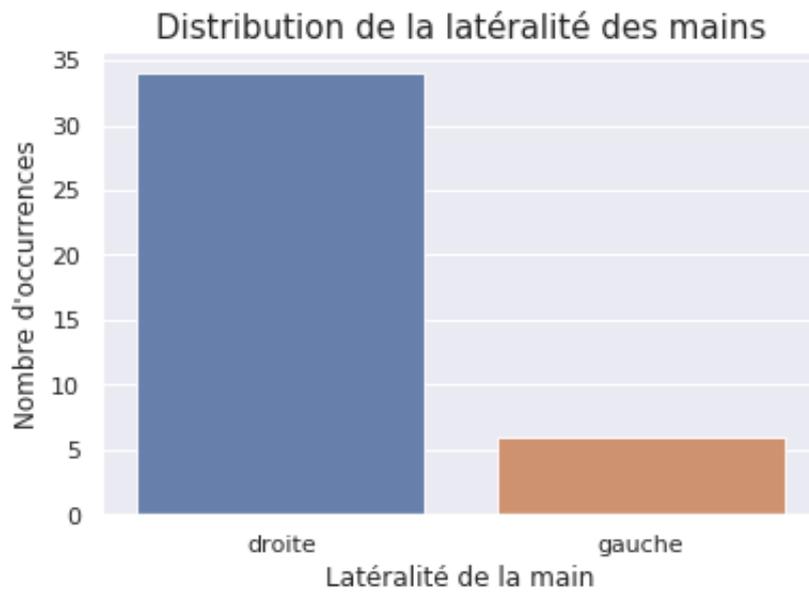


Distribution des âges

Distribution des genres dans la population des participants



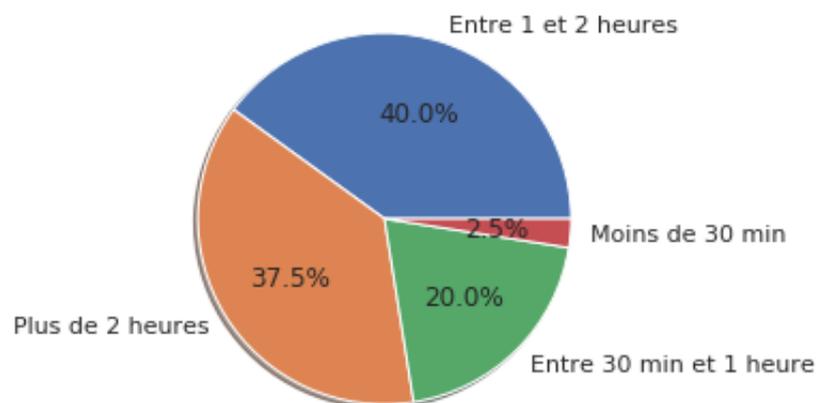
Distribution du genre



Distribution de la latéralité de la main

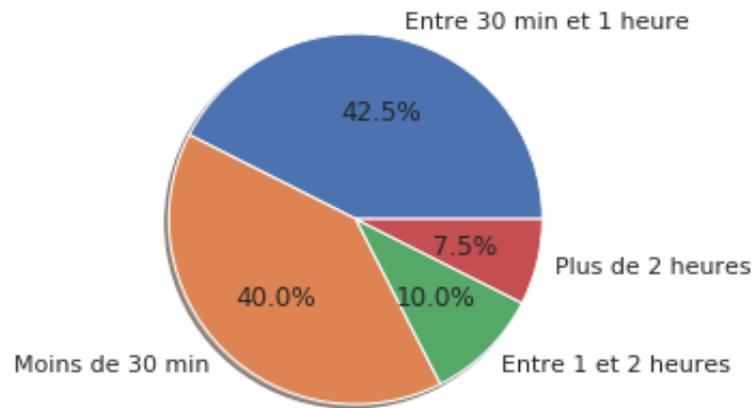
E.II Habitudes et sentiment de compétence pour l'utilisation de smartphone

Distribution des fréquences de l'usage de smartphone au quotidien (en heure)



Répartition du temps moyen estimé de l'usage quotidien du smartphone

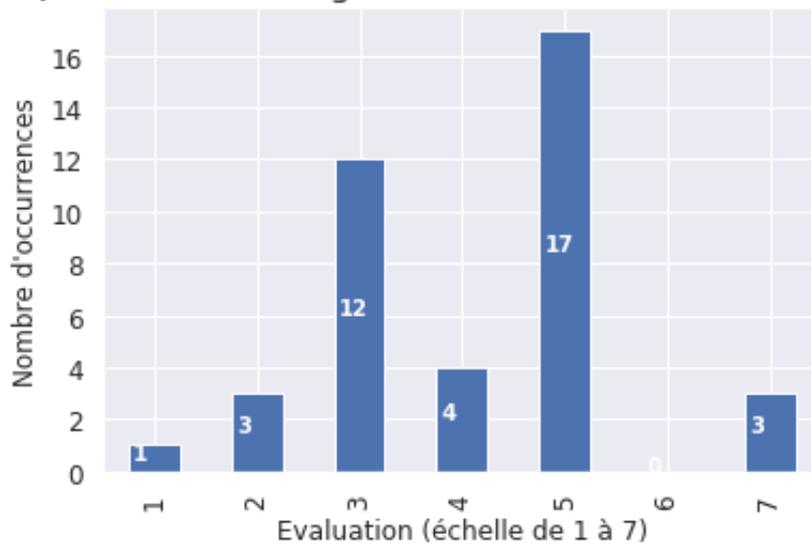
Distribution des fréquences de l'usage d'application de chat



Répartition du temps moyen estimé de l'usage d'application de messagerie instantannée sur smartphone

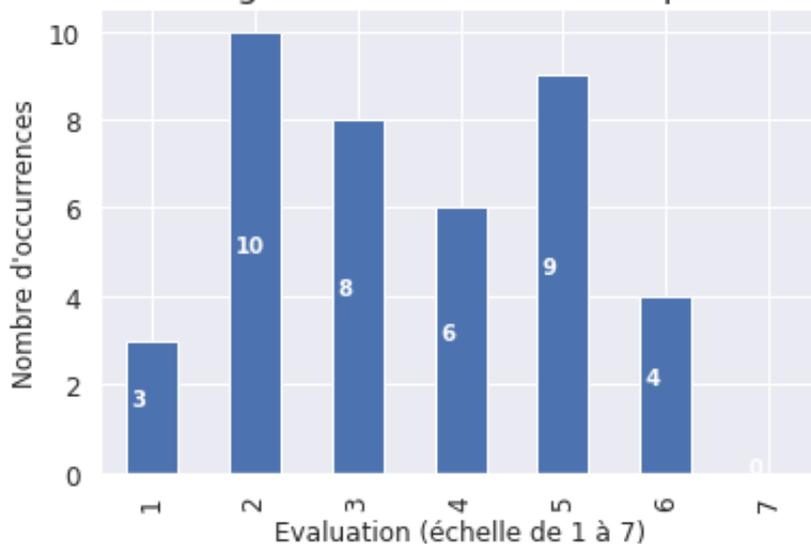
E.III Évaluation du système d'auto-évaluation des émotions

Questionnaire Siegert - identification des émotions



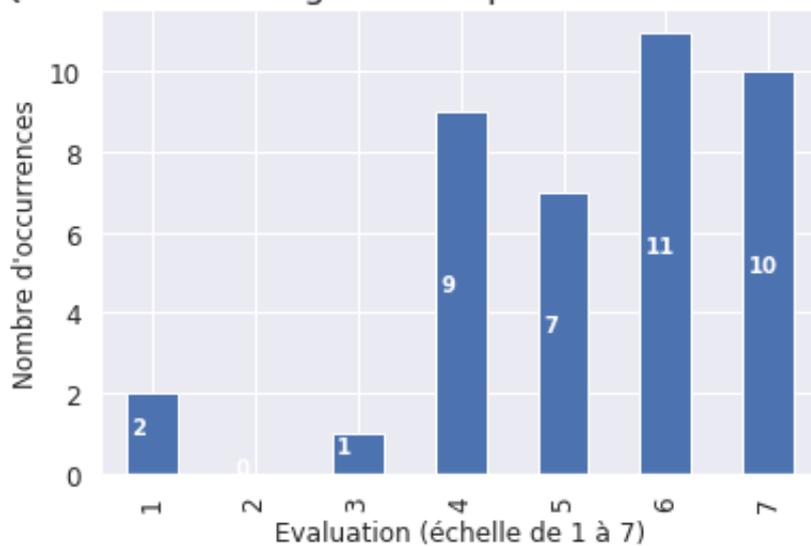
Distribution des réponses à la question : Sur la roue des émotions, il a été facile pour moi d'identifier les émotions que j'ai ressenties.

Questionnaire Siegert - Emotions et description des ressentis



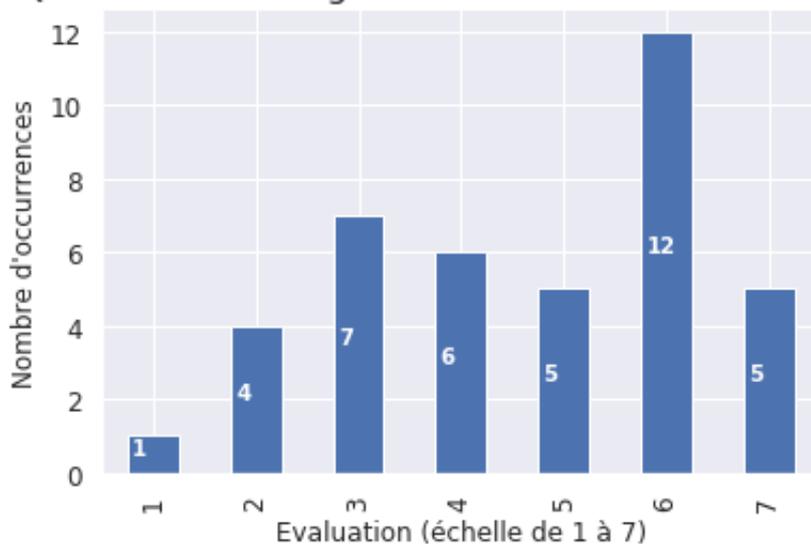
Distribution des réponses à la question : Sur la roue des émotions, les émotions proposées étaient suffisantes pour décrire ce que j'ai ressenti.

Questionnaire Siegert - Compréhension des émotions



Distribution des réponses à la question : Sur la roue des émotions, il était facile pour moi de comprendre ce que représentait les émotions proposées.

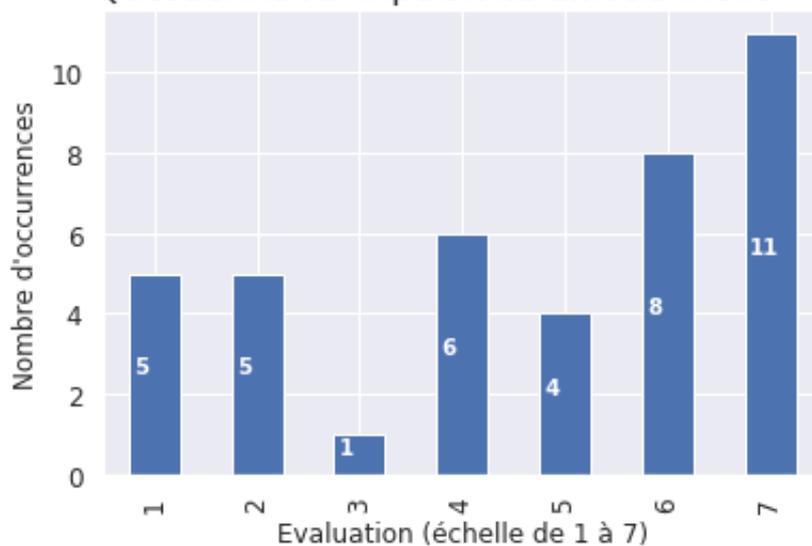
Questionnaire Siegert - Ressentis avant interaction



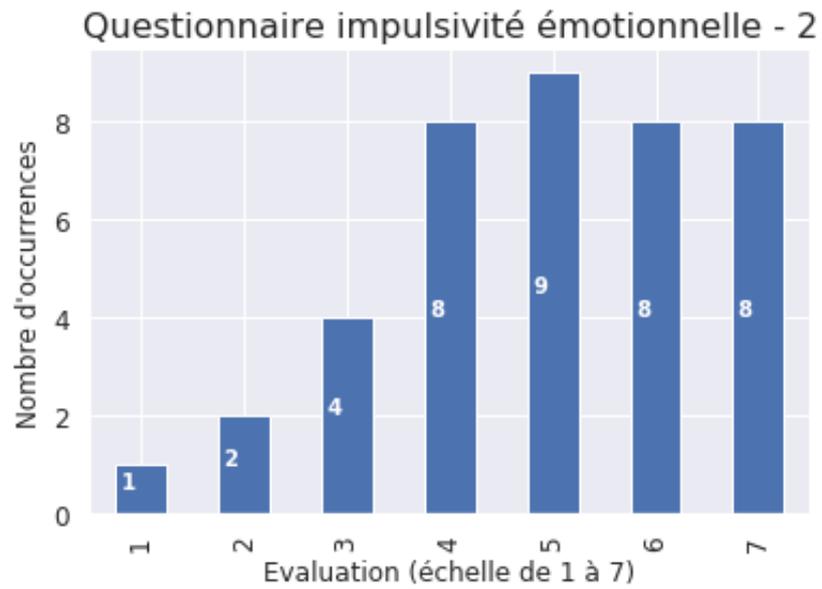
Distribution des réponses à la question : J'avais une idée claire des émotions que j'ai ressenties avant d'interagir avec la roue des émotions.

E.IV Évaluation de l'impulsivité émotionnelle

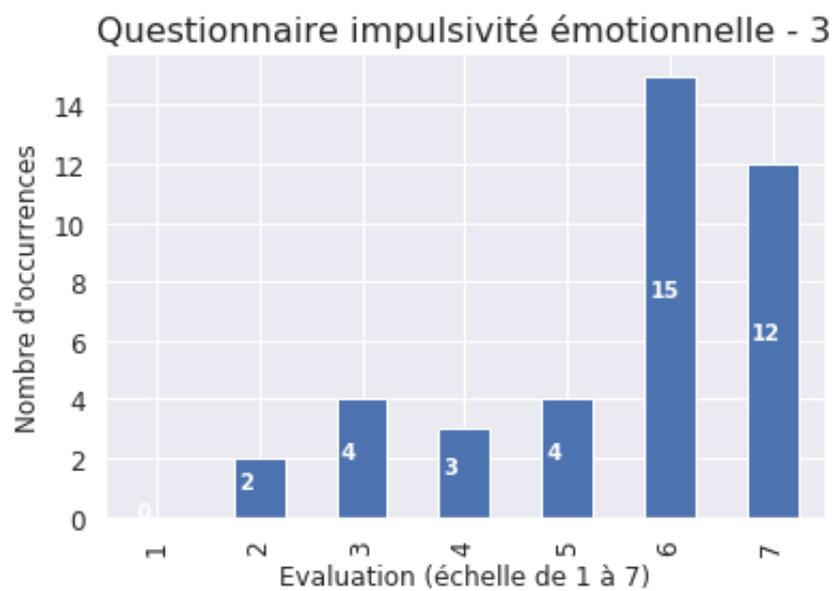
Questionnaire impulsivité émotionnelle - 1



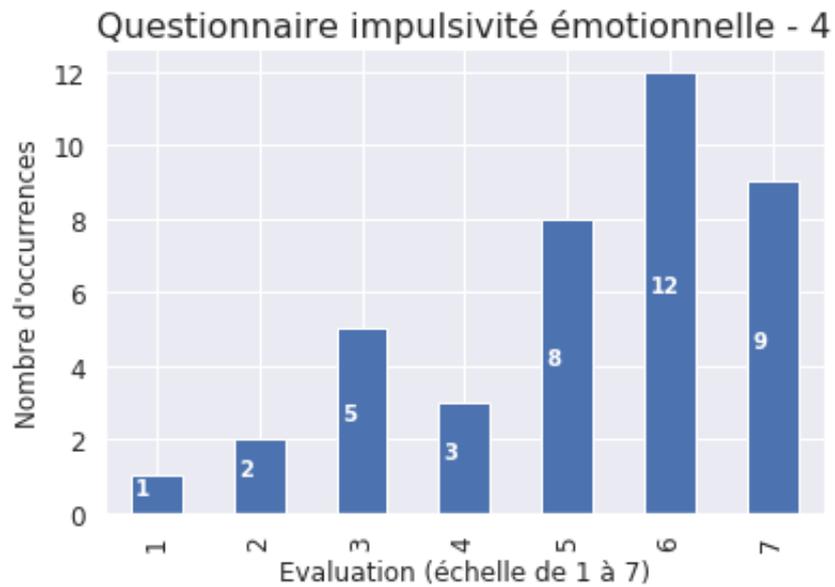
Distribution des réponses à la question : Je pleure parfois devant les films tristes.



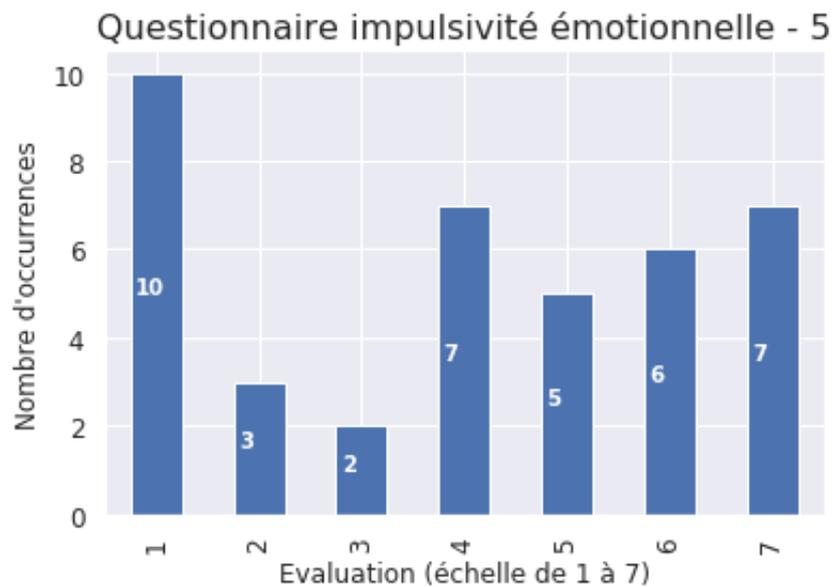
Distribution des réponses à la question : Mon corps réagit très fortement à des situations émotionnelles.



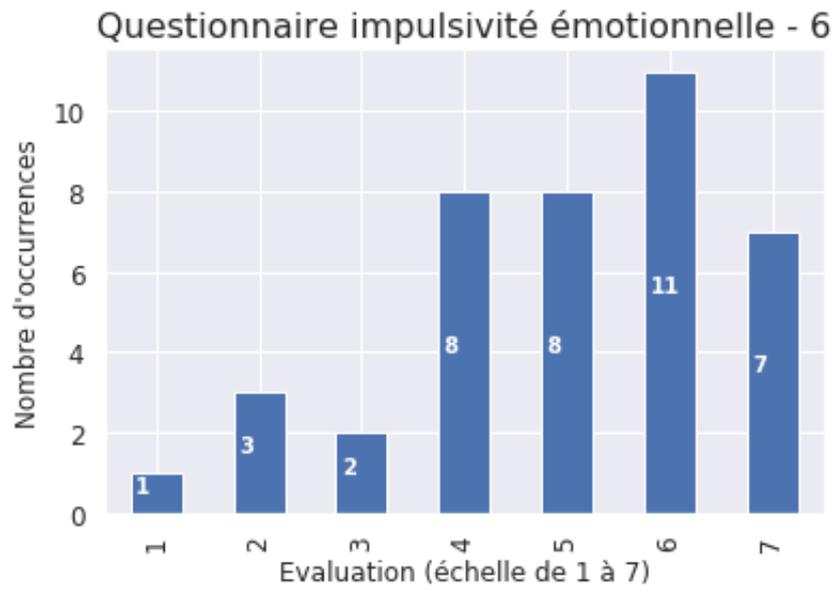
Distribution des réponses à la question : Je ressens de fortes émotions.



Distribution des réponses à la question : Je suis parfois incapable de cacher mes sentiments, même si j'aimerais pouvoir le faire.



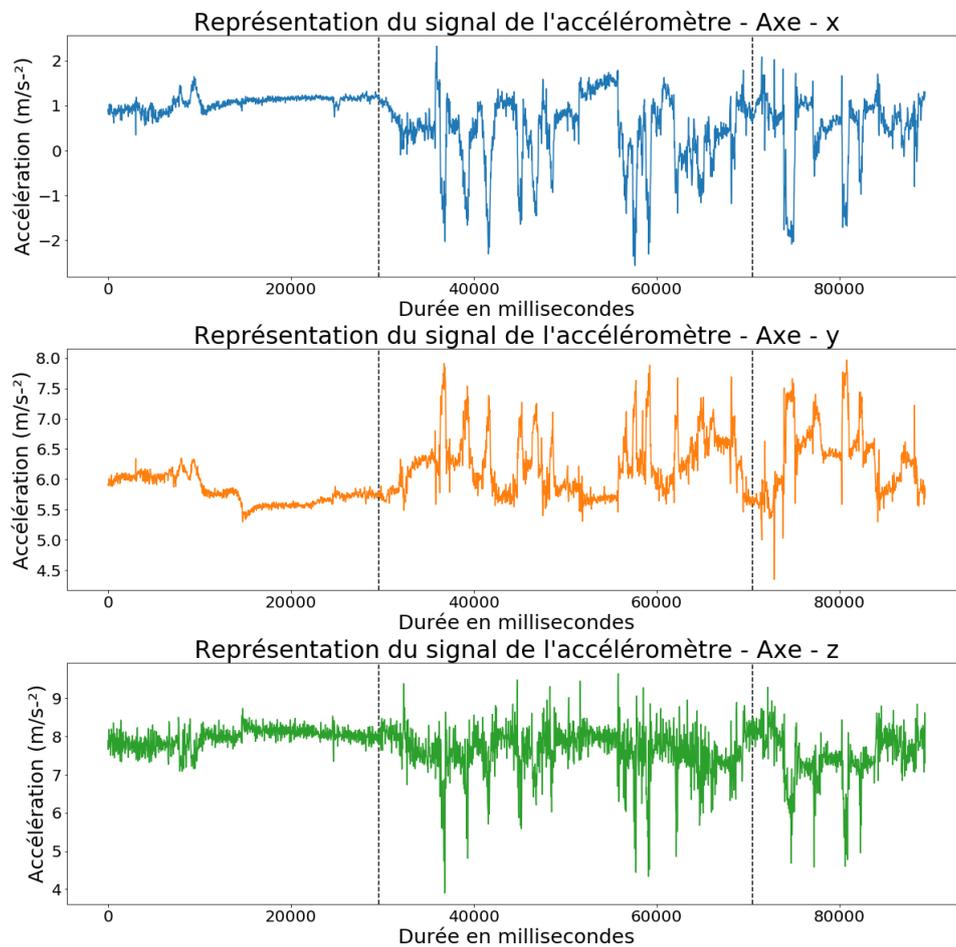
Distribution des réponses à la question : J'ai vécu des périodes dans ma vie où je n'arrivais pas à arrêter de pleurer même si j'essayais.



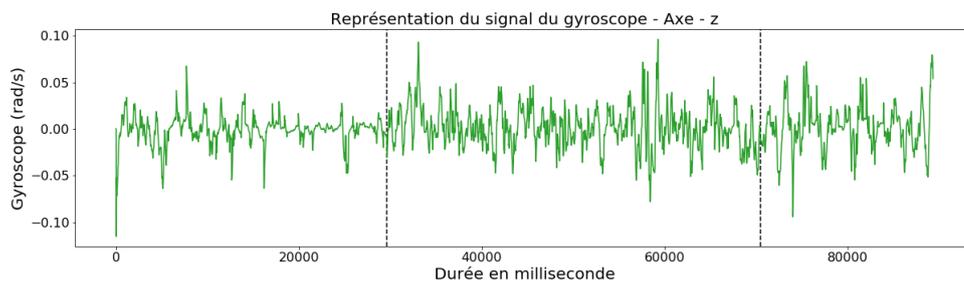
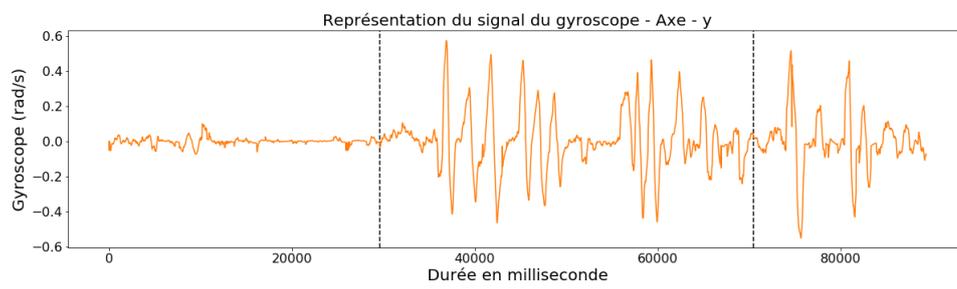
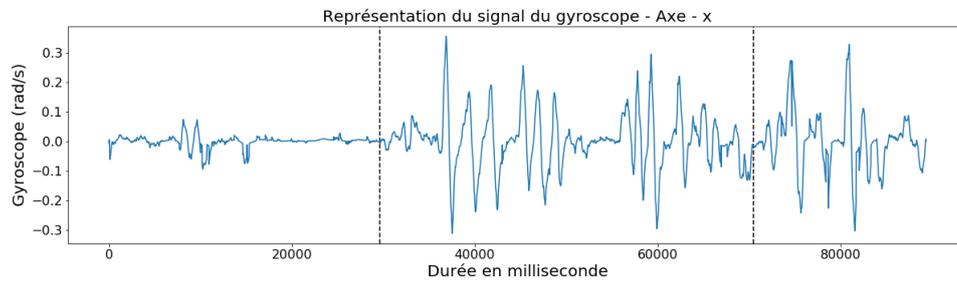
Distribution des réponses à la question : Je ressens des émotions avec beaucoup d'intensité.

F. Représentation des signaux bruts pour une tâche de l'expérimentation principale

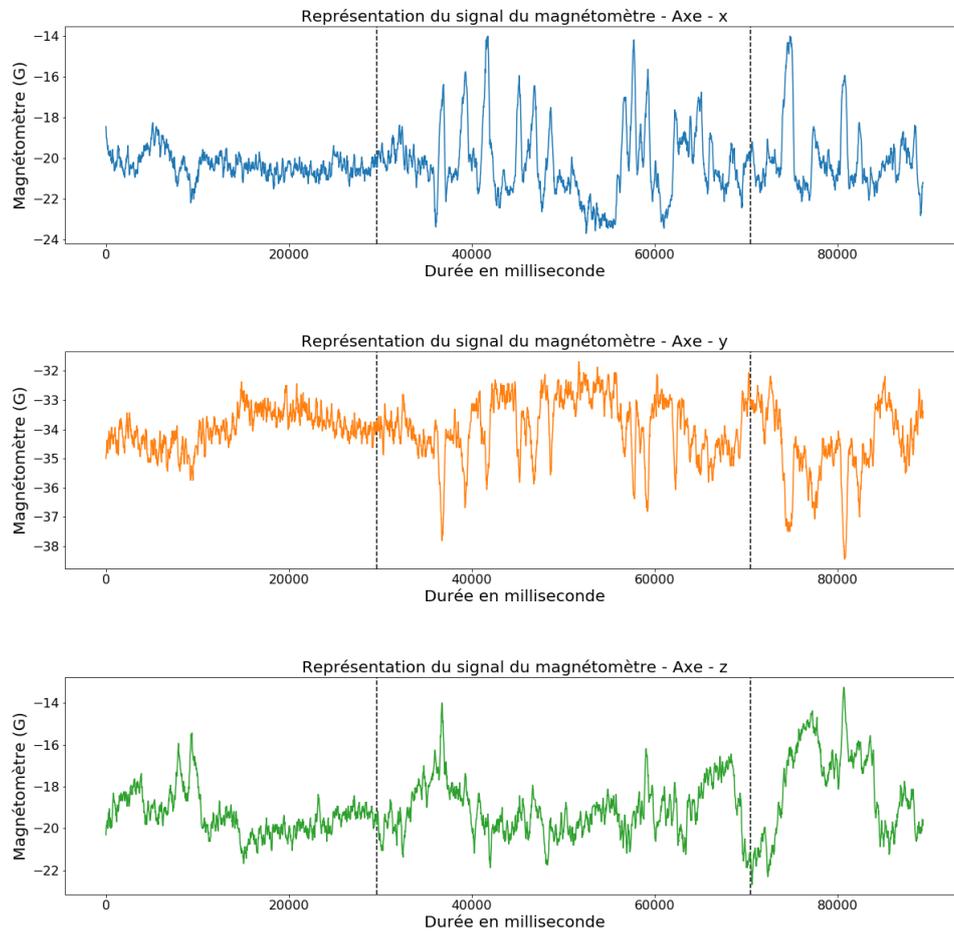
F.I Capteurs du smartphone



Représentation d'une série temporelle de signal d'accéléromètre pour une tâche

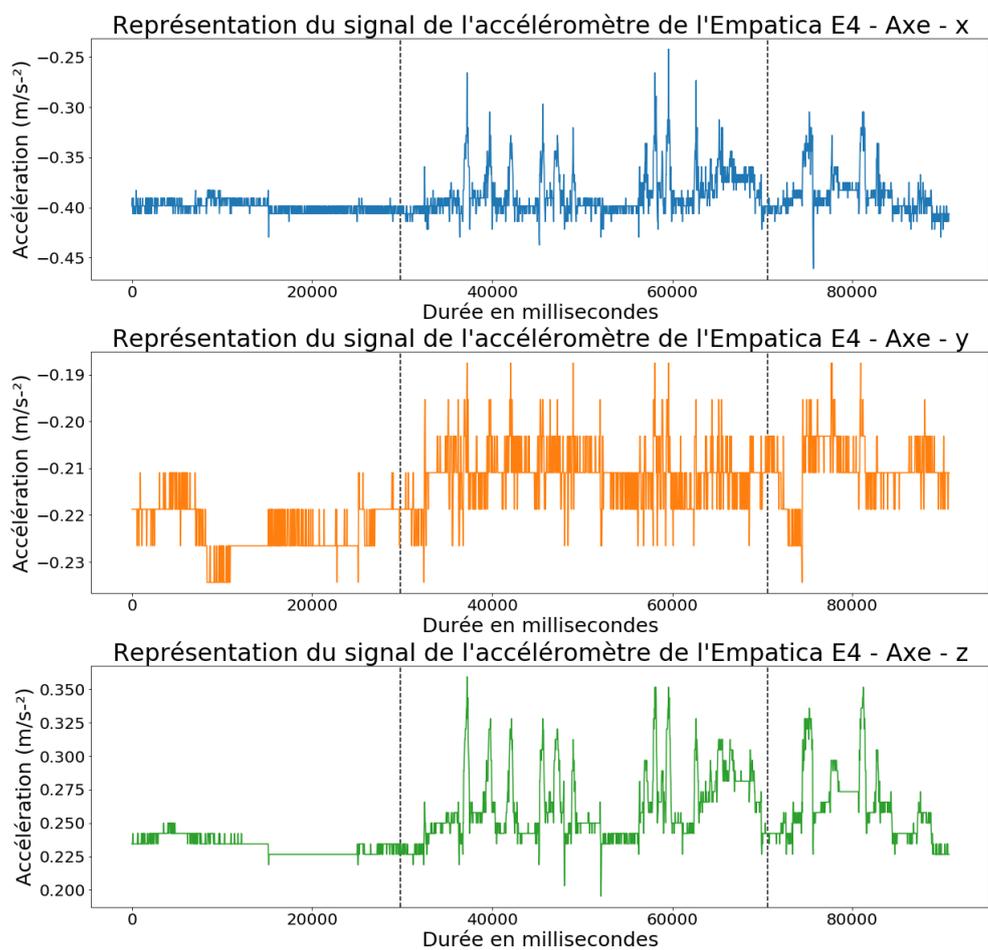


Représentation d'une série temporelle de signal de gyroscope pour une tâche

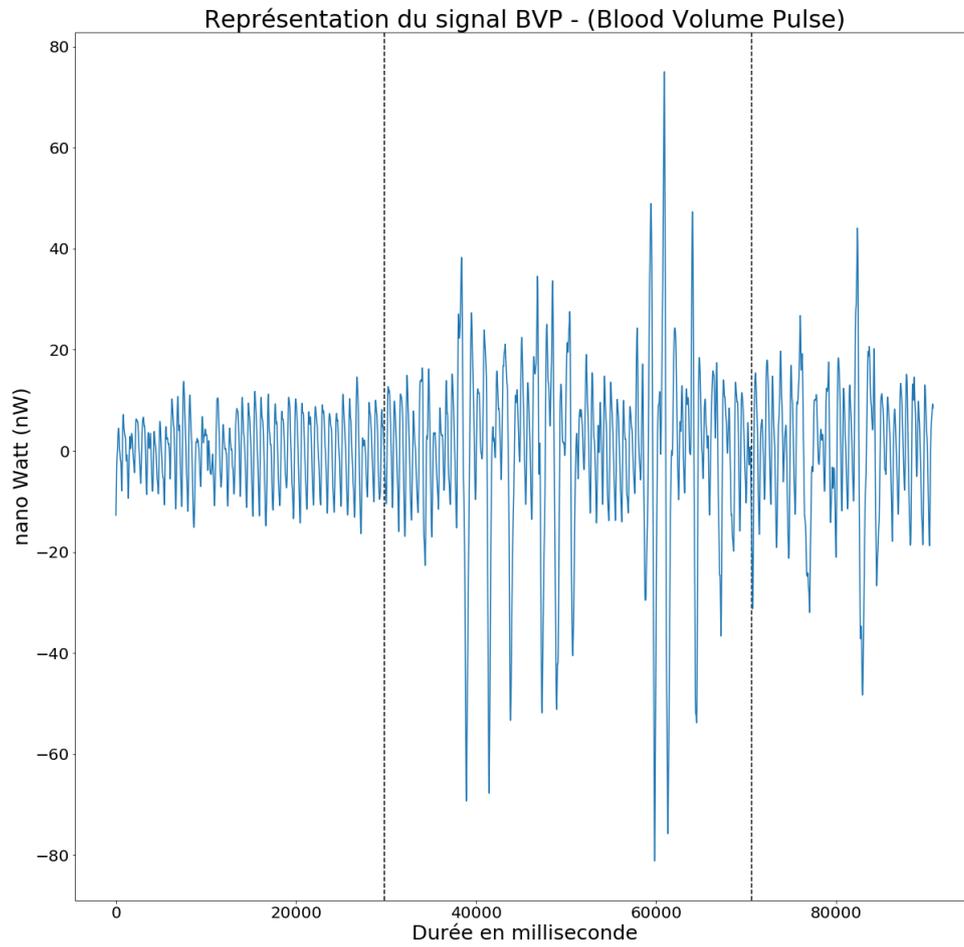


Représentation d'une série temporelle de signal de magnétomètre pour une tâche

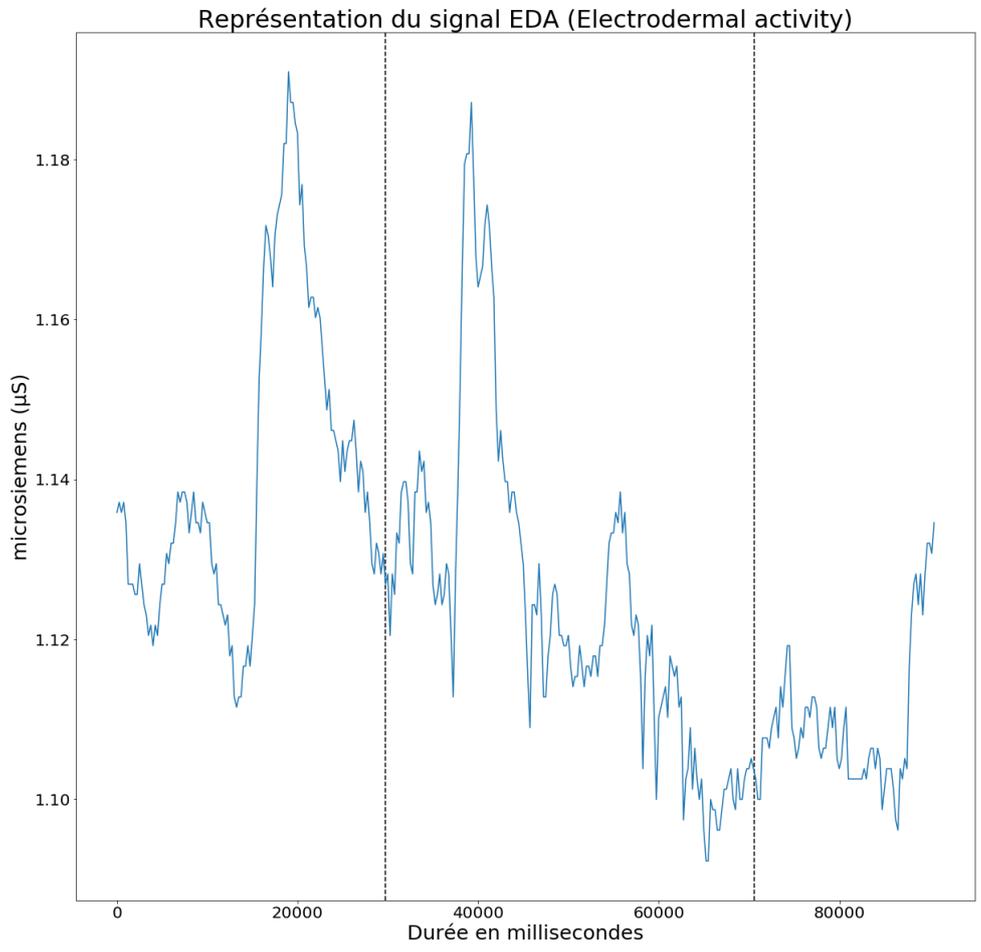
F.II Capteurs de la montre Empatica E4



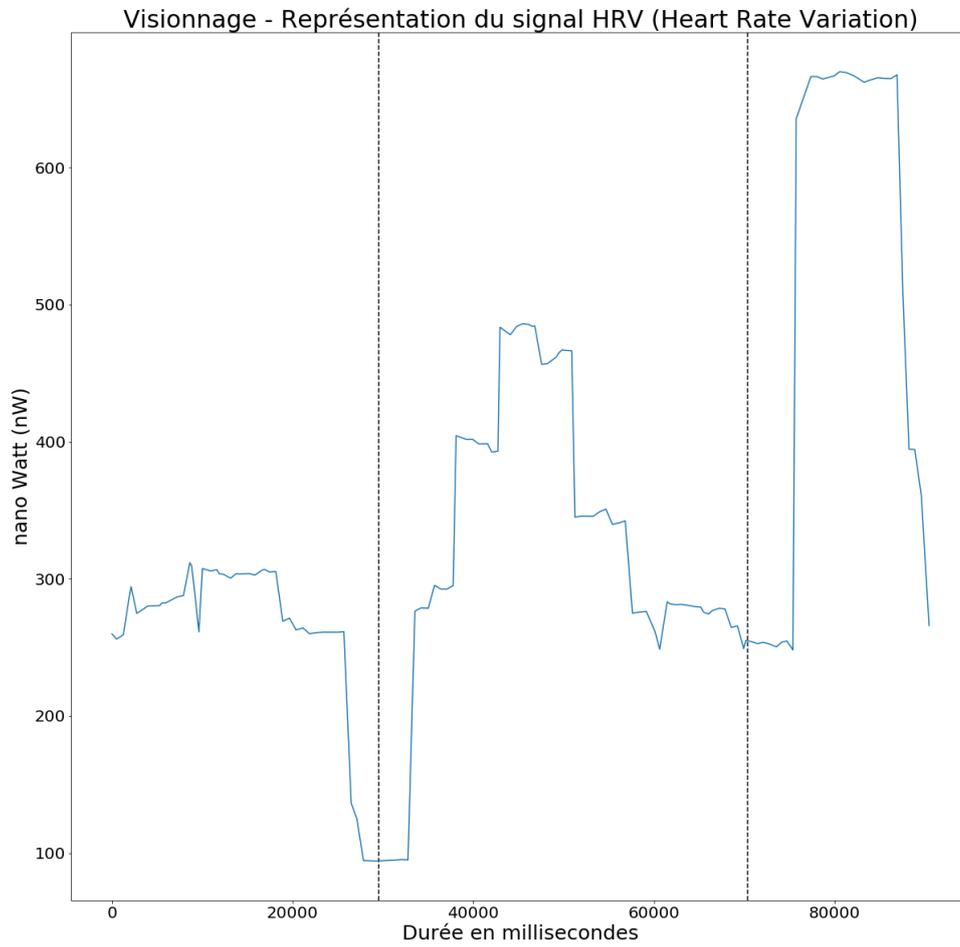
Représentation d'une série temporelle de signal d'accéléromètre pour une tâche



Représentation d'une série temporelle de signal de photopléthysmographie pour une tâche



Représentation d'une série temporelle de signal de capteur de conductance électrodermale pour une tâche



Représentation d'une série temporelle de signal de capteur de variation de fréquence cardiaque pour une tâche

G. Détails des calculs sur les séries temporelles

Descripteurs des signaux temporels

Nom du descripteurs	Description
Moyenne	$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$
Écart type	$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$
Variance	$var = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$
Médiane	$median = \frac{1}{2}(x_{(n+1)/2} + x_{(n+1)/2})$
Médiane absolue des écarts	$MAD = \frac{1}{n} \sum_{i=1}^n x_i - \bar{x} $
Maximum	$max = \max_i(x_i)$
Minimum	$min = \min_i(x_i)$
Écart Inter-quartile	$IQR = Q_3 - Q_1$
Énergie temporelle	$eng = \sum_{i=1}^n x_i^2$
Coefficient de corrélation	$Corr(X, Y) = \frac{cov(X, Y)}{\sigma_X \cdot \sigma_Y}$ $cov(X, Y) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$
Asymétrie	$s = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^3 \sigma^{-3}$
Kurtosis	$k = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^4 \sigma^{-4}$
Énergie fréquentielle	$engf = \frac{1}{n} \sum_{i=1}^n x_i^2$
Entropie	$ent = - \sum_{i=1}^n x_i \log(x_i)$
Entropie spectrale	Calcul de la Densité spectrale de puissance : $P(x_i) = \frac{1}{n} x_i ^2$ $p_i = \frac{1}{n} x_i ^2$ $ES = - \sum_{i=1}^n p_i \log(p_i)$