



HAL
open science

**Modèle Boîte-Blanche pour l'Evaluation du Stress à
Partir de Mesures Physiologiques et Applicabilité en Vie
Réelle**
Gaël Vila

► **To cite this version:**

Gaël Vila. Modèle Boîte-Blanche pour l'Evaluation du Stress à Partir de Mesures Physiologiques et Applicabilité en Vie Réelle. Traitement du signal et de l'image [eess.SP]. Université Grenoble Alpes [2020-..], 2021. Français. NNT : 2021GRALT014 . tel-03275929

HAL Id: tel-03275929

<https://theses.hal.science/tel-03275929v1>

Submitted on 1 Jul 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

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITE GRENOBLE ALPES

Spécialité : **Signal, Image, Parole, Télécoms**

Arrêté ministériel : 25 mai 2016

Présentée par

Gaël VILA

Thèse dirigée par **Christelle GODIN**,
Codirigée par **Sylvie CHARBONNIER**,
Encadrée par **Aurélie CAMPAGNE**

Préparée au sein du **Laboratoire Signal et Systèmes de Capteurs (LSSC)**, Département **Systèmes (DSYS)** du **CEA-Léti**, dans l'**École Doctorale d'Electronique, Electrotechnique, Automatique et Traitement du Signal (ED-EEATS)**

Modèle Boîte-Blanche pour l'Evaluation du Stress à Partir de Mesures Physiologiques et Applicabilité en Vie Réelle

Thèse soutenue publiquement le **18 Mars 2021**,
devant le jury composé de :

M^{me} Alice CAPLIER

Professeure à Grenoble INP - UGA, *Présidente du Jury*

M. Frédéric DEHAIS

Professeur à l'ISAE SUPAERO, *Rapporteur*

M. Mario CHAVEZ

Chargé de Recherche au CNRS, *Rapporteur*

M. Norbert NOURY

Professeur à l'Université de Lyon, *Examinateur*

M^{me} Christelle GODIN

Ingénieure-Chercheuse au CEA-Léti, *Directrice*

M^{me} Sylvie CHARBONNIER

Maître de Conférence à l'Université Grenoble-Alpes, *Co-Directrice*

M^{me} Aurélie CAMPAGNE

Maître de Conférence à l'Université Grenoble-Alpes, *Encadrante*



Remerciements

J'adresse les premiers mots de cet ouvrage à ses premiers destinataires. Merci à MM. Frédéric Dehais et Mario Chavez, rapporteurs du présent manuscrit, d'avoir accepté d'évaluer mes travaux de thèse. Merci à M. Norbert Noury et à M^{me} Alice Caplier, d'avoir accepté d'en être examinateur et examinatrice. Je vous souhaite une excellente lecture.

Je dois les chapitres qui suivent au support assidu et conjugué de Christelle Godin, ma directrice de thèse, de Sylvie Charbonnier, ma co-directrice, et d'Aurélie Campagne, mon encadrante. Pour apprécier la contribution de chacune d'elles, on pourrait comparer mon travail à un arbre fruitier. Sans le soutien de Christelle, vous (lecteur, lectrice) auriez sous les yeux une jolie prairie de pages blanches. Christelle a planté une graine dans un terreau fertile, lui a fourni les bons tuteurs et a veillé à ce qu'elle ne manque de rien : elle a permis la croissance de ce projet durant quatre belles années. Sans l'appui de Sylvie, vous auriez sous les yeux une plante sauvage : si l'on n'est pas expert en plantes, on ne sait pas bien ce que c'est. Les conseils de Sylvie m'ont permis d'élaguer beaucoup, de greffer un peu ; ils m'ont donné des idées claires pour préciser mon travail et une aide précieuse pour l'écriture du manuscrit. Sans les apports d'Aurélie, vous auriez sous les yeux un arbre cerné de ronces. Lorsqu'on ne sait pas bien ce qu'on cherche, l'étude du stress peut être un terrain très épineux ; et le savoir-faire d'Aurélie, ses retours attentifs, étaient nécessaires pour éclaircir ce sujet avec toute la rigueur que celui-ci méritait. Je les remercie toutes les trois pour avoir su encadrer mes travaux de thèse dans un juste équilibre entre l'exigence d'une production scientifique aboutie et la bienveillance d'une relation humaine. J'espère que de leur point de vue aussi, leur travail a porté ses fruits.

Au-delà de ces trois personnes, ce sont trois laboratoires qui ont collaboré pour produire ces recherches à la croisée des domaines scientifiques. Je remercie tous les membres du Laboratoire Signal et Systèmes de Capteurs au CEA-Leti, pour m'avoir chaleureusement accueilli et pour m'avoir offert un environnement de travail de qualité. Merci à Viviane Cattin, responsable du laboratoire, pour sa disponibilité et son engagement. Merci au Laboratoire de Psychologie et Neurocognition, pour ses ressources et son expertise sur les questions expérimentales. Merci au GIPSA, pour m'avoir accueilli à la fin de cette thèse et permis de mener de front l'écriture et l'enseignement dans les meilleures conditions. Merci à Anne Guérin et à Stéphane Bonnet, membres de mon comité de suivi de thèse, pour leur écoute et leurs conseils.

Au cours de cette expérience, j'ai fait de belles rencontres. A Etienne Labyt, qui a suivi ma thèse de si près que je le considère volontiers comme mon quatrième encadrant, merci pour ces mains tendues tout au long de mon parcours : elles avaient la chaleur d'une amitié. A mes camarades de classe dans l'étude du stress au laboratoire : Simon Ollander, Oumayma Sakri, Violette Hoareau, merci pour ces précieuses collaborations. A mes compagnons de répit, Alex Perreira da Silva, Chhayarith Heng Uy, merci pour les moments de convivialité partagés entre deux lignes de code ou de ce manuscrit.

Une moitié de ces pages a été rédigée au cours d'un épisode historique qui laissait peu d'harmonie entre effort et réconfort. Merci aux amis que j'ai eu la chance de voir dans ce contexte, et à ceux que je reverrai bientôt. Merci à ma famille, pour son soutien et pour son indulgence à mon égard, quand je laisse parfois l'absence se substituer à l'éloignement. Et merci à Agathe, enfin, pour être Agathe au quotidien.

Résumé en Français

Le stress aigu est une réponse adaptative du corps aux épreuves de la vie quotidienne. Les corrélats physiologiques du stress sont connus et mesurables à travers le rythme cardiaque, la conductance de la peau ou l'activité respiratoire. Récemment, l'essor des capteurs physiologiques portables a donné une perspective concrète à la détection d'épisodes de stress dans la vie quotidienne. Mais les corrélats physiologiques du stress ne lui sont pas spécifiques : ils sont également sensibles à de nombreux facteurs confondants. En outre, la mesure physiologique en vie réelle est sujette à de nombreux artefacts et les modèles actuellement proposés manquent d'une stratégie de gestion efficace de la qualité des mesures. Enfin, la littérature a tendance à promouvoir des modèles de type boîte-noire, dont la décision ne peut être interprétée dans un but thérapeutique.

Pour répondre à ces problématiques, nous proposons des méthodes de type boîte-blanche qui aboutissent à un modèle de détection du stress susceptible de fonctionner dans la vie quotidienne. Nous proposons d'abord une sélection des caractéristiques sensibles au stress et peu sensibles aux facteurs confondants. Sur la base des caractéristiques sélectionnées, nous proposons un détecteur de stress robuste aux artefacts et qui adopte une structure modulaire, afin de s'adapter en temps réel à la qualité des mesures physiologiques. L'idée derrière ce modèle est de pouvoir se séparer sans nouvel apprentissage des caractéristiques momentanément indisponibles ou trop bruitées. Pour valoriser cette capacité adaptative de notre détecteur de stress, nous proposons finalement un algorithme permettant d'estimer en temps réel la qualité des données de rythme cardiaque. Couplé à cet indice de qualité, notre modèle de détection du stress constitue un pas de plus vers une mesure objective de ce phénomène et de sa prévalence dans notre quotidien.

English Abstract

Acute stress is our body's adaptive response to the challenges of daily life. Stress has well-known physiological correlates, which can be monitored through heart rate, skin conductance or respiratory activity. Over the past decade, the rise of wearable physiological sensors has provided tangible insights into the detection of stressful episodes in daily life. Yet, physiological stress correlates are also sensitive to many confounding factors. Moreover, real-life physiological measurements involve frequent artifacts ; and current literature models lack an effective strategy for managing sensor data quality. Finally, current accuracy-based approaches tend to promote black-box models whose decision cannot be interpreted for therapeutic purposes.

To tackle these issues, we provide white-box methods to design a stress detection model able to operate in real-life settings. We first propose a feature selection method that identifies features that are sensitive to stress and relatively insensitive to confounding factors. We then develop a stress detection model designed to withstand artefactual data ; and which incorporates a modular structure to timely adapt to the quality of physiological measurements. The idea behind this model is to isolate, without need for training, features that are temporarily unavailable or too noisy for stress detection. To capitalize on such an adaptive ability, we finally provide an algorithm to timely estimate the quality of heart rate data. Combined with this quality index, our stress detection model is one more step towards an objective measurement of this phenomenon and its occurrence in our daily lives.

Table des Matières

Remerciements	3
Résumé en Français	4
English Abstract	4
Table des Matières	5
Index des Figures	8
Index des Tableaux	10
Index des Symboles & Acronymes	12
Chapitre Un - Introduction	14
I. Contexte et Motivations	14
II. Problématiques de Recherche	15
III. Plan du Manuscrit	15
Chapitre Deux - Stress : Réponse Adaptative, Concept Versatile	17
I. Définition Générique	17
II. Identité Physiologique du Stress	18
II.A. Approche Clinique : Deux Points de Vue Complémentaires	18
II.B. Marqueurs Neurophysiologiques du Stress	19
II.C. Modèle Synthétique du Stress Physiologique	21
III. Définitions Croisées et Facteurs Confondants	22
III.A. Facteurs Physiologiques	22
III.B. Facteurs Psychologiques	23
III.C. Synthèse : l'Etat de Stress et ses Facteurs Confondants	25
IV. Méthodes Expérimentales pour l'Etude du Stress	26
IV.A. Évaluer le Stress	26
IV.B. Induire le Stress	27
IV.C. Expression des Facteurs Confondants dans une Base de Données Expérimentales	29
Chapitre Trois - Conception d'un Détecteur de Stress : Etat de l'Art	33
I. Mesures Physiologiques	34
I.A. Activité Cardiovasculaire	34
I.B. Activité Electrodermale	37
I.C. Autres Mesures Disponibles : Pertinence et Perspectives	38
II. Caractéristiques Extraites des Signaux Périphériques	41
II.A. Caractéristiques Extraites du Signal de Rythme Cardiaque	41
II.B. Caractéristiques Extraites du Signal Electrodermal	44
II.C. Utilisation des Caractéristiques dans un Modèle de Stress	45
III. Apprentissage d'un Détecteur de Stress	47
III.A. Apprentissage Supervisé	47
III.B. Comparaison de Modèles Publiés dans la Littérature	52
IV. Obstacles à l'Utilisation d'un Modèle en Vie Réelle	58
IV.A. Validité d'un Modèle de Stress	58
IV.B. Qualité des Mesures dans la Vie Quotidienne	60
IV.C. Analyse des Tentatives de Détection du Stress dans la Vie Quotidienne	62

V. Positionnement de la Thèse _____	67
V.A. Principes et Limites d'une Méthode Standard pour l'Estimation du Stress en Vie Réelle _____	67
V.B. Questions de Recherche _____	69
Chapitre Quatre - Base de Données pour l'Etude du Stress en Conditions Contrôlées _____	72
I. Protocole et Campagne d'Expériences _____	72
I.A. Cadre Expérimental et Recrutement des Participants _____	72
I.B. Tâches Expérimentales _____	73
I.C. Déroulement Pratique de l'Expérience _____	74
I.D. Evaluations Psychologiques _____	75
II. Enregistrements Physiologiques _____	75
II.A. Signaux Recueillis _____	75
II.B. Caractéristiques Extraites _____	76
Chapitre Cinq - Sélection de Caractéristiques Sensibles à la Charge Affective _____	79
I. Modélisation de l'Effet du Stress et des Facteurs Confondants sur une Caractéristique _____	80
I.A. Sources d'Influence Observables sur une Caractéristique Physiologique _____	80
I.B. Hypothèses sur la Variation d'une Caractéristique en Fonction des Contributions _____	82
II. Méthodes pour la Sélection de Caractéristiques Sensibles et Spécifiques au Stress _____	85
II.A. Estimation de la Part du Stress et des Facteurs Confondants _____	85
II.B. Critères de Sélection des Caractéristiques _____	88
II.C. Stratégie de Sélection des Caractéristiques _____	91
III. Résultats et Discussion _____	94
III.A. Validation du Protocole Expérimental _____	94
III.B. Sélection des Caractéristiques _____	96
III.C. Apports du Protocole Expérimental _____	97
III.D. Limites de cette Etude _____	100
IV. Synthèse : Principes et Apports de la Méthode de Sélection Proposée _____	102
Chapitre Six - Modèle Adaptatif pour l'Identification du Stress Emotionnel _____	104
I. Principes et Limites de l'Approche Standard pour la Production d'un Modèle Boîte-Blanche _____	105
I.A. Normalisation des Caractéristiques _____	105
I.B. Apprentissage en Boîte-Blanche _____	106
II. L'indice de Stress : un Modèle Adaptatif _____	107
II.A. Principe Général _____	108
II.B. Construction de l'Indice de Stress et Synthèse des Paramètres _____	110
II.C. Apprentissage _____	111
III. Comparaison de Modèles pour la Détection du Stress _____	112
III.A. Question de Recherche et Méthodologie Générale _____	112
III.B. Mesure des Performances d'un Détecteur de Stress _____	113
III.C. Expérience 1 : Comparaison d'Algorithmes pour la Détection du Stress en Milieu Contrôlé _____	116
III.D. Expérience 2 : Application de l'Indice de Stress dans un Contexte de Mesures Dégradées _____	119
IV. Résultats et Discussion _____	121
IV.A. Expérience 1 : Comparaison des Modèles de Stress _____	121
IV.B. Expérience 2 : Application dans un Contexte de Mesures Dégradées _____	125
IV.C. Limites de cette Etude _____	128
IV.D. Discussion _____	129
Chapitre Sept - Qualité des Mesures dans la Vie Quotidienne _____	133
I. Contexte _____	134
I.A. Propriétés des Signaux Cardiaques _____	134

I.B. Défis de la Mesure Cardiaque en Condition Ambulatoire _____	135
I.C. Synthèse et Objectifs _____	137
II. Matériel : Base de Données en Vie Réelle _____	138
II.A. Protocole Expérimental _____	138
II.B. Traitements des Signaux _____	140
III. Méthodes _____	144
III.A. Estimation de la Part d'Echantillons Manquants dans un Signal IBI _____	144
III.B. Comparaison PPG-ECG _____	147
IV. Résultats _____	148
IV.A. Analyse des Segments Complets des Signaux IBI _____	148
IV.B. Fiabilité de la Mesure PPG _____	149
V. Discussion _____	153
V.A. Sur l'Observation de l'Activité Cardiaque en Vie réelle _____	153
V.B. Sur l'Indice de Carence _____	155
Chapitre Huit - Conclusions & Perspectives _____	158
I. Synthèse des Contributions _____	158
II. Principaux Résultats _____	158
II.A. Recherche de Caractéristiques Sensibles au Stress _____	158
II.B. Proposition d'un Modèle Adaptatif : l'Indice de Stress _____	159
II.C. Estimation de la Qualité d'un Signal de Rythme Cardiaque _____	159
II.D. Synthèse des Apports Méthodologiques _____	160
III. Perspectives de Recherche _____	161
Annexes _____	162
Annexe 1. Stress Physique ou Stress Mental : l'Exemple de l'Activité Cardiovasculaire ____	163
Annexe 2. Modèles Psychologiques du Stress _____	165
Annexe 3. Stratégies de Neutralisation de l'Effet du Sujet sur les Corrélats du Stress ____	172
Annexe 4. Modèles élaborés en conditions contrôlées _____	174
Annexe 5. Extraction des Caractéristiques _____	175
Annexe 6. Tests Statistiques _____	182
Annexe 7. Résultats de la Sélection des Caractéristiques _____	188
Annexe 8. Pertinence des Estimateurs : Compléments d'Analyses _____	192
Annexe 9. Apprentissage en boîte blanche _____	197
Annexe 10. Interpolation de l'Indice Simple _____	203
Annexe 11. Algorithmes d'apprentissage _____	206
Annexe 12. Disponibilité du Signal IBI pour le Calcul d'une Caractéristique Cardiaque ____	210
Annexe 13. Valorisation des Travaux _____	212
Bibliographie _____	213

Index des Figures

FIGURE 2.1 – SCHEMA DE LA REPONSE NEUROENDOCRINIENNE AU STRESS. FIGURE TIREE DE (ULRICH-LAI & HERMAN, 2009).	20
FIGURE 2.2 : SCHEMA D'UNE ECHELLE VISUELLE ANALOGIQUE A VALEURS CONTINUES (A) ET A VALEURS DISCRETES (B)	27
FIGURE 3.1 : AXES DE RECHERCHE POUR LA CONSTRUCTION D'UN MODELE DE STRESS (CAS DE L'APPRENTISSAGE SUPERVISE)	33
FIGURE 3.2 : DISPOSITIFS DE MESURE PORTABLES POUR L'ELECTROCARDIOGRAPHIE. (A) ZEPHYR BIOHARNESS 3 (B) HEXOSKIN	35
FIGURE 3.3 – COMPARAISON DE BATTEMENTS CARDIAQUES EN ECG (BLEU) ET PPG (ROUGE) (THOMAS ET AL, 2016)	35
FIGURE 3.4 – DISPOSITIFS PORTABLES DE MESURE PPG. (A) EMPATICA E4 (B) SPIRE HEALTH TAG	36
FIGURE 3.5 – COMPARAISON DE BATTEMENTS CARDIAQUES EN ECG (EN BAS) ET EN CARDIOGRAPHIE D'IMPEDANCE (ICG, EN HAUT) (VAN LIEN ET AL, 2015)	36
FIGURE 3.6 – IMPEDANCES EQUIVALENTES A L'INTERFACE ENTRE LA PEAU ET UNE ELECTRODE. LES DIFFERENCES DE POTENTIEL AUX INTERFACES NE SONT PAS REPRESENTEES.	37
FIGURE 3.7 – IMPEDANCES EQUIVALENTES A L'INTERFACE ENTRE LA PEAU ET L'ELECTRODE DANS LE CAS D'UNE MESURE NON-INVASIVE LES DIFFERENCES DE POTENTIEL AUX INTERFACES NE SONT PAS REPRESENTEES.	37
FIGURE 3.8 – DISPOSITIFS DE MESURE DE L'AED EN CONDITION AMBULATOIRE. (A) STOPSLEEP (B) EMBRACE	38
FIGURE 3.9 – SCHEMA D'UNE REPONSE EN CONDUCTANCE DE LA PEAU (SC) ET SES PARAMETRES, RELATIVEMENT A LA COMPOSANTE TONIQUE DU SIGNAL (SCL)	44
FIGURE 3.10 – CLASSEMENT DES CARACTERISTIQUES SELON LA REVUE DE (WIJSMAN, 2014)	46
FIGURE 3.11 : SCHEMA DE METHODOLOGIE GENERALE EN APPRENTISSAGE SUPERVISE	48
FIGURE 3.12 – FRONTIERES DE DECISION DANS LE CAS DE LA DETECTION PAR ANALYSE DISCRIMINANTE (A) ET PAR LES 15 PLUS PROCHES VOISINS (B), DANS UN ESPACE A DEUX CARACTERISTIQUES. D'APRES HASTIE ET COLLABORATEURS (2009).	50
FIGURE 3.13 : VALIDATION CROISEE : DIVISIONS SUCCESSIVES DE LA BASE D'APPRENTISSAGE.	51
FIGURE 3.14 : LIEN EMPIRIQUE ENTRE LA PERFORMANCE DES MODELES ET LE CONTROLE DE FACTEURS CONFONDANTS	55
FIGURE 3.15 : MATRICE DE CONFUSION DES MODELES PROPOSES PAR GJORESKI ET COLLOBORATEURS (2016) SUR DES MESURES EN VIE REELLE.	64
FIGURE 4.1 : SESSIONS EXPERIMENTALES DU POINT DE VUE D'UN PARTICIPANT	75
FIGURE 4.2 : CONFIGURATION DU DISPOSITIF DE MESURES PHYSIOLOGIQUES.	76
FIGURE 5.1 : SOURCES D'INFLUENCE POSSIBLES SUR UNE CARACTERISTIQUE PHYSIOLOGIQUE	82
FIGURE 5.2 : POSITION RELATIVE DES ESTIMATEURS	86
FIGURE 5.3 : VARIATIONS FORTEMENT COHERENTES (EN BLEU) OU FAIBLEMENT COHERENTES (EN ROUGE), DE DISTRIBUTION QUASI-GAUSSIENNE (A) OU QUELCONQUE (B)	89
FIGURE 5.4 : SENSIBILITE ET SPECIFICITE AU STRESS D'UNE CARACTERISTIQUE F SELON LES 5 ESTIMATEURS T, A, K, E, P	93
FIGURE 6.1 : DISTRIBUTION DES VALEURS DE QUATRE CARACTERISTIQUES POUR DEUX SUJETS DIFFERENTS EN SITUATION DE REPOS, CONTROLE ET STRESS.	105
FIGURE 6.2 : INDICES DE STRESS : CHAINE DE TRAITEMENTS.	108
FIGURE 6.3 : DIMENSIONS DE L'INDICE DE STRESS.	109
FIGURE 6.4: DIVISION DES DONNEES POUR LA VALIDATION CROISEE LOSOTO	115
FIGURE 6.5 : RESEAU DE NEURONES ARTIFICIEL A 1 COUCHE CACHEE DE 3 UNITES (REPRESENTATION DANS MATLAB R2017B)	117
FIGURE 6.6 : PROTOCOLE LOSOTO ADAPTE AU BRUITAGE DE LA BASE DE TEST	120
FIGURE 6.7 : DISTRIBUTION DES CLASSES STRESS ET CONTROLE EN SORTIE DE L'INDICE \mathcal{M} ET D'UNE REGRESSION LOGISTIQUE A DEUX OU TROIS CLASSES, SUR L'ENSEMBLE DES DONNEES DE TEST	122
FIGURE 6.8 : PERFORMANCE MESUREE POUR TROIS MODELES DE STRESS LORSQU'UNE CARACTERISTIQUE PRESENTE DES VALEURS ALEATOIRES	126
FIGURE 7.1 : ECG THORACIQUE (A) ET PPG DU POIGNET GAUCHE (B) ENREGISTRES SIMULTANEMENT SUR UN MEME SUJET	134
FIGURE 7.2 : TEMPS DE TRANSIT DE L'ONDE PPG	135
FIGURE 7.3 : ARTEFACTS DE MOUVEMENT SUR UN SIGNAL ECG (A) ET PPG (B)	136
FIGURE 7.4 : MATERIEL DE MESURES PHYSIOLOGIQUES ET SIGNAUX DISPONIBLES POUR CHAQUE APPAREIL	138
FIGURE 7.5 : PROTOCOLE SEMAINE DE TRAVAIL – JOURNEE DE MESURES DU POINT DE VUE D'UN PARTICIPANT	139

FIGURE 7.6 : SYNCHONISATION DES SIGNAUX D'ACCELEROMETRE POUR E4 ET BIOHARNESS _____	142
FIGURE 7.7 : RETRAIT DES VALEURS ABERRANTES SUR UN SIGNAL D'INTERVALLES INTER-BATTEMENTS _____	143
FIGURE 7.8 : PORTION COMPLETE DE SIGNAL IBI _____	144
FIGURE 7.9 : PORTION INCOMPLETE DE SIGNAL IBI _____	145
FIGURE 7.10 : IDENTIFICATION BATTEMENT A BATTEMENT DES SIGNAUX IBI_{BH3} ET IBI_{E4} SUR UNE PERIODE DE TRENTE SECONDES _	147
FIGURE 7.11 : QUALITE GENERALE DU SIGNAL IBI_{E4} _____	150
FIGURE 7.12 : QUALITE DU SIGNAL IBI_{E4} EN FONCTION DE LA QUANTITE DE MOUVEMENT _____	151
FIGURE 7.13 : ERREUR SUR LES TROIS CARACTERISTIQUES CARDIAQUES CALCULEES A PARTIR DU SIGNAL IBI_{E4} _____	152
FIGURE 0.1 – COMMANDE POSSIBLE DU RYTHME CARDIAQUE EN FONCTION DE LA DEMANDE METABOLIQUE EN SITUATION D'EXERCICE (ROWELL, 1993) _____	163
FIGURE 0.2 – COMMANDE POSSIBLE DU RYTHME CARDIAQUE EN FONCTION DE LA DEMANDE METABOLIQUE EN SITUATION D'EXERCICE (ROWELL, 1993) _____	163
FIGURE 0.3 – EXAGERATION DE LA REPOSE ALLOSTATIQUE LORS D'UN EPISODE DE STRESS MENTAL. _____	164
FIGURE 0.4 – REPRESENTATION DE LA CHARGE COGNITIVE : CAUSES (EN VERT) ET CORRELATS (EN BLEU). ADAPTEE DE (PAAS & MERRIËNBOER, 1994), COMPLETEE AVEC LES INFORMATIONS DU TEXTE. _____	169
FIGURE 0.5 : PROCEDE D'EXTRACTION DE CARACTERISTIQUES SUR LE SIGNAL CARDIAQUE _____	177
FIGURE 0.6 : PUISSANCES SPECTRALES LF ET HF MESUREES SUR UN PERIODOGRAMME DE L'INTERVALLE INTER-BATTEMENTS ____	177
FIGURE 0.7 : PROCEDE D'EXTRACTION DE CARACTERISTIQUES SUR LE SIGNAL ELECTRODERMAL _____	179
FIGURE 0.8 : DETECTION ET CORRECTION D'UN ARTEFACTS SUR LE SIGNAL DE CONDUCTANCE DE LA PEAU _____	179
FIGURE 0.9 : DETECTION ET SELECTION DES REPONSES ELECTRODERMALES SUR LE SIGNAL DE CONDUCTANCE LISSE. _____	180
FIGURE 0.10 : PROCEDE D'EXTRACTION DE CARACTERISTIQUES SUR LE SIGNAL RESPIRATOIRE _____	181
FIGURE 0.11 : ILLUSTRATION D'UN TEST BILATERAL (A) ET D'UN TEST D'EQUIVALENCE (B). DEPUIS (LAKENS, 2018) _____	185
FIGURE 0.12 : ESTIMATION DU NOMBRE D'EXEMPLES STANDARD POUR DEMONTRER LA PRESENCE D'UN EFFET EN FONCTION DE LA TAILLE D'EFFET θ (A), OU L'ABSENCE D'EFFET EN FONCTION DE LA LIMITE D'EQUIVALENCE Δ_{EA} LORSQUE $\theta=0$ (B). _____	187
FIGURE 0.13 : LIEN EMPIRIQUE ENTRE LES EFFETS TACHE (θ_T), ACTIVITE (θ_A), CONTRAINTE (θ_K), CONTEXTE (θ_E) ET PRESSION (θ_P). _____	194
FIGURE 0.14 : CONSTRUCTION DE L'INDICE SIMPLE SUR DONNEES SIMULEES _____	204
FIGURE 0.15 : TAUX DE DISPONIBILITE MESURE EN FONCTION DU SEUIL DE TOLERANCE SUR CHAQUE CRITERE DE QUALITE _____	211

Index des Tableaux

TABLEAU 3.1 – MESURES D’INTERET POUR LA DETECTION DU STRESS AIGU (SHARMA & GEDEON, 2012)	34
TABLEAU 3.2 – ÉVOLUTION ATTENDUE DES CARACTERISTIQUES CARDIAQUES STANDARD EN SITUATION DE STRESS ET INFLUENCE PREDOMINANTE DE L’ACTIVITE AUTONOME	43
TABLEAU 3.3 – ÉVOLUTION ATTENDUE DES CARACTERISTIQUES ELECTRODERMALES STANDARD EN SITUATION DE STRESS ET INFLUENCE PREDOMINANTE DE L’ACTIVITE AUTONOME	45
TABLEAU 3.4 : ALGORITHMES D’APPRENTISSAGE COMMUNS	49
TABLEAU 3.5 : ÉCHANTILLON DE MODELES POUR LA DETECTION DU STRESS ET VALIDATION DE TROIS CRITERES D’OBSERVATION.	54
TABLEAU 3.6 : PRECISION DES MODELES DE STRESS EN FONCTION DE CHAQUE CRITERE D’OBSERVATION (MOYENNE ± ECART-TYPE)	55
TABLEAU 3.7 : VALIDITE DU MODELE DE HEALEY ET PICARD (2005) SELON BROEK ET COLLABORATEURS (2009)	59
TABLEAU 3.8 : TENTATIVES DE DETECTION DU STRESS DANS LA VIE QUOTIDIENNE	63
TABLEAU 4.1: RESUME DES TACHES DE LA CAMPAGNE EXPERIMENTALE EN LABORATOIRE	74
TABLEAU 4.2 : RESUME DES CARACTERISTIQUES RETENUES DANS LE CADRE DE CE CHAPITRE	77
TABLEAU 5.1 : COMPOSANTES DU MODELE	83
TABLEAU 5.2 : CALCUL DES CINQ ESTIMATEURS	86
TABLEAU 5.3 : CRITERES STATISTIQUES SUR LA PRESENCE D’UN EFFET DANS UN ECHANTILLON DE DONNEES	90
TABLEAU 5.4 : STATISTIQUES SUR LES ESTIMATEURS POUR LA SELECTION DES TACHES – ECHELLES VISUELLES ANALOGIQUES DE STRESS (EVAS)	94
TABLEAU 5.5 : STATISTIQUES SUR LES ESTIMATEURS POUR LA SELECTION DES TACHES – STAIT-TRAIT ANIETY INVENTORY (STAI-Y-A)	94
TABLEAU 5.6 : CARACTERISTIQUES SELECTIONNEES – SYNTHESE DES TACHES VALIDES	96
TABLEAU 5.7 : CARACTERISTIQUES SELECTIONNEES – CONGRUENCE INTER-TACHES.	97
TABLEAU 5.8 : CARACTERISTIQUES NON SELECTIONNEES PAR NOS CRITERES	98
TABLEAU 5.9 : MATRICE DE CORRELATION POUR LES QUATRE CARACTERISTIQUES FINALEMENT SELECTIONNEES	100
TABLEAU 6.1 : TAILLE DE LA BASE D’APPRENTISSAGE POUR CHAQUE TACHE ET CHAQUE CLASSE	114
TABLEAU 6.2 : TAILLE DES ENSEMBLES D’APPRENTISSAGE ET DE TEST A CHAQUE ETAPE DU PROTOCOLE LOSOTO	116
TABLEAU 6.3 : ERREUR SUR LA MESURE ORIGINALE EN FONCTION DU BRUIT INTRODUIT POUR UNE CARACTERISTIQUE F QUELCONQUE.	119
TABLEAU 6.4 : DETAIL DES MODELES ET DES PRECISIONS MESUREES PAR VALIDATION CROISEE	123
TABLEAU 6.5: COEFFICIENTS DE LA FRONTIERE DE DECISION CALCULES POUR LES MODELES \mathcal{R}^2 ET \mathcal{N}^1 DURANT LA VALIDATION CROISEE	124
TABLEAU 6.6: SEUIL DE REFERENCE ET SEUIL DE CERTITUDE POUR CHAQUE CARACTERISTIQUE DURANT LA VALIDATION CROISEE	124
TABLEAU 7.1 : SELECTION DES SEGMENTS COMPLETS POUR CHAQUE INDICE DE QUALITE ET NIVEAU MAXIMAL D’ACTIVITE SUR CES SEGMENTS	149
TABLEAU 0.1 – CARACTERISATION D’UNE SITUATION SELON LA THEORIE TRANSACTIONNELLE	166
TABLEAU 0.2 – CATEGORIES DE STRESS ET EMOTIONS ASSOCIEES. ADAPTE DE (FOLKMAN & LAZARUS, 1985)	166
TABLEAU 0.3 : STRATEGIES DE NEUTRALISATION DE L’EFFET SUJET.	172
TABLEAU 0.4 : SYNTHESE DE L’ETAT DE L’ART EN MATIERE DE DETECTION DU STRESS SUR DES DONNEES EXPERIMENTALES ENREGISTREES EN CONDITIONS CONTROLEES	174
TABLEAU 0.5 : ÉVOLUTION ATTENDUE DES CARACTERISTIQUES STANDARD DU SIGNAL CARDIAQUE	176
TABLEAU 0.6 – ÉVOLUTION ATTENDUE DES CARACTERISTIQUES ELECTRODERMALES STANDARD EN SITUATION DE STRESS.	178
TABLEAU 0.7 : ETALONNEMENT DE A EN FONCTION DU NOMBRE N D’EXEMPLES DISPONIBLES	187
TABLEAU 0.8 : TAILLES D’EFFET MESUREES POUR LES 32 CARACTERISTIQUES DE L’ENSEMBLE DE DEPART SUR LES TACHES TSST ET TA	188
TABLEAU 0.9 : EFFET TACHE POUR CHAQUE TACHE DU PROTOCOLE EXPERIMENTAL SUR LES CARACTERISTIQUES PRE-SELECTIONNEES	189

TABLEAU 0.10 : EFFET CONTRAINTE POUR CHAQUE TACHE DU PROTOCOLE EXPERIMENTAL SUR LES CARACTERISTIQUES PRE-SELECTIONNEES _____	189
TABLEAU 0.11 : EFFET PRESSION POUR CHAQUE TACHE DU PROTOCOLE EXPERIMENTAL SUR LES CARACTERISTIQUES PRE-SELECTIONNEES _____	189
TABLEAU 0.12 : EFFET ACTIVITE POUR CHAQUE TACHE DU PROTOCOLE EXPERIMENTAL SUR LES CARACTERISTIQUES PRE-SELECTIONNEES _____	190
TABLEAU 0.13 : EFFET TACHE POUR LES TACHES EXPERIMENTALES TSST ET TA SUR LES CARACTERISTIQUES STANDARD NON SELECTIONNEES _____	190
TABLEAU 0.14 : EFFET PRESSION POUR LES TACHES EXPERIMENTALES TSST ET TA SUR LES CARACTERISTIQUES STANDARD NON SELECTIONNEES _____	190
TABLEAU 0.15 : EFFET CONTRAINTE POUR LES TACHES EXPERIMENTALES TSST ET TA SUR LES CARACTERISTIQUES STANDARD NON SELECTIONNEES _____	191
TABLEAU 0.16 : EFFET ACTIVITE POUR LES TACHES EXPERIMENTALES TSST ET TA SUR LES CARACTERISTIQUES STANDARD NON SELECTIONNEES _____	191
TABLEAU 0.17 : CORRELATIONS (SPEARMAN) MESUREES ENTRE LES TAILLES D'EFFET DE LA FIGURE 0.13 ET CONTRIBUTION COMMUNE AUX ESTIMATEURS CORRESPONDANTS. _____	195

Index des Symboles & Acronymes

A, T, K, E, P	
Estimateurs du Chapitre Cinq.....	86
AED	
Activité électrodermale	37
BH3	
Ceinture thoracique Zephyr BioHarness 3	138
C, R, S	
Conditions Contrôle, Repos, Stress	73
D2	
Tâche d'attention visuelle D2.....	73
E4	
Bracelet de mesures Empatica E4	138
ECG	
Electrocardiographie / gramme	34
EVA	
Echelle Visuelle Analogique	26
f	
Caractéristique physiologique quelconque	82
HF	
Hautes fréquences du rythme cardiaque	42
HPA	
Voie de sécrétion du cortisol.....	21
HR	
Rythme cardiaque.....	41
hrMean, hrStd, hrRmssd	
Moyenne, écart-type, RMSSD du signal HR	77
HRV	
Variabilité cardiaque.....	41
IBI	
Intervalle inter-battements (cardiaques)	41
LF	
Basses fréquences du rythme cardiaque	42
LOSO	
<i>Leave-One-Subject-Out</i>	51
LOSOTO	
<i>Leave-One-Subject-and-One-Task-Out</i>	115
LOTO	
<i>Leave-One-Task-Out</i>	52
p	
p-valeur	89
PPG	
Photopléthysmographie / gramme	35
\mathcal{R}	
Régression logistique.....	117
Resp	
Signal respiratoire.....	38
respAutocorr	
Autocorrélation du signal respiratoire	77
RMS	
Racine du carré moyen	43
RMSSD	
RMS des différences successives (IBI, HR)	43
SAM	
Voie de sécrétion de l'adrénaline.....	21
SCL	
Niveau tonique en conductance cutanée	44
SCR	
Réponse électrodermale en conductance	44
SECPST	
<i>Socially Evaluated Cold Pressor Test</i>	28, 73
STAI	
State-Tait Anxiety Inventory	27
TA	
Tâche Arithmétique	73
TSST	
<i>Trier Social Stress Test</i>	28, 73
ϑ	
Taille d'effet	89
τ	
Instant de mesure d'une caractéristique	82, 105
\mathcal{N}	
Réseau de neurones	118
\mathcal{S}, \mathcal{M}	
Indice de Stress (Simple, Multiple).....	108
χ	
Indice de Carence	145

Chapitre Un - Introduction

I. Contexte et Motivations

En tant qu'expérience intime aux manifestations capricieuses, le phénomène qui nous intéresse dans cet ouvrage est difficile à observer. Pourtant, « stress » est un terme qui parle à tout le monde.

Après les douleurs et la fatigue, le stress est le quatrième symptôme spontanément reporté par les travailleurs européens (European Foundation for the Improvement of Living and Working Conditions, 2007). L'impact du stress lié au travail sur la santé mentale et physique des individus fait largement consensus, dans la littérature scientifique comme dans les institutions publiques (World Health Organization, 2003). Au début du XXI^e siècle, la Commission Européenne a évalué le coût de cet impact à un minimum de 20 milliards d'euros, parmi les 15 pays alors membres de l'UE (European Agency for Safety and Health at Work, 2014).

Les effets pathologiques du stress dit « chronique » résulteraient de l'accumulation d'épisodes courts et spontanés d'un stress dit « aigu », relatif aux événements de la vie courante. Le stress aigu est une réponse adaptative de l'organisme à une situation exigeante sur le plan physique, physiologique et/ou psychologique. Dans la majorité des situations stressantes, cette réponse est à notre avantage : son objectif est de mobiliser au mieux les ressources qui nous permettront de « faire face ». Ainsi, un état de stress bien géré peut constituer un bénéfice essentiel dans la pratique des métiers à risque, comme pour les pompiers, les pilotes, le personnel hospitalier, les athlètes de haut niveau – mais également dans les épreuves de la vie quotidienne.

Les bons comme les mauvais effets du stress sont profondément liés à ses manifestations physiologiques, que l'on peut mesurer de manière non-invasive. Des changements dans l'évolution du rythme cardiaque, dans la conductance de la peau, dans l'activité respiratoire ont été fréquemment observés en situation de stress aigu. Avec le développement de l'apprentissage automatique, depuis le début des années 2000, des mesures en laboratoire ont permis de produire des modèles capables d'identifier des épisodes de stress sur la base de ces corrélats. Avec l'essor des capteurs physiologiques portables, depuis le début des années 2010, on a désormais l'opportunité d'adapter ces modèles aux mesures de la vie réelle. L'utilisation de signaux physiologiques pour identifier des épisodes de stress dans la vie quotidienne ouvre la porte à une mesure objective de ce phénomène invisible, et à de nouvelles stratégies de gestion du stress à travers le « soi quantifié ».

La perspective de détecter le stress implique de nombreuses disciplines scientifiques – les sciences cognitives, les sciences expérimentales, la physiologie, le traitement du signal, l'apprentissage automatique – et pose des défis majeurs dans chacune d'entre elles. Un terme a été récemment proposé pour désigner la fusion des compétences issues de ces domaines dans l'objectif de mesurer ou reproduire un phénomène émotionnel : l'informatique affective (Picard, 1997). Sur ce nouveau terrain de recherche, les apports à l'état de l'art peuvent prendre de nombreuses formes : depuis la réalisation d'une campagne expérimentale jusqu'à la proposition d'un nouveau modèle, en passant par la proposition de corrélats pertinents ou d'une mesure de la qualité des données. Nos apports s'inscrivent dans ces quatre catégories.

II. Problématiques de Recherche

Notre travail de recherche a pour objectif de lever un certain nombre de verrous qui s'opposent à la détection du stress dans la vie quotidienne. Ceux-ci sont de plusieurs types.

- Les corrélats physiologiques du stress (ses « caractéristiques ») ne sont pas spécifiques au stress : ils réagissent à de nombreux facteurs dits confondants. En vie réelle, l'omniprésence de ces facteurs compromet fortement la prise de décision pour un détecteur qui n'a pas été entraîné à les prendre en compte.
- Les modèles de détection du stress développés dans la littérature ont souvent comme objectif premier la « performance », définie comme la capacité du détecteur à reconnaître des épisodes de stress et de non-stress. Cette recherche de performance a tendance à produire des détecteurs de type « boîte-noire », rendant l'interprétation de la décision prise inaccessible à l'utilisateur ou à un thérapeute.
- Les modèles de la littérature ne permettent pas de prendre en compte certaines particularités de la mesure en vie réelle, comme la mauvaise qualité des signaux enregistrés. Les signaux réels issus de capteurs portables engendrent fréquemment des problèmes de fiabilité des données (périodes de temps où les caractéristiques sont mal estimées), voire de disponibilité des données (périodes de temps où certaines caractéristiques ne peuvent être calculées).

Afin de répondre à ces problématiques, nous avons développé une méthode permettant de sélectionner les caractéristiques physiologiques les plus sensibles au stress tout en restant relativement peu sensibles aux facteurs confondants. En s'appuyant sur ces caractéristiques, nous avons proposé un détecteur de stress de type « boîte-blanche », c'est-à-dire permettant une interprétation de la décision prise. La particularité de ce modèle est d'être robuste aux valeurs extrêmes, mais aussi modulable, de sorte qu'il puisse s'adapter à l'indisponibilité de certaines caractéristiques. Afin de mettre en valeur cette capacité adaptative de notre détecteur de stress, nous avons proposé un estimateur en temps réel de la qualité du signal de rythme cardiaque. Cet estimateur peut être couplé avec notre détecteur de stress, afin de s'adapter en temps réel aux caractéristiques disponibles. L'ensemble de ce travail est présenté dans ce manuscrit suivant le plan détaillé ici.

III. Plan du Manuscrit

Avant de chercher à estimer le stress, nous avons besoin d'une définition claire de notre concept cible. Cette problématique est traitée dans le Chapitre Deux - , où l'on montre que le sens même du mot « stress » ne fait pas l'unanimité dans la littérature scientifique. Ces divergences justifient l'emploi d'une définition large (le stress est la réponse de l'organisme à une situation perçue comme pertinente et difficile par la personne), ce qui ne nous dispense pas de prendre en compte des facteurs confondants potentiels comme l'effort mental faible.

Le Chapitre Trois - propose une revue des méthodes concrètement mises en œuvre pour aboutir à une détection du stress dans la vie quotidienne. Une analyse critique des modèles trouvés dans la littérature nous permettra de définir notre positionnement. En plus d'une approche en *boîte-blanche*, deux contraintes seront retenues pour l'élaboration de notre détecteur de stress : (i) la prise en compte des facteurs confondants connus et (ii) la conservation des performances du détecteur dans un

contexte de mesures dégradées. Comme on l'observera dans ce chapitre, ces deux contraintes font l'objet d'un traitement inégal (et incomplet) dans la littérature actuelle.

En l'absence d'un modèle de détection standard, la conception d'un détecteur de stress passe par l'apprentissage sur des données expérimentales. Dans cette perspective, une campagne de mesures en laboratoire ayant précédé nos travaux de thèse sera décrite au Chapitre Quatre - . Les signaux physiologiques et les évaluations psychologiques recueillis durant cette expérience serviront de base aux analyses du Chapitre Cinq - et du Chapitre Six - .

A partir de ces données expérimentales, la conception d'un « bon » modèle commence par la sélection de « bonnes » caractéristiques parmi celles que la littérature a pu définir comme corrélats du stress. L'identification de caractéristiques sensibles au stress et peu sensibles à certains facteurs confondants est le sujet du Chapitre Cinq - . Dans cette démarche, nous constaterons que la majorité des caractéristiques couramment utilisées dans la littérature ne permettent pas de distinguer le stress d'une activité cognitive modérée.

Les caractéristiques sélectionnées pourront servir de base à l'élaboration d'un détecteur de stress dans le Chapitre Six - . Comme on l'a suggéré plus haut, un atout de notre modèle par rapport à ceux de la littérature est qu'il peut s'adapter à un ensemble variable de caractéristiques physiologiques. Pour réaliser cette fonction, notre détecteur adopte une structure modulaire qui lui permettra de se séparer temporairement des caractéristiques dont les mesures ne sont pas suffisamment fiables, ou sont temporairement indisponibles.

Mais le potentiel d'un tel modèle ne peut être exploité que si l'on dispose d'une information pertinente sur la fiabilité des caractéristiques utilisées. Dans cet objectif, nous proposons un algorithme qui estime la quantité d'information manquante sur une portion quelconque du signal de rythme cardiaque. Cet algorithme est développé dans le Chapitre Sept - , à partir d'une nouvelle base de données issues d'une campagne expérimentale en vie réelle et recueillie dans le cadre de nos travaux de thèse.

Pour terminer, le Chapitre Huit - présente nos conclusions principales et propose de nouvelles perspectives de recherche. Les travaux présentés dans ce manuscrit ont fait l'objet d'articles et de brevets qui sont listés en Annexe 13.

Chapitre Deux - Stress : Réponse Adaptative, Concept Versatile

Dans ce chapitre, on s'applique à définir notre concept cible : le stress émotionnel. La section I propose une définition intuitive du phénomène, basée sur le sens commun. La section II dessine l'identité physiologique du stress, au moyen d'un certain nombre de marqueurs qui font consensus dans la littérature scientifique. Mais le prérequis d'une bonne définition du stress est aussi de connaître ce qu'il n'est pas. Ainsi, la section III dépasse cette identité consensuelle pour définir un certain nombre de facteurs confondants. Pour finir, la section IV détaille les méthodes expérimentales disponibles pour constituer une base de données propice à l'étude du stress en laboratoire.

I. Définition Générique

Dans le dictionnaire en ligne de l'agence française Larousse (Larousse, s. d.), on trouve au mot « stress, n.m. » la définition suivante:

État réactionnel de l'organisme soumis à une agression brusque.

Il est précisé que ce terme provient de son homonyme anglais, qui signifie « effort intense ». Ici, terme « réactionnel » est proposé comme un attribut essentiel de « l'état » de stress. C'est par sa *fonction* qu'on cherche à expliquer *l'état*. Comme on le verra dans les sections suivantes, cette approche semble indispensable pour le définir clairement.

Dans la définition précédente, le système « organisme » (ici proposé comme objet du stress), ne peut pas être isolé : il est toujours en interaction avec son environnement. Cet environnement impose à l'organisme une condition nécessaire à l'état de stress : « l'agression brusque ». Nous (organismes évoluant fréquemment en état de stress) avons tous en mémoire des circonstances qui évoquent cette notion d'*agression*, que l'on peut généraliser à n'importe quelle situation perçue comme menaçante. La présence d'un jury attentif et exigeant illustre efficacement la condition *stressante* d'une situation d'examen. En particulier, le souvenir de sensations corporelles intenses (tétanie, nervosité ou fébrilité) associées à des sentiments négatifs ou neutres (impuissance, anxiété ou détermination) suffit à ancrer l'état *stressé* dans l'imaginaire collectif. Ainsi que les comportements (paralysie, esquive ou confrontation) qu'un tel état participe à engendrer.

Mais dans les mêmes circonstances, une personne préparée à l'épreuve (habituee au regard du jury et confiante dans ses capacités), pourra nous opposer un phénomène rival : le *bon stress* (parfois appelé *eustress*), qui nourrit notre volonté et hisse nos actions à la hauteur de nos attentes. Ce stress-là s'accompagne souvent d'un sentiment positif, loin de « l'agression brusque » dont on parlait un peu plus haut. Ainsi l'état de *stress* émerge d'expériences radicalement distinctes, liées à des perceptions différentes d'un même environnement. Qu'est-ce qui nous autorise, dans ce cas, à le concevoir comme un phénomène universel et récurrent ? C'est que sa fonction reste, chaque fois, identique. Le stress est une *réponse adaptative* aux manifestations multiples, dont l'enjeu est unique : augmenter nos capacités à faire face aux circonstances qui l'ont déclenchée.

La définition du Larousse capturerait cependant l'essentiel de notre concept cible. Un phénomène de *stress* est la somme d'un attribut exogène : l'environnement *stresseur*, et d'un attribut endogène :

l'organisme *stressé*. On explique le second attribut par ce qu'il augmente nos capacités à faire face au premier. Si cette définition fonctionnelle du stress a le mérite d'être univoque, elle ne suffira pas pour le modéliser. Il faut encore préciser d'une part, les conditions d'existence du stress et d'autre part, ses manifestations sur le corps humain. Comme on le verra dans ce chapitre, l'ensemble des facteurs du stress est à géométrie variable : celui-ci dépend du rapport personnel qu'un individu entretient avec son environnement.

II. Identité Physiologique du Stress

Dans ces conditions, existe-t-il un ensemble mesurable d'effets sur notre organisme qui suffise à caractériser l'état de stress ? Si oui, dans quelles conditions ? Ces deux questions doivent précéder toute recherche valide d'un modèle du stress humain. Dans cette perspective, la neurophysiologie propose des clés de compréhension incontournables sur l'état de stress en tant que réponse générique de l'organisme à une demande environnementale.

II.A. Approche Clinique : Deux Points de Vue Complémentaires

Un précurseur du concept de *stress* fut introduit par H. Selye (1936), qu'il nomma d'abord « syndrome général d'adaptation ». Selye observait alors, sur des rats exposés à des dommages divers tels que le froid, la chirurgie ou des agents chimiques, une réaction pathologique récurrente et non-spécifique au dommage infligé. La réaction, qui durait autant que le « stresseur » en question, conduisait en quelques semaines à la mort de l'animal. Ce travail d'analyse se limitait cependant aux stresseurs dits *physiques*, qui agressaient littéralement l'organisme touché. Chez l'être humain, ce sont les stresseurs dits *psychologiques* qui causent aujourd'hui la plupart des pathologies liées au stress (McEwen, 2000).

Ces stresseurs sont particulièrement étudiés en psychologie du travail, où l'on cherche à déterminer les facteurs responsables de ces pathologies. Certains d'entre eux sont recensés dans le modèle Demande-Contrôle de Karasek et Theorell (1990). La « Demande » intègre les besoins du travail en termes de charge mentale, d'éveil psychologique et d'interaction sociale. Le « Contrôle » reflète les compétences personnelles du travailleur et sa capacité à les développer. D'autres facteurs ont été proposés par des modèles concurrents, comme le Déséquilibre Effort-Récompense (Siegrist & Peter, 1994). En-dehors du lieu de travail, les événements majeurs de la vie personnelle (comme la perte d'un proche) sont également vecteurs des pathologies liées au stress (Dohrenwend, 2006).

Mais d'autres équipes ont aussi démontré l'intérêt d'étudier le stress à l'échelle individuelle, sur des fenêtres de temps plus courtes. Par exemple, Kanner et al. (1981) ont montré un lien entre l'apparition des troubles mentaux liés au stress et la répétition de « petits tracas » auxquels les personnes étaient exposées tous les jours. Les résultats de cette étude ont été étendus à l'impact général du stress sur la santé (DeLongis et al., 1982; Jandorf et al., 1986). Pour ces auteurs, ce ne sont pas de « grands événements » qui provoquent les maladies liées au stress, mais la multiplication des tracas liés à ces événements.

Dans l'étude clinique du stress, on peut ainsi distinguer deux points de vue complémentaires. Le premier s'intéresse au *stress chronique*, c'est-à-dire la forme durable et pathogène de cette réponse adaptative. Dans les études citées plus haut, une approche épidémiologique était employée pour dégager des *macro-stresseurs* en lien avec des pathologies répandues comme les maladies

coronariennes. Le deuxième point de vue s'intéresse au *stress aigu*, la forme éphémère déclenchée par un événement identifiable et contingent dans le temps. Dans cette approche, on s'intéresse aux *micro-stresseurs* et à leurs effets sur la personne et ses comportements.

II.B. Marqueurs Neurophysiologiques du Stress

D'après ce qui précède, le stress chronique (et les pathologies associées) serait provoqué par l'accumulation de stress aigu. Ainsi, les travaux cités montrent le potentiel préventif d'un détecteur de stress dans la vie quotidienne. Mais d'un point de vue physiologique, comment justifier qu'une succession de phénomènes transitoires aboutisse à l'apparition d'un état pathogène permanent, distribué dans l'organisme et non-spécifique à ces phénomènes ? Une réponse à cette question se trouve dans les réponses du système neuroendocrinien.

II.B.1. Activité nerveuse autonome

Introduisons d'abord quelques notions fondamentales. On utilise le terme anglais *arousal* pour décrire un niveau général d'éveil de la personne. Sur le plan psychologique, l'*arousal* se traduit par un état d'alerte et une motivation accrue ; il accompagne notamment les épisodes émotionnels intenses. Sur le plan physiologique, l'*arousal* s'accompagne (par exemple) d'une augmentation de l'activité cardiaque, de la pression sanguine et de l'activité sudomotrice. Dans la littérature, les notions de *stress* et d'*arousal* sont si intimement liées que Boucsein (2012) a défini le stress de la manière suivante :

Un état d'arousal généralisé et une émotion négative mais non spécifique, qui apparaissent comme conséquence de stressseurs agissant sur les individus.

L'*arousal* physiologique est commandé par des circuits cérébraux connus, et propagé dans le corps par le système nerveux périphérique. Pour mémoire : le *système nerveux* forme une arborescence fine et distribuée sur la totalité du corps humain. Cette arborescence peut être divisée en deux parties : le *système nerveux central* (cerveau, cervelet, tronc cérébral, moelle épinière) en constitue le « tronc » ; et le *système nerveux périphérique* en constitue les « branches ». À son tour, le système nerveux périphérique se subdivise en deux systèmes : le *système nerveux somatique* et le *système nerveux autonome*, qu'on distingue fonctionnellement par leur action sur l'organisme. Le système nerveux somatique coordonne le fonctionnement des muscles squelettiques et transmet l'information sensorielle (somesthésique, proprioceptive) au système nerveux central. De son côté, le système nerveux autonome régule les fonctions végétatives, comme l'activité cardiovasculaire, la respiration, la digestion. Contrairement au système nerveux somatique, l'absence (relative) de contrôle sur l'activité nerveuse autonome en fait un marqueur privilégié des états mentaux (Boucsein, 2012).

Les fibres du système nerveux autonome se divisent, à leur tour, en deux catégories fonctionnellement définies. On décrit généralement le système nerveux (ortho)*sympathique* comme favorisant l'action, et le système nerveux *parasympathique* comme favorisant le retour à un état d'équilibre ou de repos. Ces fonctions s'appliquent notamment sur le cœur, les vaisseaux sanguins, la peau, les poumons. Elle permet une commande binaire, de type activation/inhibition, sur certaines fonctions de ces organes telles que l'activité cardiaque, la vasodilatation, la transpiration, le cycle respiratoire. Ainsi, l'état de *repos* se caractérise par un certain équilibre entre l'activité nerveuse sympathique et parasympathique. À l'inverse, une augmentation de l'*arousal* se traduit par une activité prédominante de la branche sympathique sur la branche parasympathique. Cet état est

directement observable au travers du phénomène de « transpiration émotionnelle », à l'origine des mesures électrodermales détaillées en Chapitre Trois - I. Ce sont des fibres nerveuses sympathiques qui activent les glandes sudoripares eccrines, responsables de l'essentiel de l'activité sudomotrice. Or, ces fibres nerveuses sont largement stimulées en situation de stress.

II.B.2. Réponse hormonale

Lors d'un épisode de stress, l'activité nerveuse autonome s'accompagne d'une réponse spécifique du système endocrinien. Celui-ci régule la sécrétion des hormones, ces messagers chimiques qui circulent dans les vaisseaux sanguins. Parmi eux, le cortisol et l'adrénaline mobilisent les ressources de l'organisme et le préparent à l'effort.

La sécrétion de *cortisol* par la glande surrénale figure parmi les réponses les plus spécifiques au stress (Dedovic et al., 2009). Tout comme l'activité nerveuse autonome, la sécrétion de cette molécule est majoritairement régulée par l'hypothalamus (système nerveux central). Le cortisol agit sur le système immunitaire, sur la concentration sanguine en glucides et en lipides, sur la sensibilité de certaines cellules à l'adrénaline, et sur la synthèse de cette dernière (McEwen, 2000). À l'instar du cortisol, l'adrénaline (également sécrétée par la glande surrénale) permet de mobiliser les ressources énergétiques de l'organisme, comme la glycémie. Elle augmente également le rythme cardiaque, la contractilité du muscle cardiaque et la pression artérielle (McEwen, 2000), qui favorisent le démarrage et le maintien d'une activité physique ou mentale.

La sécrétion d'adrénaline se fait sur stimulation directe du système nerveux autonome. Cette voie permet une action plus rapide que la voie purement hormonale qui est celle du cortisol (Wetherell et

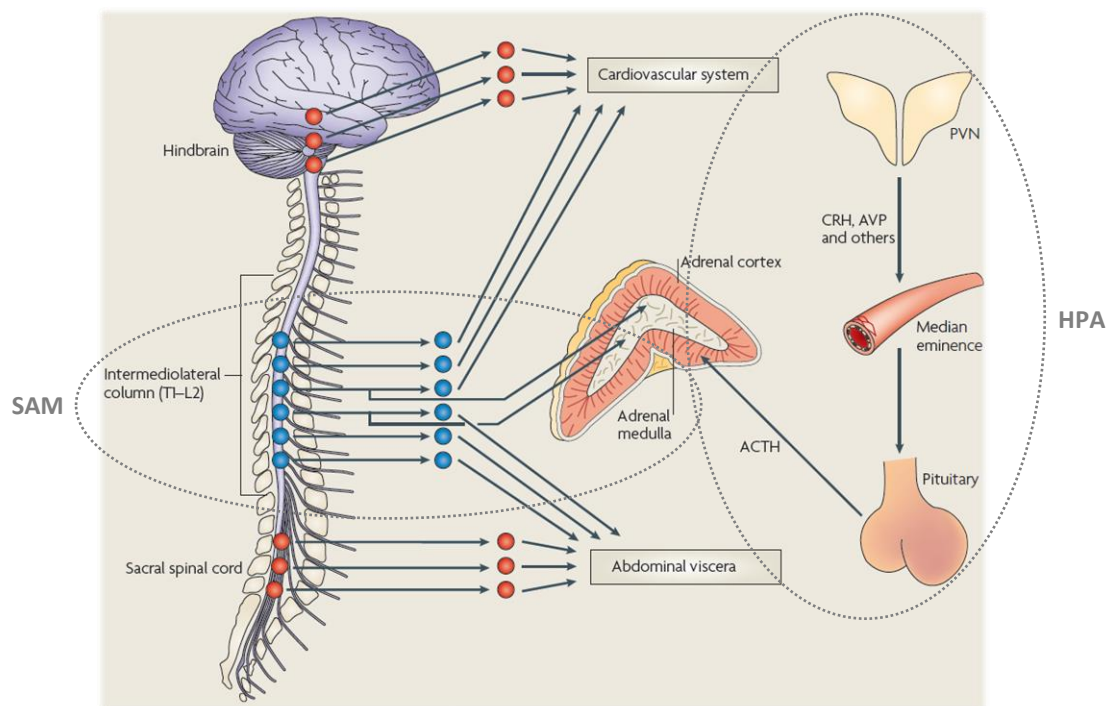


Figure 2.1 – Schéma de la réponse neuroendocrinienne au stress. Figure tirée de (Ulrich-Lai & Herman, 2009).

Les flèches associées aux puces bleues représentent le système nerveux sympathique. Les flèches associées aux puces rouges représentent le système parasympathique.

Avec ajout des ellipses légendées : SAM (axe sympatho-surrénalien-médullaire, à gauche), HPA (axe hypothalamo-hypophysaire-surrénalien, à droite).

PVN : noyau paraventriculaire (de l'hypothalamus) ; CRH : corticolibérine (intermédiaire de l'axe HPA) ; AVP : vasopressine ; ACTH : adrénocorticotrophine (intermédiaire de l'axe HPA).

al., 2006). Un épisode de stress aigu voit donc deux mécanismes s'activer sur des temporalités différentes. Le premier, dit « sympatho-surrénalien-médullaire » (dénomination standard : **SAM**, de l'anglais *sympathetic-adrenal-medullary*), augmente le taux d'adrénaline et permet une réaction rapide au stress. Le second, dit « hypothalamo-hypophysio-surrénalien » (dénomination standard : **HPA**, de l'anglais *hypothalamo-pituitary-adrenal*), augmente le taux de cortisol et provoque une réaction plus spécifique au danger. La Figure 2.1 est tirée de (Ulrich-Lai & Herman, 2009). Elle représente ces deux réponses (SAM à gauche, HPA à droite), ainsi que les actions directes du système nerveux sympathique (représenté par les puces bleues) et parasympathique (puces rouges) sur certains organes vitaux tels que le système cardiovasculaire et les viscères abdominaux.

II.C. Modèle Synthétique du Stress Physiologique

En résumé : la modification de l'activité cardiaque, de l'activité respiratoire, de la pression sanguine, de la vasomotricité, de la transpiration et même de la réponse immunitaire constituent un ensemble de symptômes de l'*arousal* physiologique lié au stress aigu. Cette réponse automatique permet à l'organisme d'anticiper et d'entretenir une réponse active à une demande de l'environnement extérieur (« *environmental challenge* », (McEwen, 2000)). De fait, ce phénomène d'adaptation est systématiquement lié au stimulus qui l'a fait naître, si bien qu'un certain nombre d'auteurs préfèrent appeler *stress* la situation ou l'événement responsable de cet état, plutôt que les changements physiologiques qui l'accompagnent (Dedovic et al., 2009; McEwen, 2000).

Un paradigme simple permet de penser ces changements physiologiques sur le court comme sur le long terme. On peut considérer que l'ensemble des paramètres vitaux sont régulés par deux mécanismes complémentaires : l'homéostasie et l'allostasie. L'*homéostasie* désigne la tendance du métabolisme à maintenir constants (ou quasi-constants) un certain nombre de paramètres, car leur stabilité est nécessaire à la survie de l'individu (Cannon, 1929). Parmi ces paramètres se trouvent (par exemple) la température corporelle et la pression partielle en O₂ dans le sang. On peut dès lors appeler *allostasie* l'ensemble des systèmes du métabolisme qui se modifient sur une plus large amplitude, afin de permettre l'homéostasie (McEwen, 2000). Parmi ceux-ci, on peut citer : l'activité cardiaque et la vasomotricité, pour préserver la pression en O₂, ou encore la transpiration, pour maintenir la température corporelle. Dans la suite de ce manuscrit, le terme *allostasie* (ou bien *réponse allostatique*) désignera l'ensemble des réponses physiologiques qui caractérisent un épisode de stress.

La *charge allostatique* persiste un certain temps après la disparition du stress. Après un court épisode de stress mental, une période variable est nécessaire pour un retour à la « normale » des paramètres biologiques : quelques minutes seulement pour le rythme cardiaque, mais plus d'une heure pour la concentration (salivaire) en cortisol (Rimmele et al., 2009). Par cette persistance ou par l'anticipation de menaces futures, la répétition de nos stressors quotidiens installe ainsi sur la durée la présence active d'hormones du stress dans notre organisme (Smyth et al., 1998).

Lorsque cette présence perdure (McEwen (2005) parle de *surcharge allostatique*), ses effets sur l'organisme deviennent néfastes. Le cortisol et l'adrénaline ont des effets destructeurs à long terme sur leurs organes cibles. Une augmentation trop fréquente de la pression sanguine, par exemple, provoque de l'artériosclérose ; la réponse immunitaire, d'abord stimulée, s'effondre ; l'atrophie de certains neurones impacte la mémoire (McEwen, 2005). Dans le système cardiovasculaire, un déséquilibre persistant entre l'activité nerveuse sympathique et l'activité parasympathique joue un rôle important dans l'apparition des maladies coronariennes ou d'arythmies létales (Sztajzel, 2004).

Ainsi lorsque le stress aigu devient chronique, la réponse adaptative devient « syndrome d'adaptation » et l'organisme s'abîme peu à peu dans l'épuisement.

III. Définitions Croisées et Facteurs Confondants

Sur le court terme, les mécanismes biologiques décrits plus hauts sont systématiques, mesurables et analogues chez tous les individus. Ils confèrent à l'état de stress une identité physiologique indépendante du stresser ; et permettent sa détection à partir de mesures physiologiques standard. Mais ces marqueurs *corrélatifs* du stress ne lui sont pas *spécifiques* : ils accompagnent des phénomènes bien différents. Les sections qui suivent cherchent à redéfinir notre concept-cible à partir de ses *facteurs confondants*, ces phénomènes non-stressants qui agissent sur les corrélats physiologiques du stress.

III.A. Facteurs Physiologiques

L'exercice physique, les conditions extérieures ou encore l'heure du jour, déclenchent les marqueurs physiologiques identifiés en section II. Ces réplicats bien connus de l'allostasie liée au stress nous permettront de donner une définition physiologique à la notion de « stress mental ».

III.A.1. Il existe des allostasies non-stressantes

Souvent considérée comme le marqueur le plus spécifique du stress, la concentration sanguine en cortisol est pourtant largement impliquée dans la préparation du corps à l'éveil (Buijs et al., 2003). La molécule est massivement sécrétée durant les dernières phases du sommeil, pour atteindre un pic de présence dans l'organisme peu après le réveil, avant de retomber progressivement au fil de la journée. La concentration en cortisol se modifie ainsi naturellement sur une très large amplitude, suivant un rythme circadien. Ce mécanisme permet un contrôle du rythme cardiaque et de la température corporelle de nature à prévenir les demandes physiologiques liées à l'activité diurne.

De même, la température extérieure est un déterminant bien connu de l'activité sympathique à la surface de la peau. Sous l'effet de la chaleur, les fibres sympathiques activent les glandes sudoripares eccrines. Sous l'effet du froid, elles provoquent la vasoconstriction des vaisseaux sanguins. Ainsi, les mesures physiologiques focalisées sur ces phénomènes cutanés (*e.g.* : activité électrodermale, photopléthysmographie) seront affectées par des variations de température en l'absence de stress. Un détecteur de stress seulement basé sur le débit sudatoire, ou sur la concentration sanguine en cortisol, pourrait ainsi *confondre* un état de stress avec une variation circadienne ou liée à la température extérieure. Par chance, le « bruit » généré par ces facteurs confondants est à variations lentes ; et touche une partie seulement des manifestations physiologiques du stress.

L'exercice physique, en revanche, déclenche une réponse forte dans chacun des corrélats connus du stress. Le cortisol et l'adrénaline sont tous deux sécrétés pour mobiliser les ressources énergétiques de l'organisme. L'activité nerveuse autonome est bouleversée pour adapter le flux sanguin à la consommation métabolique en dioxygène (Taylor et al., 1999). Après l'exercice physique, les paramètres biologiques ainsi modifiés ne reviennent que progressivement à la normale (Martins-Pinge, 2011). Du point de vue développé en section II, l'exercice physique agit ainsi comme un *stresseur* intense et durable sur l'organisme en activité. De la même manière que le stress, une séquence

d'exercices excessive peut même engendrer une altération temporaire et pathologique des fonctions biologiques : le syndrome de surentraînement (Hackney, 2006). Lorsqu'il n'est pas pratiqué en conditions extrêmes, cela dit, l'exercice physique n'engendre pas d'effets pathologiques sur le long terme (Carroll et al., 2009).

III.A.2. Le stress engendre une réponse excessive aux besoins de l'organisme

Bien qu'ils aient des effets similaires sur le système neuroendocrinien, ni l'exercice physique, ni les rythmes circadiens, ni la température ambiante ne sont des *stresseurs* au sens proposé en section I. L'existence de ces facteurs condants nous impose de préciser « l'identité physiologique » du stress décrite en section II. D'un point de vue physiologique, par exemple, on pourrait commencer par répondre à la question suivante : comment se fait-il que la répétition de bouleversements identiques au stress aigu n'aboutisse jamais à un état de stress chronique ?

Plusieurs auteurs (Carroll et al., 2009; Sherwood et al., 1986; Turner & Carroll, 1985) ont proposé une différence fondamentale entre la charge allostatique liée à l'exercice physique et celle liée au stress. A niveau égal de consommation en dioxygène, l'*amplitude* de la réponse cardiovasculaire est bien supérieure en situation d'effort mental qu'en situation d'effort physique. Une démonstration de ce phénomène est proposée en Annexe 1. D'autre part, la *persistance* de la réponse hormonale est bien plus importante après un effort mental qu'après un effort physique (Brosschot, 2010). A la différence de l'activité physique, cette amplitude et cette persistance ne rencontrent pas un besoin réel de l'organisme en situation d'effort mental. Lorsqu'elle est trop fréquente, cette « exagération » de la charge allostatique serait à l'origine des pathologies cardiovasculaires liées au stress.

Ainsi, la réponse physiologique décrite en section II.B peut répondre à deux sollicitations bien distinctes. Dans le premier cas, cette réponse est issue d'un réflexe physiologique et proportionnée aux besoins du corps. Dans le second cas, elle est déclenchée par un processus cognitif et surévaluée par rapport aux besoins du corps. Ce dernier type de réponse caractériserait l'état de *stress mental*.

III.B. Facteurs Psychologiques

Au regard de la section I, cette conception purement somatique du stress ne suffit pas à produire une définition pertinente. En particulier, le vaste concept de *stress mental* ne permet pas d'expliquer ce qu'est un *stresseur*, ou bien une activité *stressante*. Par exemple, peut-on définir un type d'activité mentale qui soit non-stressant ? Ou encore, peut-on distinguer un « bon » d'un « mauvais » stress sur le plan physiologique ? Ces questions suggèrent de repenser le stress d'un point de vue psychologique ; et de compléter éventuellement la liste des facteurs confondants. Pour y répondre, les paragraphes qui suivent réalisent la synthèse de plusieurs modèles qui sont détaillés en Annexe 2. Parmi ces modèles, la nature stressante de l'effort mental et les sentiments générés par un stresseur cristallisent des discordances fortes sur la notion de *stress*.

III.B.1. Il existe un type d'effort mental non-stressant

Dans la section précédente, on distinguait implicitement les notions d'effort *physique* et d'effort *mental* en associant ce dernier à une situation de stress. Cette analogie est-elle toujours valide ?

En psychologie, il n'existe pas de définition unique pour le concept *d'effort mental* car les recherches sur le sujet recoupent des domaines d'étude bien distincts (stress, mémoire de travail, sciences éducatives...). Parmi ces recherches, cependant, un paradigme solide a pu s'établir autour de l'existence de deux niveaux d'effort (Fairclough & Mulder, 2011; Hockey, 1997).

1. L'individu est en situation *d'effort faible* lorsque les ressources nécessaires pour faire face à la situation sont immédiatement disponibles.
2. L'individu est en situation *d'effort compensatoire* lorsque la situation nécessite de recruter des ressources cognitives qui ne sont pas immédiatement disponibles.

Les auteurs ayant directement comparé les notions de « *stress* » et « *d'effort/charge* » mental(e) affirment que *l'effort faible* ne s'accompagne pas de stress (Gaillard & Wientjes, 1994; Hockey, 1997). En revanche, celui-ci enclenche une réponse neuroendocrinienne similaire à celle du stress (Fairclough & Mulder, 2011; Hockey, 1997).

La notion *d'effort faible* ne correspond ni à la définition fonctionnelle du stress proposée en section I (car il n'augmente pas nos capacités à faire face à un stresser), ni à la définition clinique que l'on a vue en section II.A.II (car il ne serait pas porteur de stress chronique). Tous ces arguments appuient l'idée qu'un effort faible est *non-stressant*. En revanche, l'existence d'un *effort compensatoire* non-stressant dépend de la définition que le chercheur attribue à la notion de stress. Cela revient à questionner l'existence d'un « bon » et d'un « mauvais » stress (voir aussi section I).

III.B.2. La réponse physiologique au stress pourrait dépendre de l'état émotionnel

Pour certains auteurs, en effet, le *stress* se résume à ses expériences négatives (Boucein, 2012) ; au décalage entre les ressources disponibles et celles nécessaires pour faire face à la situation (Hockey, 1997) ; ou à une réponse physiologique inefficace pour faire face au stresser (Gaillard & Wientjes, 1994). Ces auteurs associent implicitement le « bon stress » à un effort compensatoire non-stressant.

De leur côté, Lazarus et Folkman (1985) ont proposé la définition suivante :

*Une relation entre une personne et son environnement,
lorsque celui-ci est évalué par la personne comme pertinent pour son bien-être –
et lorsque les ressources de la personne sont sollicitées ou dépassées.*

Cette conception du stress mental a le mérite d'exprimer ce qu'est un *stresseur*, à savoir toute situation perçue comme *pertinente* et *difficile* par la personne. Ce cadre théorique est relativement large : il inclut à la fois les notions de « bon » stress, de « mauvais » stress et d'effort compensatoire.

Pour trancher la question, un certain nombre d'auteurs a étudié la possibilité de distinguer plusieurs *types* de stress mental sur le plan physiologique (Frankenhaeuser, 1982; Tomaka et al., 1997). Un consensus solide dans ces modèles se trouve dans l'absence de sécrétion de cortisol en cas d'effort perçu comme non-stressant (Frankenhaeuser, 1982; Hockey, 1997). Ainsi, l'activation de l'axe HPA suffirait à définir la présence de stress dans l'effort mental. Mais comme on le verra en section IV.A, cette propriété est difficile à vérifier empiriquement.

En ce qui concerne les mesures non-invasives considérées dans ce manuscrit, les partisans d'un stress différencié sont seulement parvenus à montrer une différence *relative* dans les paramètres physiologiques (ex. : une augmentation *plus ou moins* importante du rythme cardiaque). Ces résultats

n'ont pas permis d'identifier une réponse spécifique à chaque type de stress. Autrement dit : la conception du stress comme une réponse adaptative polymorphe manque encore de support empirique pour s'insérer dans un modèle physiologique du stress émotionnel.

Cette conclusion nous a conduit à abandonner les notions de « bon » et de « mauvais » stress, au profit de l'approche inclusive proposée par Lazarus et Folkman (1985). Dans cette approche, la perception d'un même *stresseur* peut évoluer au cours du temps : celui-ci peut générer un sentiment négatif, positif ou neutre, sans remettre en cause le caractère *stressé* de l'individu. Comme ce sentiment peut aussi moduler l'activité nerveuse autonome (Ekman et al., 1983; Schachter & Singer, 1962), la « couleur émotionnelle » du stress doit être considérée comme un effet du second ordre sur les corrélats physiologiques qui nous intéressent.

III.C. Synthèse : l'Etat de Stress et ses Facteurs Confondants

La notion de *stress* est recouverte par une multitude de concepts : charge allostatique, effort mental, détresse, défi, menace, préjudice, réaction somatique inefficace... Parmi ces concepts, il apparaît un dénominateur commun qui est cohérent avec la définition proposée en section I : une situation de stress implique de recruter des « ressources » (cognitives, énergétiques) supplémentaires pour faire face. Cet impératif semble déclencher une réponse physiologique excessive aux besoins de l'organisme, probablement à l'origine de l'apparition du stress chronique.

En écho à la proposition de Lazarus et Folkman (1985), nous définissons *l'état de stress* comme le faisceau des réponses physiologiques qui accompagnent le vécu d'une situation perçue comme *difficile* par la personne et *pertinente* pour son bien-être. Par exemple, une prise de parole en situation d'examen. Dans cette configuration, on peut nommer *charge affective* la réponse physiologique qui ne répond pas à un réflexe de l'organisme pour l'activité considérée. Les corrélats du stress sont les paramètres (ou *caractéristiques*) mesurables qui présentent une telle charge affective. Par exemple, une augmentation du rythme cardiaque que l'on peut lier aux enjeux de la prise de parole.

Mais ces corrélats ne sont pas spécifiques : ils répondent à d'autres phénomènes que l'on peut qualifier de non-stressants. On appelle *facteurs confondants* ces phénomènes qui font que l'absence ou la présence d'une réponse physiologique ne signifie pas l'absence ou la présence de stress. Les sections précédentes nous ont permis de définir trois types de facteurs confondants.

(i) Les corrélats du stress peuvent répondre à un besoin réel du métabolisme. Par exemple, la mobilisation des ressources énergétiques et la régulation thermique nécessaires à une marche lente dans une pièce non climatisée. Lors d'un effort physique, la réponse de ces corrélats est intense et (pour l'heure) indiscernable de celle du stress. Hors d'un effort physique, les besoins métaboliques s'expriment par des variations lentes et *spontanées* dans les paramètres biologiques : la *ligne de base*.

(ii) Les corrélats du stress peuvent montrer une réponse réflexe de l'organisme au traitement cognitif d'une tâche quelconque. Par exemple, une augmentation du rythme cardiaque que l'on peut lier aux processus psychomoteurs de la parole. Dans la suite de ce manuscrit, on appellera *charge mentale* cette réponse qui correspond à la notion d'effort mental « faible » développée section III.B.

(iii) La charge affective peut être modulée soit par le type, soit par l'intensité du stress ressenti par la personne. Par exemple, une prise de parole peut être perçue successivement comme un défi et une menace. Les connaissances actuelles ne permettent pas de distinguer ces variations d'un second ordre dans l'état de stress à partir de mesures non-invasives.

IV. Méthodes Expérimentales pour l'Étude du Stress

Bien que le concept de stress manque d'une définition standard, la plupart des études expérimentales ont recouru à des méthodes de référence pour induire et mesurer du stress chez leurs participants. Comme on pourra le constater dans le Chapitre Trois - , la conception du protocole expérimental détermine fortement les résultats obtenus dans l'estimation du stress. En effet, les méthodes expérimentales définissent *implicitement* le concept de « stress » que l'on pourra observer dans les mesures. Les paragraphes qui suivent nous permettront de comprendre pourquoi.

IV.A. Évaluer le Stress

Afin de valider le caractère *stressant* d'une condition expérimentale, on a besoin d'une *vérité de terrain* qui confirme une variation du niveau de stress chez les participants. Les paragraphes qui suivent développent les mesures de stress les plus courantes dans la littérature.

IV.A.1. Concentrations hormonales

En particulier, la réponse hormonale (notamment celle en cortisol) peut être mesurée de manière non-invasive par prélèvement salivaire. Un certain nombre d'études a directement utilisé les concentrations salivaires ou sanguines de ces hormones (notamment en cortisol) comme mesure de référence pour le niveau de stress. Des méta-analyses ont confirmé le déclenchement de cette réponse hormonale en conditions standardisées, les principaux facteurs de réponse étant la pression sociale et le manque de contrôle sur l'environnement (Dickerson & Kemeny, 2004). Cependant, il semble que les résultats divergent entre les études sur la corrélation entre la concentration en cortisol et le subjectif de stress (Hjortskov, Garde, et al., 2004).

En effet, cette mesure est aussi perturbée par divers facteurs tels que le sexe de l'individu, les cycles menstruels et la contraception orale (Hellhammer et al., 2009), les rythmes circadiens (voir III.A.III.C). Cette réponse souffre également d'un fort temps de latence : un pic de cortisol salivaire ne peut être mesuré que 20 à 30 minutes après l'apparition d'un stressor (Smyth et al., 1998). Ces inconvénients rendent difficile l'utilisation de cette mesure pour l'évaluation du stress en conditions réalistes.

IV.A.2. Évaluation subjective directe : les échelles visuelles

Paradoxalement, le sentiment *subjectif* de stress pourrait constituer une *vérité de terrain* plus fiable qu'une mesure hormonale dite objective. Du point de vue du sujet, la plus simple de ces mesures consiste à donner une « note » à son propre stress. On demande au sujet de placer un marqueur librement sur une échelle fixe, dont les deux extrémités représentent ses niveaux minimum et maximum de stress (voir Figure 2.2). Cette échelle peut être continue (a) ou discrète (b). Ce type d'échelle couramment utilisée pour quantifier le sentiment subjectif de stress, est dite « visuelle analogique » (**EVA**) (Skoluda et al., 2015). Cette mesure a été corrélée avec les scores issus de questionnaires standards (voir section ci-après) et s'est montrée susceptible de distinguer différents niveaux de stress, même à l'échelle de plusieurs individus (Lesage et al., 2012).

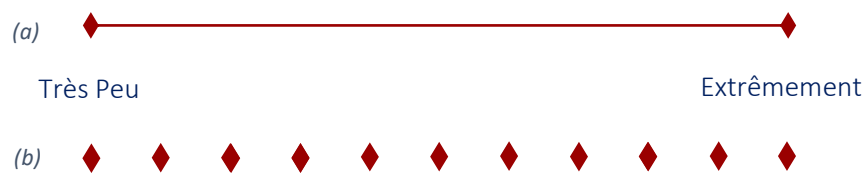


Figure 2.2 : Schéma d'une échelle visuelle analogique à valeurs continues (a) et à valeurs discrètes (b)

De manière générale, une EVA peut être utilisée pour quantifier n'importe quel sentiment subjectif, comme celui de douleur ou d'effort (Bullinger et al., 1984), ou encore des émotions telles que la joie, la peur, la colère (Ekman, P. et al., 1983). Elle est également utilisée dans les mesures *dimensionnelles* de l'émotion (Russell & Mehrabian, 1977).

IV.A.3. Évaluation subjective indirecte : les questionnaires standardisés

Des questionnaires ont également été élaborés dans une méthodologie standard, afin de permettre une mesure robuste du stress psychologique. Plutôt qu'une échelle visuelle, le sujet considère un certain nombre d'*items* lexicaux : des phrases ou mots à caractériser, souvent à connotation négative ou positive et liés à une humeur générale ou un type d'événement particulier. À chaque item doit correspondre un score, qui est un indicateur de fréquence (Parfois/Souvent/...), de quantité (Peu/Beaucoup/...) ou encore d'adhésion (D'accord / Pas d'accord).

Un exemple de ces questionnaires est spécifiquement dédié à l'anxiété : il s'agit du *State-Trait Anxiety Inventory (STAI)*, dont la deuxième version (STAI-Y : (Spielberger, 1983)) est fréquemment utilisée pour mesurer le stress à valence négative. Ce questionnaire quantifie l'anxiété en tant qu'état d'humeur (forme STAI-Y-A, aussi appelée STAI-S) et en tant que trait de personnalité (forme STAI-Y-B, ou STAI-T). Les scores des items positifs (faible anxiété) sont inversés, puis sommés aux items négatifs pour générer un indicateur global. Néanmoins, la taille d'un tel questionnaire (2x20 items) le rend inadéquat pour la mesure *instantanée* du sentiment d'anxiété : sa résolution temporelle est limitée. Ainsi, un certain nombre d'études utilisent le STAI-S en version abrégée (Marteau & Bekker, 1992).

Un questionnaire semblable se nomme *Positive And Negative Affect Schedule (PANAS)*, dont les 10 items positifs quantifient l'état d'humeur positive, et les 10 items négatifs l'état d'humeur négative (Watson et al., 1988). Ce questionnaire présente l'avantage de rester fiable en rétrospective, depuis l'instant de mesure jusqu'à deux mois auparavant : son échelle de temps dépendra seulement de la consigne donnée au sujet. Cependant, il n'est pas spécifique au stress. Un autre indicateur plus fréquemment appliqué au stress, est donné par le questionnaire *Perceived Stress Scale* (Cohen et al., 1983). Celui-ci demande au sujet de noter 14 événements sur une échelle de fréquence, au cours d'une période donnée. À l'origine élaborés pour l'étude du stress sur le long terme, ce questionnaire reste cependant peu adapté à la mesure du stress aigu. Un dernier questionnaire fréquemment utilisé en sciences affectives est le *NASA-TLX* (Hart & Staveland, 1988), qui mesure différentes composantes de la charge mentale subjective. Le stress est l'une de ces composantes.

IV.B. Induire le Stress

Au travers d'une synthèse des tâches standard, les paragraphes qui suivent tentent de répondre à la question suivante : comment introduire un stress réaliste dans une expérience de laboratoire ?

IV.B.1. Approche historique

En sciences affectives, on a souvent exploité les sens de la personne (vision, ouïe, douleur) pour générer un sentiment, une émotion ou une humeur en particulier. Dans le cas du stress, on peut utiliser des scènes de films (Grundlehner et al., 2009), des images à contenu émotionnel riche (van Stegeren et al., 2008), des sons désagréables (Mandruck et al., 2016) ; ou l'anticipation d'une douleur physique, autrefois induite par des chocs électriques (Averill & Rosenn, 1972).

Ici, l'expérimentateur contrôle l'apparition, la durée et l'intensité de chaque stimulus qui fait office de stressor. On a ainsi accès à des caractéristiques fines sur les enregistrements physiologiques, comme le temps de réaction ou de récupération du signal après chaque micro-événement. Mais une telle finesse d'analyse ne peut être obtenue qu'au détriment du réalisme des tâches proposées. Un choc électrique inévitable ou le visionnage forcé d'images désagréables ne correspondent pas à une situation de la vie quotidienne. En particulier, l'individu n'a aucun moyen d'action sur le stressor (ici l'entrée sensorielle) qui est imposé par le protocole expérimental, ce qui le prive d'un large spectre de réponses affectives face au stressor.

IV.B.2. Tâches standard

a) Stress physique

Dans les protocoles modernes, on retrouve le stress d'origine physique dans une tâche de résistance au froid : le *Cold Pressor Test* (CPT), qui demande au participant de plonger sa main plusieurs minutes dans une eau à 4°C. Cette tâche a été liée à une activation de l'axe endocrinien HPA et du système nerveux autonome (Skoluda et al., 2015). Un sentiment de douleur s'accompagne d'une sécrétion de cortisol, mais pas nécessairement d'endorphines (Bullinger et al., 1984). On a ainsi suggéré que le stress physiologique observé serait provoqué par l'évaluation cognitive d'une douleur (Bullinger et al., 1984), ou par l'anticipation d'une blessure par l'eau froide (Dedovic et al., 2009).

b) Stress psychosocial et pression temporelle

Il existe une variante du CPT dans laquelle le sentiment de douleur s'accompagne d'une évaluation sociale : **SECPT**, pour *Socially Evaluated Cold Pressor Test* (Schwabe et al., 2008). Cette variante consiste à informer le participant que sa réaction sera filmée ; et que ses performances seront évaluées relativement aux autres participants. On introduit ainsi un objectif dont la réussite dépend de la réaction du participant face à l'eau froide. En élargissant le spectre des réponses affectives possibles, l'évaluation sociale installe les conditions d'une réponse réaliste au stressor, tout en augmentant le sentiment subjectif de stress.

Le facteur social est ainsi présent dans la majorité des études actuelles qui cherchent à reproduire des conditions « réalistes » de stress. Parmi les tâches les plus utilisées dans la littérature, le *Trier Social Stress Test*, (**TSST**) et la tâche arithmétique sont fréquemment enchaînées au cours d'une même session expérimentale. Le TSST simule un entretien d'embauche de courte durée (quelques minutes), avec un temps de préparation tout aussi court. De son côté, la tâche arithmétique consiste à répéter un calcul le plus vite et le plus justement possible (exemple : soustraire de manière sérielle, le nombre de 13 de 1022). Ces deux tâches se déroulent sous pression temporelle et sociale. Dans la littérature, l'enchaînement de l'une et de l'autre a montré d'excellentes performances en termes d'activation des systèmes endocrinien et autonome liés au stress (Kirschbaum et al., 1993).

c) Attention et charge mentale

Pour finir, il n'est pas rare que des tâches initialement prévues pour l'étude de la mémoire de travail ou de l'attention soient adaptées aux protocoles d'étude sur le stress. Le test de Stroop (1935) est l'une de ces tâches. Il consiste à présenter des noms de couleurs, typographiés chacun dans une couleur qui peut être soit congruente (exemple : le mot *ROUGE* écrit en rouge), soit incongruente (*ROUGE* écrit en bleu). Le sujet doit lire le mot, ou bien donner sa couleur : il s'agit d'un test d'attention sélective. Une élévation des indicateurs psychologiques et physiologiques du stress a déjà été observée dans la condition incongruente par rapport au repos (Castaldo et al., 2015; Tulen et al., 1989). De même, la tâche arithmétique fréquemment utilisée pour l'étude du stress est également utilisée pour l'étude de la charge cognitive (Paas et al., 1994; Tomaka et al., 1993).

IV.C. Expression des Facteurs Confondants dans une Base de Données Expérimentales

Des tâches expérimentales efficaces permettent d'enregistrer des variations de stress sur un ensemble de sujets dans un milieu contrôlé. Dans le Chapitre Trois - , on verra comment les données récoltées permettent d'élaborer un modèle physiologique du stress mental. Mais la section III.C nous a aussi permis d'identifier des facteurs confondants théoriques qui doivent être pris en compte lorsqu'on veut attribuer une variation physiologique au stress. Dans les paragraphes qui suivent, on a donc tenté de répondre à la question suivante : comment ces facteurs se traduisent-ils concrètement dans une base de données expérimentale ?

IV.C.1. Facteurs confondants physiques

L'exercice physique, les conditions extérieures (ex : température) et le contexte somatique (ex : rythmes circadiens) font partie de ces facteurs confondants du stress dont l'existence fait largement consensus dans la littérature. Pour contrôler leur influence sur les mesures physiologiques, un protocole expérimental comprend généralement deux contraintes. (i) Outre leur caractère *stressant*, les tâches précédentes ont en commun l'absence d'effort physique. (ii) Avant l'apparition du stress, une tâche de *repos* permet de mesurer la ligne de base pour chaque signal enregistré.

Présente dans la plupart des études en sciences affectives, la tâche de repos peut durer entre 3 et 30 minutes selon le type de mesure considéré. Durant cette période, l'activité du participant et son interaction avec l'environnement doivent demeurer minimales : on enregistre ainsi le niveau *spontané* des paramètres physiologiques. Lors de l'analyse des données, ces enregistrements permettent de soustraire la ligne de base aux signaux mesurés pendant les tâches stressantes.

IV.C.2. L'effet de l'activité

Si l'on pousse plus loin le raisonnement précédent, on observera que le *stress* n'est pas l'unique facteur divergent entre la tâche de repos et une tâche expérimentale stressante.

La *parole*, par exemple, est présente dans la plupart des tâches présentées dans le paragraphe IV.B.2. Or, des études s'accordent pour montrer l'effet activateur de la parole sur l'activité cardiaque, entre une tâche arithmétique pratiquée en silence et la même tâche pratiquée à voix haute (Sloan et al., 1991; Tomaka et al., 1994). En elle-même, la parole n'est pas un *stresseur* ; mais elle a recours à

des processus cognitifs et moteurs qui contribuent à l'activité physiologique mesurée lors de la tâche stressante. Considérée seule, elle correspond à une situation *d'effort faible* : à la charge mentale définie en section III.C s'ajoutent les besoins métaboliques liés à la tâche en cours.

Dans la suite de ce manuscrit, on appelle *effet de l'activité* cette réponse réflexe de l'organisme au traitement cognitif et moteur de la tâche en cours. Si l'on adopte la définition du stress proposée en section III.C, cet effet de l'activité doit être considéré comme un facteur confondant. Une recherche pertinente des corrélats du stress doit permettre de le distinguer de la charge affective.

Comme la ligne de base, on peut mesurer cet effet au moyen d'une tâche expérimentale non-stressante, dite tâche *contrôle*. Celle-ci est assimilable à la tâche stressante par l'activité mentale et motrice requise, mais sans le stresser associé. Cette tâche contrôle permet ainsi d'observer les facteurs confondants liés à l'activité, qui n'étaient pas présents dans la tâche de repos. Un exemple de tâche contrôle a été introduit par Dedovic et collaborateurs (2005) dans leur tâche arithmétique. Leur protocole inclut trois conditions. Dans la condition Repos, le sujet regarde un point fixe. Dans la condition Contrôle, le sujet effectue une série de calculs simples. Dans la condition Stress, les calculs atteignent progressivement les capacités limites du sujet. Dans la même démarche que pour la tâche arithmétique, une tâche contrôle adaptée au SECPT consiste à plonger sa main dans une eau tiède (35 à 37°C), sans pression sociale (Schwabe et al., 2008).

IV.C.3. L'effet de la tâche

Le dernier facteur d'erreur relevé dans la section III.C se trouve dans la modulation de la charge affective, par le type ou par l'intensité du stresser. Pour illustrer comment cet effet se manifeste dans une base de données, on peut étudier un exemple : l'expérience de Skoluda et collaborateurs (2015). Ces auteurs ont comparé l'efficacité du TSST, du CPT (sans évaluation sociale) et du test de Stroop pour induire du stress chez leurs participants. Les trois tâches ont montré une augmentation significative du sentiment subjectif de stress (EVA) par rapport à une tâche de repos. En revanche, une augmentation du cortisol salivaire a été observée seulement pour le TSST. Une augmentation du rythme cardiaque a été constatée pour TSST et le test de Stroop, mais pas pour CPT. Ainsi, trois conditions expérimentales différentes ont montré des effets différents sur les corrélats du stress.

Une critique que l'on pourrait formuler sur cette étude est l'absence de contrôle sur l'effet d'activité. Si l'on reprend l'exemple précédent, la simple présence de *parole* dans le test de Stroop et le TSST pourrait justifier la réponse cardiaque mesurée dans ces deux tâches ; et non dans le CPT. Cependant, il est à noter que l'augmentation mesurée pour le rythme cardiaque apparaît plus importante durant le TSST (qui simule un entretien d'embauche) que durant le test de Stroop (qui consiste à nommer des couleurs). D'après Linden (1987), en effet, l'augmentation du rythme cardiaque lors d'une prise de parole serait d'autant plus prononcée que le contenu du discours est pertinent pour la personne. Tous ces résultats montrent qu'il est possible d'observer des variations différentes d'un même paramètre selon la tâche expérimentale utilisée. Dans la suite de ce manuscrit, on appellera *effet de la tâche* cette différenciation de la réponse physiologique potentiellement observée à l'échelle de plusieurs sujets entre des tâches stressantes différentes.

IV.C.4. L'effet du sujet

Dans l'étude de Skoluda et collaborateurs (2015), l'effort faible dû à la parole n'explique pas l'augmentation de cortisol constatée pour le TSST, qui marque un effet de la tâche sur la charge affective des participants. A ce sujet, Harvey et collaborateurs (2010) ont observé que le cortisol n'était sécrété par l'organisme que lorsque l'individu percevait le stressueur comme une *menace*. Conçu pour l'étude de l'attention sélective, le test de Stroop peut être stressant sans être menaçant ; ce qui explique l'augmentation du sentiment subjectif de stress, mais pas de la concentration en cortisol salivaire, observée dans notre exemple.

Lors de la conception d'un protocole expérimental, il faut ainsi s'attendre à ce que l'ensemble des manifestations du stress dépendent de la *perception* des tâches par chaque participant. A ce sujet, Tomaka et Blaskovitch (1994) ont montré que l'activité cardiaque en situation de stress était modulée par la « croyance en un monde juste » de leurs participants lors d'une même tâche arithmétique. Loin d'une réponse universelle, une tâche expérimentale stressante est ainsi susceptible de présenter un *spectre* de réactions physiologiques qui dépend du vécu des différents sujets. Dans la suite de ce manuscrit, on appellera *effet du sujet* cette différenciation de la réponse physiologique observée entre plusieurs individus pour une même tâche stressante.

IV.C.5. Synthèse : le rôle du protocole expérimental dans l'étude du stress

Ce que nous appelons *effet de l'activité*, *effet de la tâche* et *effet du sujet* constitue la face observable des facteurs confondants du stress d'origine psychologique (section III.C), à l'intérieur d'une base de données recueillie en laboratoire. Dans un protocole expérimental pour la caractérisation du stress, *l'effet de l'activité* et *l'effet de la tâche* peuvent être pris en compte en utilisant des tâches de contrôle et différentes tâches stressantes. Lorsqu'on cherche à identifier les corrélats du stress, *l'effet du sujet* peut être pris en compte en analysant les données d'individus différents sans s'attendre à une réponse uniforme sur l'ensemble des participants. Un effet du sujet peut être observé aussi bien dans la ligne de base, que dans la réactivité des paramètres physiologiques des participants lors des tâches expérimentales.

Ces trois effets sont connus dans la littérature en sciences affectives. Mais contrairement à la ligne de base, ils ne sont pas systématiquement pris en compte pour la conception d'un détecteur de stress – comme on pourra l'observer dans le Chapitre Trois - . Or, la base de données sur laquelle on élabore un détecteur de stress définit *implicitement* le concept que celui-ci pourra détecter. Pour un détecteur de stress conçu seulement sur la base du TSST, par exemple, les corrélats du stress sont seulement ceux d'une prise de parole en public. Mais ces corrélats pourraient ne pas se reproduire avec la même intensité dans d'autres situations stressantes : le détecteur de stress pourrait donc être inefficace dans un nouveau contexte. De même, si le protocole expérimental manquait d'une tâche contrôle, *l'effet de l'activité* ne pourra pas être pris en compte dans l'élaboration de ce détecteur de stress pour lequel la parole, quel que soit le contexte, sera toujours un acte stressant.

Pour finir, on peut noter que le caractère *stressant* des tâches classiquement utilisées (tâches présentées dans le paragraphe IV.B.2) est souvent obtenu au prix d'un effort mental élevé. Pour s'affranchir de cette contrainte, on pourrait chercher une tâche qui impose du stress sans effort. Mais à la différence de l'effort physique, il est délicat d'induire en laboratoire un stress réaliste qui ne s'accompagne pas d'effort mental (voir IV.B.1). Ainsi, la notion de *stress* ciblée par les tâches standard

est intrinsèquement liée à la notion d'effort mental. Celui-ci fait partie de la définition *implicite* donnée au stress par la majorité des recherches de la littérature. Lors d'une analyse des données expérimentales, la question pertinente pour le chercheur en informatique affective n'est plus seulement : « Qu'est-ce que le stress ? » – mais : « Qu'est-ce que je peux observer à partir d'une certaine base de données ? » Nous reprendrons cette question à notre compte au Chapitre Cinq - .

A RETENIR

- L'état de stress est une réponse de l'organisme au vécu d'une situation jugée comme difficile par la personne et pertinente pour son bien être.
- Les corrélats physiologiques du stress sont connus et mesurables. Cependant, ils ne sont pas spécifiques et restent sensibles à de nombreux facteurs confondants.
- Dans une base de données expérimentales ces facteurs confondants sont de quatre types : l'allostasie spontanée, l'effet de l'activité, l'effet de la tâche et l'effet du sujet.

Chapitre Trois - Conception d'un Détecteur de Stress : Etat de l'Art

Dans le chapitre précédent, on présentait différentes conceptions du stress ainsi que les méthodes expérimentales qui permettent de l'observer. Ce type de protocole permet de collecter des données qui seront utilisées pour concevoir un modèle de détection du stress.

La démarche de recherche qui permet d'aboutir à un modèle de stress peut être décomposée en cinq axes de travail. Le recueil de données expérimentales est le premier d'entre eux (axe I). Le travail de modélisation commence ensuite, avec le traitement des signaux mesurés (axe II.A) et l'extraction de caractéristiques sur les signaux traités (axe II.B). Il se poursuit avec une sélection de ces caractéristiques (axe III.A) ; et se termine avec l'apprentissage et la validation du modèle (axe III.B). Ces différents axes sont représentés sur la Figure 3.1.

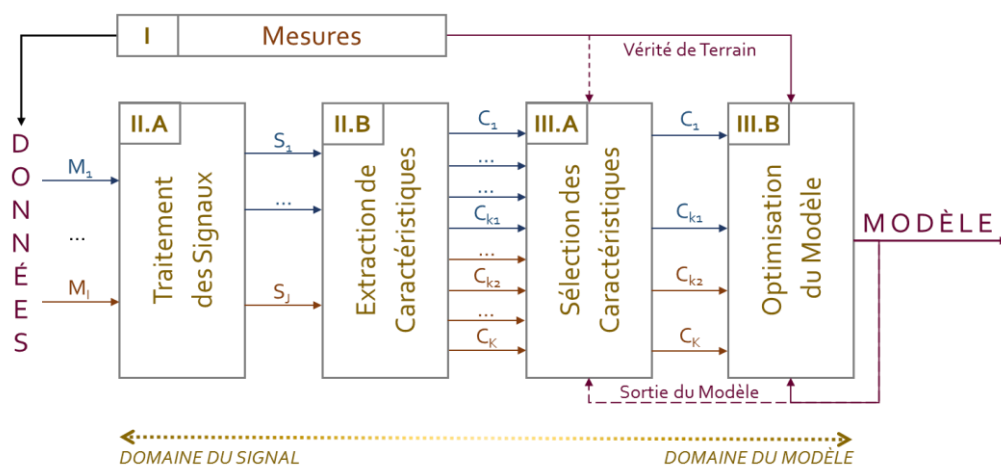


Figure 3.1 : Axes de recherche pour la construction d'un modèle de stress (cas de l'apprentissage supervisé)

Ce chapitre présente les principales orientations possibles pour conduire chacune des cinq étapes. Comme sur la Figure 3.1, la section I recensera les mesures physiologiques communément utilisées dans la littérature et leur lien avec le stress mental. La section II détaillera les traitements mis en œuvre et les caractéristiques extraites des mesures les plus pertinentes. La section III exposera les étapes conduisant à l'apprentissage d'un modèle de stress à partir de données de laboratoire ; et la section IV développera les obstacles qui s'opposent à l'application de ce modèle dans la vie quotidienne. Enfin, la section V donnera le positionnement de nos travaux de thèse à la lumière des éléments développés dans ce chapitre.

Dans cette démarche, nous serons amenés à questionner les choix de méthode qui ont permis de concevoir la majorité des détecteurs proposés dans la littérature scientifique. En particulier, nous montrerons que la conception d'un modèle *performant* (*i.e.*, qui détecte des épisodes de stress sans se tromper) ne garantit pas que celui-ci soit *valide* (*i.e.*, qui détecte effectivement du stress), ni même *robuste* (*i.e.*, qui résiste à un environnement de mesures dégradé) dans les situations de la vie quotidienne. Ces problématiques justifient l'emploi d'une approche en boîte-blanche parmi les méthodes qui seront proposées dans la suite de ce manuscrit.

I. Mesures Physiologiques

Le Tableau 3.1 (ci-dessous) donne le classement proposé par Sharma et Gedeon (2012), des mesures actuellement disponibles (et d'intérêt) pour la détection du stress aigu. Elles ont été classées selon leur corrélation avec le stress dans la littérature, leur fréquence d'utilisation dans la littérature sur le stress et le caractère intrusif du dispositif de mesure associé. A noter que parmi ces mesures, l'expression vocale et faciale (en gris) sont éliminées d'office car elles ne sont pas des mesures physiologiques.

Tableau 3.1 – Mesures d'intérêt pour la détection du stress aigu (Sharma & Gedeon, 2012)

Rang	Mesure	Rang	Mesure
1	Variabilité Cardiaque	8	Pression Sanguine
2	Activité Electrodermale	9	Température de la Peau
3	Électroencéphalographie	10	Oxymétrie de Pouls
4	Diamètre de la Pupille	11	Clignements de l'œil
5	Voix	12	Respiration
6	Direction du Regard	13	Électromyographie
7	Expression Faciale		

Parmi les mesures restantes, l'activité cardiovasculaire (variabilité cardiaque, pression sanguine, oxymétrie du pouls) et l'activité électrodermale méritent une attention particulière car elles sont particulièrement représentées dans la littérature sur la détection du stress. Elles feront ainsi l'objet des deux premières parties de cette section (I.A et I.B). On s'intéressera ensuite au potentiel des mesures restantes (sensibilité, applicabilité) pour l'estimation du stress en condition ambulatoire (I.C). On s'intéresse ici seulement aux dispositifs disponibles et à leur principe de mesure. Le lien de ces mesures avec le stress sera développé davantage en section II, hormis pour les mesures non considérées dans le cadre de cette thèse.

I.A. Activité Cardiovasculaire

Dans le Tableau 3.1, le mot *variabilité cardiaque* désigne un ensemble de propriétés du signal de rythme cardiaque. Celui-ci est accessible à partir de plusieurs types de signaux physiologiques.

I.A.1. Électrocardiographie

a) Principe

La mesure la plus commune dans la littérature sur la détection du stress est l'électrocardiographie (ECG), une mesure de la dynamique de l'activité du cœur basée sur le recueil des cycles de dépolarisation et repolarisation des muscles cardiaques (van Lien et al., 2015). Les appareils et techniques de mesure électrocardiographique sont très divers, avec des conséquences distinctes sur le signal enregistré, selon la localisation stratégique des électrodes sur la peau (Mulder, 1992). Un point commun entre toutes ces mesures se trouve dans le pic saillant *R* (ou complexe *QRS*), qui marque sans ambiguïté l'instant de chaque battement cardiaque. Ces battements peuvent être extraits automatiquement et en temps réel, *via* des méthodes standard de traitement du signal. Un algorithme

de référence opère par filtrages successifs : proposé par Pan et Tompkins (1985), il a été validé depuis sur de nombreuses bases de données.

b) Application à la mesure du rythme cardiaque en vie réelle

La mesure ECG est déjà largement exportée hors du laboratoire. Depuis plusieurs décennies, la technologie *Holter* permet la mesure ambulatoire au moyen d'électrodes humides semblables à celles d'un ECG classique (Kennedy, 2006). Depuis peu, l'ECG est également accessible au moyen de *textrodes*, des électrodes textiles désormais intégrables dans un vêtement quelconque (Gruetzmann et al., 2007). Un exemple de dispositif commercial utilisant la technologie *textrode* se trouve dans la ceinture thoracique Zephyr BioHarness 3 (Zephyr™, 2012). L'électrocardiographie peut aussi être intégrée dans un T-shirt connecté, tel qu'Hexoskin (Carre Technologies inc, 2012). Ces deux appareils, qui mesurent également la respiration, sont représentés dans la Figure 3.2. Plus récemment, une mesure ECG a été développée dans la version 4 de l'Apple Watch actuellement commercialisée aux Etats-Unis (Apple, 2018). A notre connaissance, cette technologie n'utilisant qu'une seule électrode n'a pas encore été validée par une étude indépendante.



Figure 3.2 : Dispositifs de mesure portables pour l'électrocardiographie.
(a) Zephyr BioHarness 3
(b) Hexoskin

Pour leur part, Johnstone et collaborateurs (2012) ont comparé la mesure de rythme cardiaque de BioHarness 3 avec celle d'un autre dispositif de mesure portable (Polar T31) au cours d'un protocole impliquant marche et jogging en conditions réelles. Une corrélation totale de 0.61 a été trouvée sur l'ensemble des données, et de 0.98 sur les données nettoyées. La mesure ECG peut être considérée comme relativement robuste, mais n'est pas infaillible en condition d'activité physique.

I.A.2. Photopléthysmographie

a) Principe

La photopléthysmographie (**PPG**), ou « oxymétrie de pouls » (Sinex, 1999), est une mesure conjointe de l'activité cardiaque et vasculaire. Le spectre d'absorption de la lumière par l'hémoglobine est différent selon que la molécule est chargée ou non en dioxygène. L'émission-réception d'un couple de rayons monochromes sur la peau (dans le proche infrarouge, le rouge ou le vert) permet d'estimer la saturation en dioxygène dans le sang artériel de l'épiderme. Le signal résultant est oscillatoire et synchronisé avec les battements cardiaques. Le graphique de la Figure 3.3 est issu de (Thomas et al., 2016) : il compare la mesure de deux battements de cœur selon les mesures d'un signal PPG (en rouge) et l'ECG (en bleu) sur deux battements cardiaques. La mesure en PPG est très sensible au mouvement, donc propice aux artéfacts ; et les oscillations lentes du signal le rendent moins adapté que l'ECG à une localisation temporelle précise des battements cardiaques.

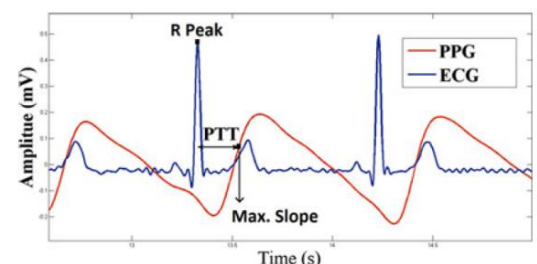


Figure 3.3 – Comparaison de battements cardiaques en ECG (bleu) et PPG (rouge) (Thomas et al, 2016)

Bien qu'un certain nombre d'algorithmes aient été proposés (Periyasamy et al., 2017; Schäfer & Vagedes, 2013), il n'existe pas de méthode de référence pour l'extraction du rythme cardiaque à partir du signal PPG.

b) Application à la mesure du rythme cardiaque en vie réelle

La mesure PPG est accessible au moyen d'un équipement léger (deux diodes côte-à-côte) qui peut être intégré à de nombreux dispositifs portés. Elle est courante dans les appareils connectés de type montre ou bracelet, comme par exemple, dans la montre Galaxy Watch (Samsung, 2018) et le bracelet Inspire HR (FitBit, 2019). Certains dispositifs sont développés spécialement pour la recherche et ont fait l'objet d'études comparatives, comme le bracelet multi-capteurs Empatica E4 (Empatica, 2013). Enfin, la mesure PPG ne se limite pas nécessairement au poignet : depuis peu, la technologie de patches Spire Health (SpireHealth, 2018) peut être embarquée sur n'importe quel vêtement. La Figure 3.4 ci-dessous montre ces deux derniers dispositifs.



Figure 3.4 – Dispositifs portables de mesure PPG.
(a) Empatica E4
(b) Spire Health Tag

Une étude publiée seulement sur le site d'Empatica (Empatica, 2014), compare les mesures de leur technologie PPG (sur bracelet Empatica E3) avec celles d'une instrumentation ECG standard, au cours d'un protocole impliquant peu de mouvement. La corrélation entre les caractéristiques extraites (9 sujets, 1 fenêtre de temps supérieure à 10 minutes) est proche de 1, mais cette étude n'a pas été reproduite en condition ambulatoire.

I.A.3. Cardiographie d'impédance

La cardiographie d'impédance mesure le volume sanguin à l'intérieur de la cage thoracique, à partir des fluctuations de la résistance électrique de cette dernière (Sherwood et al., 1990). En plus du rythme cardiaque, elle permet la mesure de paramètres hémodynamiques tels que le débit du cœur ou la vasomotricité dans les muscles squelettiques. À la différence de l'ECG, ce n'est pas une mesure passive : elle nécessite l'injection d'un faible courant à travers le corps de l'individu. Le graphique de la Figure 3.5 – Comparaison de battements cardiaques en ECG (en bas) et en Cardiographie d'impédance (ICG, en haut) (van Lien et al, 2015) ci-contre, issu de (van Lien et al., 2015), compare la mesure d'un battement de cœur par cardiographie d'impédance (au-dessus) et par ECG (en-dessous). On y observe l'instant réel de la contraction cardiaque (B-point), qui succède au pic R de quelques millisecondes.

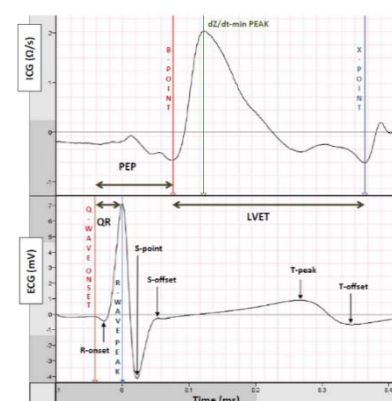


Figure 3.5 – Comparaison de battements cardiaques en ECG (en bas) et en Cardiographie d'impédance (ICG, en haut) (van Lien et al, 2015)

Bien que des tentatives existent pour mettre au point des dispositifs portables (Yazdanian et al., 2016), cette mesure n'est pas encore accessible en condition ambulatoire.

I.B. Activité Electrodermale

Le terme « activité électrodermale » (**AED**) regroupe toutes les mesures d'impédance de la peau. Cette mesure reflète l'activité sudomotrice (transpiration) entre les deux électrodes de mesure. Elle doit sa popularité au phénomène de « transpiration émotionnelle », que l'on peut mesurer sur des sites privilégiés du corps humain.

I.B.1. Principe et lieux privilégiés

La surface de la peau est divisée en dermatomes, des régions délimitées par les faisceaux nerveux qui affleurent en ces endroits (van Dooren et al., 2012). La densité des glandes sudoripares eccrines varie selon le dermatome ; cette densité est plus élevée sur la paume des mains et la plante des pieds. Ces deux régions, différenciées dès les premiers stades de l'ontogénèse, sont principalement innervées par des fibres sympathiques (Boucsein, 2012). Si leur implication dans la thermorégulation est négligeable au niveau de ces deux régions, elles s'activent en revanche lors d'un effort mental ou physique (Kerassidis, 1994). C'est donc dans les régions palmaires et plantaires que la mesure électrodermale est considérée comme un index valide de l'arousal physiologique (Picard et al., 2016).

Le signal électrodermal est obtenu en imposant, entre deux électrodes sur la surface de la peau, soit un potentiel (mesure en conductance ou admittance), soit un courant très faible (mesure en résistance ou impédance) qui peut être alternatif ou continu selon l'appareil de mesure.

La Figure 3.6 ci-contre est inspirée de Boccaletti et collaborateurs (2008) et complétée avec des éléments de Boucsein (2012). Sans représenter les échanges de courant entre deux électrodes, elle modélise l'impédance de la peau aux abords d'une glande sudoripare et permet de comprendre le lien entre activité électrodermale et sudation. Les couches inférieures de la peau (en bas) ont une impédance essentiellement résistive (R_D). Cette résistivité est faible et comparable à celle des tissus plus profonds. En revanche, les parties supérieures de l'épiderme (partie médiane de la figure) ont une résistivité forte (R_E) et des propriétés capacitives (C_E), qui varient selon l'environnement. Lorsque le canal d'une glande sudoripare se remplit (à droite de la figure), son impédance très faible ($C_G R_G$) fait l'effet d'un court-circuit vis-à-vis de l'épiderme. En diminuant la résistance de la peau, elle facilite le passage du courant vers l'électrode (en haut de la figure).

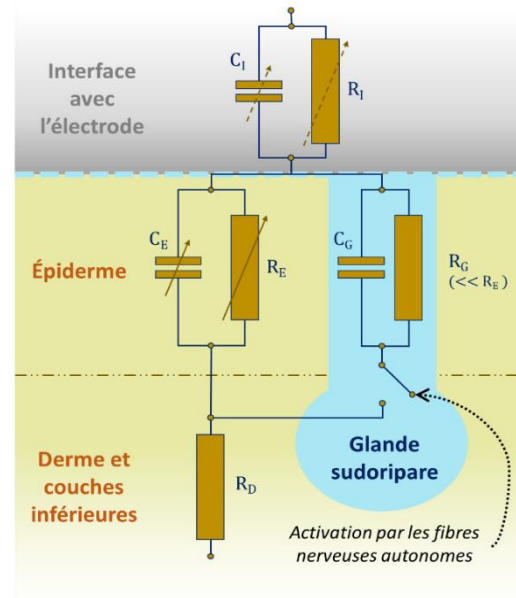


Figure 3.6 – Impédances équivalentes à l'interface entre la peau et une électrode.

Les différences de potentiel aux interfaces ne sont pas représentées.

À son tour, l'impédance à l'interface peau-électrode doit être considérée car elle dépend du capteur : on la modélise par le dipôle $R_i C_i$ sur la Figure 3.6. Cette impédance varie selon l'électrode de mesure et la quantité d'électrolyte, ce qui n'est pas sans effet sur la mesure électrodermale. En laboratoire, on privilégiera souvent des électrodes humides, mais celles-ci ne sont pas adaptées à une utilisation dans la vie quotidienne.

I.B.2. Applications pour la mesure en vie réelle

En vie réelle, la mesure d'AED peut être embarquée au moyen d'électrodes sèches dans un dispositif compact tel qu'un ensemble de bagues ou un bracelet de mesures. Par exemple, l'appareil StopSleep propose deux bagues jointes contenant quatre électrodes chacune, et un logiciel embarqué supposé surveiller l'état de vigilance d'un conducteur (Stop Sleep France, 2013). D'autres dispositifs optent pour une mesure sur le poignet, comme la montre Empatica E4. La technologie AED d'Empatica a également été intégrée dans Embrace, un bracelet spécialisé dans la prévention de l'épilepsie (Empatica, 2015a). StopSleep et Embrace sont toutes deux représentées dans la Figure 3.8 ci-dessous.



Dans une étude publiée seulement sur le site d'Empatica (2015b), leur technologie de mesure AED (sur bracelet Empatica E3) a été comparée avec celle d'une instrumentation de laboratoire. Les expérimentateurs ont disposé sur la même main un jeu d'électrodes Ag/AgCl sèches identiques pour chacun des deux dispositifs d'enregistrement. Les expérimentateurs ont montré une similarité des courbes d'AED et un coefficient de corrélation proche de 1 pour le nombre de réponses électrodermales détectées chez 9 des 10 participants enregistrés. Cette étude n'a pas été reproduite en condition ambulatoire.

I.C. Autres Mesures Disponibles : Pertinence et Perspectives

I.C.1. Activité respiratoire

L'activité respiratoire (parfois abrégée **Resp** dans la suite de ce manuscrit) est fréquemment mesurée en même temps que l'activité cardiaque. Des mesures de référence sont accessibles par spiromètre : rythme respiratoire, volume d'air inspiré ou encore (en combinaison avec d'autres dispositifs) consommation en dioxygène de l'organisme (Backs & Seljos, 1994). Mais cette technique implique une instrumentation lourde qui est inutilisable en vie réelle. Une mesure plus adaptée se trouve dans la pléthysmographie respiratoire à variation d'inductance. Elle repose sur l'effet piézoélectrique dans une jauge de contrainte intégrée à une ceinture élastique, généralement attachée autour du torse ou de l'abdomen. On la trouve, par exemple, dans les dispositifs cardiaques représentés Figure 3.2. Cette mesure fournit un signal oscillatoire dont on peut extraire le rythme respiratoire, voire estimer le volume d'air inspiré (Leutheuser et al., 2017).

Le rythme respiratoire, en particulier, a montré une augmentation en situation de stress et de charge cognitive (Backs & Seljos, 1994; Mehler et al., 2009; Tulen, Moleman, van Steenis, et al., 1989). Certaines situations de stress pourraient également déréguler le rythme respiratoire de sorte à élargir le spectre du signal (Gandhi et al., 2015). Cependant, les résultats pour ces caractéristiques ne sont pas concordants entre les études et cette mesure est fortement influencée par la parole (J. L. P. Wijsman, 2014).

I.C.2. Pression sanguine

La pression sanguine (systolique et diastolique) a été solidement corrélée au stress aigu (Sharma & Gedeon, 2012; Vrijkotte et al., 2000; J. L. P. Wijsman, 2014). Elle est accessible directement au moyen de dispositifs cliniques standard, l'étalon-or étant le sphygmomanomètre à mercure (Pickering Thomas G. et al., 2005). Cette instrumentation est encombrante et compresse le membre examiné, ce qui rend impossible son utilisation dans la vie quotidienne. Certains dispositifs permettant une mesure quotidienne de la pression sanguine commencent à voir le jour (H2CARE, 2019), mais à l'heure actuelle, aucun d'entre eux n'a fait l'objet d'une validation dans la littérature scientifique. De même, leur validité sur le terrain reste encore à établir et constitue un axe de recherche futur en traitement du signal physiologique. Enfin, un certain nombre de méthodes tentent de mesurer la pression sanguine au moyen du signal PPG, ou en fusionnant les informations issues des signaux entre la PPG et l'ECG (Y. Li et al., 2014; Thomas et al., 2016).

I.C.3. Température de la peau

La température de la peau est accessible par contact au moyen de thermistances ou de thermocouples (A. D. H. Smith et al., 2010), disponibles dans des dispositifs portables ou de laboratoire. La montre Empatica E4 intègre une solution quasi sans contact : une thermopile infrarouge, assortiment de thermocouples réagissant à l'énergie infrarouge renvoyée par la peau.

Un lien entre le stress et la température de la peau a été montré en laboratoire, sur le visage (Kataoka et al., 1998) et sur le poignet en vie réelle (J. L. P. Wijsman, 2014). Cette mesure reflèterait la vasomotricité périphérique (Reisman, 1997). Néanmoins, l'influence du dermatome sur ces résultats reste à déterminer ; d'autant plus que les études ne s'accordent pas sur le sens de variation de cette mesure relativement au stress. Les effets spécifiques de la charge mentale sur cette mesure n'ont pas non plus fait l'objet d'études rigoureuses.

I.C.4. Électromyographie

L'électromyographie (EMG) mesure les potentiels électriques générés par la tension musculaire. Certaines caractéristiques dérivées de la mesure EMG sur le muscle trapèze ont été corrélées au stress aigu (Wijsman, 2013), et l'activité myoélectrique du muscle en tension est supérieure lorsque du stress mental est ajouté à l'effort (Larsson et al., 1995). Ceci démontre une activité accrue du système nerveux somatique sur les muscles squelettiques durant un épisode de stress. La mesure de cette activation sur certains muscles du visage (corrugator supercilii, zygomatiques supérieurs) permet d'enregistrer une tendance à l'expression faciale ; elle est souvent utilisée dans la détection d'émotions en général (Broek et al., 2009). Cela dit, l'influence majeure de l'activité motrice sur une telle mesure (et son intrusivité lorsqu'elle est recueillie sur le visage) rend impossible son utilisation en vie réelle.

I.C.5. Oculométrie

De même, les mesures oculaires sont accessibles dans la vie quotidienne au moyen de lunettes instrumentalisées, comme les Pro Glasses (Tobii, 2015). En condition ambulatoire, ces dispositifs d'oculométrie reposent sur un enregistrement vidéo de l'œil humain. Une analyse des images permet d'extraire la direction du regard et le diamètre de la pupille, par exemple, et de repérer les clignements

oculaires. Le diamètre pupillaire a notamment montré son potentiel pour la détection du stress en laboratoire (Zhai & Barreto, 2006a). A la différence des capteurs cardiaques et électrodermaux, cependant, ces dispositifs restent encore onéreux et sont rarement utilisés par le grand public. Leur usage est pour l'instant réservé à la recherche dans certains domaines particuliers – comme la vision par ordinateur, l'ergonomie, l'industrie automobile.

I.C.6. Électroencéphalographie

Pour finir, l'électroencéphalographie de surface (EEG) mesure l'activité électrique du cerveau au moyen d'électrodes positionnées sur le crâne. L'activité EEG constitue une des mesures clé en neurosciences cognitives pour caractériser spécifiquement les états mentaux, en raison de sa résolution temporelle élevée, et de la diversité des caractéristiques pouvant être extraites de ce signal. En lien avec notre problématique de recherche, différents travaux ont identifié des biomarqueurs de la charge cognitive aussi bien dans des contextes de laboratoire que dans des contextes applicatifs (Roy et al., 2016; Roy et al., 2016). Dans la perspective d'estimation du stress, des caractéristiques simples comme une activité asymétrique dans certaines bandes de fréquence du signal EEG (*e.g.* bande alpha) entre les deux hémisphères, ont été identifiées comme corrélats (Lewis et al., 2007). Des essais ont été réalisés avec succès pour distinguer le stress du repos à partir de la simple mesure EEG (Hou et al., 2015). L'EEG peut également être utilisée en combinaison avec des mesures périphériques comme la PPG et l'AED pour détecter l'état de stress (Kalimeri & Saitis, 2016).

Il existe différents dispositifs de mesure EEG utilisables en ambulatoire. Depuis quelques années, certains appareils proposent une mesure EEG intégrée dans un casque audio ou dans une structure plus discrète, proposant un nombre réduit d'électrodes (Melomind, 2019; Neurosky, 2014). Cela dit, des perspectives de recherche résident dans l'acceptabilité par grand le public, de cette technologie dont le support (casque) ne fait pas l'objet d'un port systématique dans la vie quotidienne (Knight & Baber, 2005; Spagnolli et al., 2014). Bien qu'elle soit prometteuse dans un contexte applicatif (comme l'écoute de musique), la mesure EEG n'est pas encore susceptible d'identifier des épisodes de stress dans un contexte quelconque de la vie quotidienne.

I.C.7. Synthèse

Dans la perspective de développer un détecteur de stress, l'éventail des mesures disponibles est vaste. Les solutions proposées dans cette section ne couvrent d'ailleurs pas l'ensemble des recherches sur le sujet. Des mesures centrales comme l'imagerie par résonance magnétique (IRM), par exemple, ou la Tomographie par Emission de Positons (TEP), ont été (et sont encore) largement utilisées pour comprendre les corrélats neuronaux du stress (Dedovic et al., 2005, 2009). L'usage de ce type d'instrumentation n'est cependant pas envisageable en-dehors des environnements cliniques. Plus récemment, la spectroscopie dans l'infrarouge proche (NIRS, pour *Near Infrared Spectroscopy*) a prouvé son utilité pour la détection du stress en laboratoire (Gurel et al., 2019; Hakimi & Setarehdan, 2018). Une particularité de cette mesure est de permettre à la fois des analyses propres à l'imagerie cérébrale fonctionnelle et une estimation du rythme cardiaque. Comme pour l'EEG, cela dit, il n'existe pas encore de dispositif proposant une mesure de type NIRS et qui puisse faire l'objet d'un port systématique dans la vie quotidienne. L'indisponibilité actuelle des mesures centrales dans les applications grand-public, a ainsi justifié la focalisation de cette thèse sur la réponse allostatique médiée par le système nerveux périphérique.

Parmi les mesures périphériques précédemment citées, en effet, quatre sont accessibles en vie réelle *via* une instrumentation peu intrusive et déjà déployée sur le terrain : le rythme cardiaque, l'activité électrodermale, l'activité respiratoire et la température de la peau. Les dispositifs permettant actuellement ces mesures sont de type bracelet, ceinture thoracique ou T-shirt connecté, premiers ambassadeurs de l'Internet des Objets dans le marché des capteurs physiologiques. L'introduction de ces appareils a ouvert la voie à la détection d'épisodes de stress dans la vie quotidienne (voir section IV.C), à condition d'en extraire l'information pertinente sous forme de caractéristiques physiologiques.

II. Caractéristiques Extraites des Signaux Périphériques

L'extraction de caractéristiques consiste à calculer des paramètres du signal physiologique (*ex.* : niveau moyen, variance, puissance...) sur des fenêtres temporelles successives. On peut alors établir un lien entre la valeur de ces paramètres et l'état de stress de la personne. Il n'existe pas de procédure « standard » pour extraire des caractéristiques pertinentes de l'activité respiratoire ou de la température cutanée. En revanche, la littérature propose des paramètres de référence pour le rythme cardiaque et la conductance de la peau : ces caractéristiques font l'objet des sections qui suivent. Dans la suite ce manuscrit, on adoptera leur dénomination standard (en *anglais*) pour faciliter les références à la littérature.

II.A. Caractéristiques Extraites du Signal de Rythme Cardiaque

II.A.1. Rythme moyen – Rythme instantané

Un épisode de stress est connu pour entraîner une élévation du rendement cardiaque bien au-delà de la demande métabolique (voir Chapitre Deux - III.A et Annexe 1). Hors période d'effort, le rythme cardiaque moyen est un paramètre efficace pour distinguer le stress du repos, et parmi différents niveaux de stress (J. L. P. Wijsman, 2014). Sur les signaux oscillatoires tels que la PPG, on peut estimer cette caractéristique *via* l'autocorrélation ou la transformée de Fourier (Schäfer & Vagedes, 2013). Une méthode plus précise consiste à déterminer le rythme cardiaque *instantané* (dénomination standard : **HR**, de l'anglais *heart rate*), en mesurant l'intervalle temporel qui sépare deux battements de cœur sur une mesure quelconque (**IBI**, de l'anglais *interbeat interval*). Le signal HR, exprimé en battements par minute ([bpm]), est inversement proportionnel au signal IBI exprimé en secondes ([s]) (équation (3-1)). Il n'est pas échantillonné régulièrement dans le temps, mais à chaque fois qu'un nouveau battement est détecté. C'est la précision atteinte durant la détection qui fait de l'électrocardiographie un étalon-or pour mesurer le rythme cardiaque instantané.

$$HR(t) = 60 / IBI(t) \quad (3-1)$$

II.A.2. Variabilité

Le rythme cardiaque est un signal non stationnaire. Son parcours temporel, ou *variabilité cardiaque* (dénomination standard : **HRV**, de l'anglais *heart rate variability*), permet de mesurer l'équilibre des activités sympathique et parasympathique (Acharya et al., 2006). L'activité nerveuse parasympathique diminue lorsque le rythme cardiaque doit augmenter rapidement (voir aussi Annexe 1). L'activité nerveuse sympathique adapte (à plus long terme) le flux sanguin à une forte demande métabolique.

Le muscle cardiaque est ainsi continuellement soumis à des stimulations parasympathiques et sympathiques antagonistes, qui modulent son activité sur des fréquences différentes. D'où l'intérêt d'extraire des caractéristiques fréquentielles du signal de rythme cardiaque.

a) Caractéristiques fréquentielles

◆ Calcul

Afin de mesurer la contribution de chaque système, on peut intégrer la densité spectrale de puissance du signal HR dans certaines bandes de fréquences. Certains auteurs normalisent ces caractéristiques en les divisant par la puissance totale du signal (Kim et al., 2018). La densité spectrale de puissance peut être estimée de plusieurs manières. La plus répandue est l'analyse de Fourier, mais on peut aussi avoir recours à l'estimation autorégressive (AR) ou à la transformée en ondelettes (Acharya et al., 2006). Toutes ces techniques nécessitent soit d'interpoler le signal, soit de considérer (par approximation) qu'il est échantillonné régulièrement. Une alternative réside dans la méthode Lomb-Scargle, adaptée aux signaux dont l'échantillonnage est irrégulier (Muaremi et al., 2013).

◆ Hautes fréquences

Pour un enregistrement cardiaque, on appelle « hautes fréquences » (**HF**) l'intervalle compris entre 0,15Hz et 0,5Hz (ce dernier chiffre correspond à la fréquence de Nyquist pour un signal HR à 60bpm, le rythme moyen d'un individu au repos). Comme l'influence sympathique est lente, la puissance dans cette bande de fréquence serait seulement liée à l'activité autonome parasympathique, ou *tonus vagal* (Acharya et al., 2006; Kim et al., 2018). Ainsi, une diminution de cette puissance spectrale indique un retrait de l'activité nerveuse parasympathique. Par conséquent, la puissance HF serait négativement corrélée au stress aigu (Hjortskov, Rissén, et al., 2004). Cette caractéristique a parfois été considérée comme un indice de « stress » à part entière (Porges, 1995).

◆ Basses fréquences

On appelle « basses fréquences » (**LF**, de l'anglais *low frequencies*) l'intervalle compris entre 0,04Hz et 0,15Hz. La puissance dans cette plage de fréquence a été attribuée à l'activité nerveuse sympathique (Acharya et al., 2006; Kim et al., 2018), mais aussi à de l'activité parasympathique (Fairclough & Mulder, 2011). Elle serait influencée par un mécanisme de régulation de la pression sanguine par le rendement cardiaque (« baroréflexe »), qui fonctionne par inhibition vagale et activation sympathique (Fairclough & Mulder, 2011; Paas et al., 1994). Par conséquent, la puissance LF serait positivement corrélée au stress aigu (Taelman et al., 2011).

◆ Autres caractéristiques

Lorsque la fenêtre temporelle est suffisamment longue, on peut aussi considérer les « très basses fréquences » (VLF, de l'anglais : *very low frequencies*), situées entre 0,003Hz et 0,04Hz. Cette plage est influencée par les fonctions très lentes du système nerveux autonome comme la thermorégulation (Acharya et al., 2006; Paas et al., 1994). Les « ultra basses fréquences » (ULF), inférieures à 0,003Hz, ne sont adaptées qu'aux fenêtres de très longue durée. Ces deux caractéristiques ont été peu ou pas corrélées au stress aigu.

Il est enfin possible de combiner ces caractéristiques fréquentielles ; par exemple, en divisant la puissance LF par la puissance HF. Ce calcul permettrait de mesurer l'activité sympathique en s'affranchissant du tonus vagal (Kim et al., 2018). En toute cohérence, cette caractéristique a été positivement corrélée au stress aigu (Hjortskov, Rissén, et al., 2004).

b) Caractéristiques temporelles

La variabilité cardiaque se mesure aussi directement sur le signal temporel HR (Kim et al., 2018). Une mesure simple est l'écart-type du signal IBI sur la fenêtre de temps. Mais on mesure plus fidèlement la *variabilité* du signal avec les différences successives entre deux intervalles inter-battements. On estime l'énergie de cette composante en calculant la racine du carré moyen (**RMS**, de l'anglais *Root Mean Square*), d'après l'équation (3-2) ci-dessous où $n \in \{1..N\}$ repère les échantillons successifs de la fenêtre de temps. La dénomination standard de cette caractéristique est : **RMSSD**, de l'anglais : *root mean square of the successive differences*.

$$RMSSD_{IBI} = \sqrt{\frac{1}{N} \sum_{n=1}^N (IBI_{n+1} - IBI_n)^2} \quad (3-2)$$

Cette caractéristique a été négativement corrélée au niveau de stress (Vrijkotte et al., 2000). Elle est fortement corrélée à la puissance HF : ces deux mesurent reflètent le tonus vagal (Sztajzel, 2004). La RMSSD et la puissance spectrale sont presque toujours calculées sur le signal IBI, mais il n'est pas incorrect de les calculer sur le signal HR, si l'on garde en tête que la puissance de ce signal est modulée par son niveau moyen (Fairclough & Mulder, 2011).

II.A.3. Effet du Stress et de la charge mentale sur les caractéristiques cardiaques

Le Tableau 3.2 ci-dessous résume et schématise l'évolution attendue des principales caractéristiques cardiaques, ainsi que les mécanismes autonomes justifiant cette évolution.

Tableau 3.2 – Évolution attendue des caractéristiques cardiaques standard en situation de stress et influence prédominante de l'activité autonome

Caractéristique	Évolution attendue	Influence supposée
Moyenne (IBI)	↘	P↘ + S↗
RMSSD	↘	P↘
Puissance HF	↘	P↘
Puissance LF	↗	P↘ + S↗
Rapport LF/HF	↗	S↗

S : sympathique – P : parasympathique

Dans leur revue de littératures, Kim et collaborateurs (2018) ont observé un manque de consensus entre les études quant à la réactivité au stress des caractéristiques HRV. Cette inconsistance pourrait s'expliquer par la forte diversité des protocoles expérimentaux utilisés pour induire et mesurer l'influence du stress, dont certains seront détaillés plus tard dans cette partie.

Par ailleurs, la variabilité cardiaque est une mesure privilégiée de l'effort mental. Les caractéristiques HRV seraient sensibles à l'effort mental (même faible), mais seraient incapables de distinguer plusieurs niveaux d'effort (Paas et al., 1994; Veltman & Gaillard, 1998). Ces variations feraient partie d'une réaction de « défense » de l'organisme à une activité mentale quelconque (Fairclough & Mulder, 2011). Par ailleurs, Mandrick et collaborateurs (2016) ont observé que la RMSSD diminue lorsque la charge mémorielle augmente en l'absence de stress, mais pas en présence de

stress. Ainsi, la variation des caractéristiques HRV serait déclenchée par la simple présence d'une charge mentale – avec un effet additionnel de la charge affective, cet effet étant encore mal connu.

II.B. Caractéristiques Extraites du Signal Electrodermal

II.B.1. Composante tonique et composante phasique

Comme pour le signal cardiaque, le signal électrodermal comprend des composantes à variations lentes et rapides. On les appelle, respectivement, l'activité tonique et l'activité phasique (Benedek & Kaernbach, 2010). Celles-ci sont sous-tendues par l'activité des fibres nerveuses sympathiques.

La composante *tonique*, à variations lentes, correspond au niveau de base de l'AED à un instant donné. Elle dépend des propriétés physiques de l'épiderme (composition, température, humidité), ainsi que de l'état d'éveil général (*arousal*) de la personne (Boucsein, 2012). Dans le cas d'une mesure en conductance (SC, de l'anglais : *skin conductance*), sa dénomination standard est **SCL** (*skin conductance level*).

La composante *phasique*, à variations rapides, est constituée de pics (ou séries de pics) appelés *réponses électrodermales*. Celles-ci reflètent les impulsions électriques des fibres nerveuses sympathiques. Dans le cas d'une mesure en conductance, par exemple, elle se traduit par des pics positifs dont la dénomination standard est **SCR** (de l'anglais : *skin conductance responses*). Sur le signal électrodermal, la fréquence des SCRs est un indicateur notoire de la réactivité de l'activité sympathique cutanée (Boucsein, 2012).

Une réponse électrodermale se caractérise par différentes dimensions, représentées sur la Figure 3.9 ci-contre. Son amplitude A est défini par la différence du niveau électrodermal entre le pic (phasique) et le niveau de base (tonique). Elle est proportionnelle à la réactivité de la fibre nerveuse efférente (Benedek & Kaernbach, 2010). Sa période d'ascension T_A est le temps qui sépare le pic du début d'apparition de la SCR. Elle est généralement inférieure à cinq secondes (Green et al., 2014) et dépend de l'amplitude de la SCR. Sa période de recouvrement T_R est le temps qui sépare le pic de l'instant où le signal électrodermal rejoint son niveau de base. Ces trois caractéristiques peuvent être combinées : $T_T = T_A + T_R$ est la durée totale d'une réponse, et $A * T_T$ (parfois nommé « produit du pic ») est une estimée de son aire. Si une séparation des composantes est disponible, cette aire peut être calculée en intégrant entre les courbes SC et SCL. En laboratoire, on peut également observer le temps de latence entre l'apparition d'un stimulus et celui d'une réponse électrodermale.

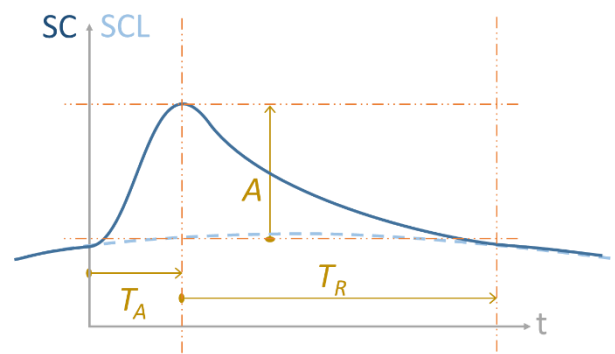


Figure 3.9 – Schéma d'une réponse en conductance de la peau (SC) et ses paramètres, relativement à la composante tonique du signal (SCL)

La séparation des composantes tonique et phasique ne comprend pas de méthodologie standard. L'activité phasique peut être isolée par simple filtrage (C. Setz et al., 2010) ; mais cette méthode peut dénaturer la forme caractéristique des réponses électrodermales. Une autre approche consiste à détecter les SCRs à partir des dimensions caractéristiques mentionnées précédemment (Green et al., 2014). Enfin, l'activité phasique peut être estimée par des grandeurs dérivées, comme la puissance de la dérivée seconde du signal électrodermal (J. Wijsman et al., 2013).

II.B.2. Effet du stress et de la charge mentale sur les caractéristiques électrodermales

Une revue de Lazarus (1966) montrait déjà une forte corrélation entre le niveau de stress et le niveau SCL moyen ou la fréquence des SCRs. Dans leur tentative de modélisation, Healey & Picard (2005) ont déterminé la SCL moyenne comme caractéristique la plus corrélée, parmi d'autres mesures de rythme cardiaque, d'électromyographie et de respiration, au niveau de stress de participants effectuant une tâche de conduite. L'activité électrodermale phasique a été considérée comme la mesure par excellence de l'*arousal* (Picard et al., 2016), parfois comme un indicateur du stress mental (Kocielnik et al., 2013). Enfin, les dimensions caractéristiques d'une réponse électrodermale sont susceptibles de refléter l'état affectif de la personne (Boucsein, 2012; Broek et al., 2009). Cependant, l'efficacité de ces caractéristiques en estimation du stress manque de support empirique. Aussi, le Tableau 3.3 prédit l'évolution des deux caractéristiques électrodermales considérées comme standard.

Tableau 3.3 – Évolution attendue des caractéristiques électrodermales standard en situation de stress et influence prédominante de l'activité autonome

Caractéristique	Évolution attendue	Influence supposée
Moyenne (SCL)	↗	S↗
Fréquence des SCRs	↗	S↗

S : sympathique – P : parasympathique

À nouveau, cependant, ces deux caractéristiques ont été liées à l'effort mental. Par exemple, une augmentation de la fréquence des SCRs et de la SCL ont été observées lors d'un épisode de conduite sur autoroute, lorsqu'un effort mental supplémentaire était induit par une tâche annexe simple (Mehler et al., 2009; Verwey & Veltman, 1996). La proximité de ces analyses avec celles de Healey & Picard (2005) illustre la réciprocité des notions de stress et d'effort mental. Kilpatrick (1972) a suggéré que l'AED tonique serait liée à la charge mentale, tandis que l'AED phasique serait liée à la charge affective. À nouveau, cette hypothèse manque de support empirique consistant.

II.C. Utilisation des Caractéristiques dans un Modèle de Stress

II.C.1. Caractéristiques les plus corrélées au stress dans l'état de l'art

Les caractéristiques précédentes sont fréquemment rencontrées dans les détecteurs de stress élaborés durant les quinze dernières années. Dans ce contexte, il est tentant de proposer une revue de littérature qui mettrait en évidence les caractéristiques les « plus corrélées » au stress. Par exemple, Wijsman (2014) a rassemblé une vingtaine d'études publiées entre 2003 et 2013 ; et recherché les liens établis entre le stress et certaines caractéristiques cardiaques, électrodermales et respiratoires. Une adaptation graphique de ses résultats est représentée dans la Figure 3.10. Les caractéristiques sont classées par ordre de congruence entre les études, c'est-à-dire la proportion d'études ayant montré une évolution similaire de ces caractéristiques en situation de stress.

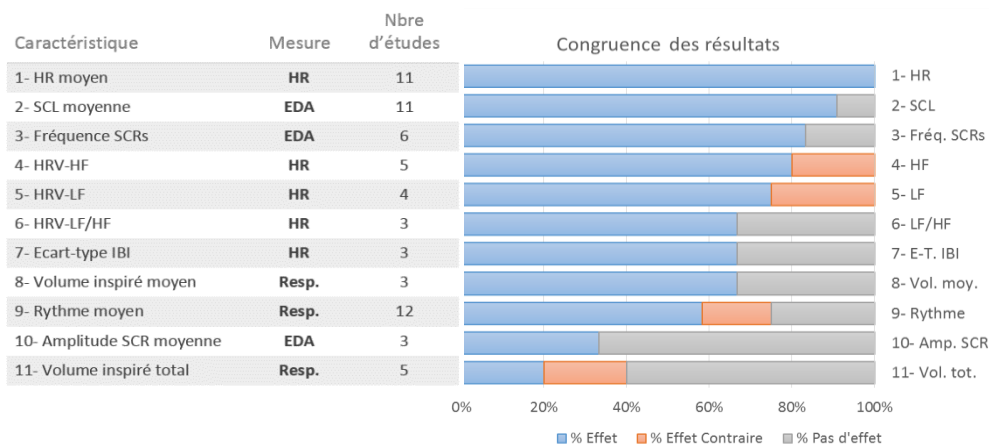


Figure 3.10 – Classement des caractéristiques selon la revue de (Wijsman, 2014)

Pour chaque caractéristique, la partie bleue du graphique en barre indique la proportion des études qui ont montré une évolution significative dans le même sens. La partie rouge indique la proportion d'études ayant montré un effet contraire. La parties grise indique la proportion d'études n'ayant pas montré d'effet.
 HR : rythme cardiaque, EDA : Activité électrodermale, Resp. : Signal Respiratoire

On constate que la littérature peine à mettre en évidence des corrélats universels, si ce n'est les trois premières caractéristiques : le rythme cardiaque moyen, le niveau électrodermal moyen et fréquence des réponses électrodermales. Ce type d'analyse analyse permet de repérer les caractéristiques stables à travers les conditions expérimentales. Cependant, il ne rend pas compte de l'effet de l'activité dans les variations mesurées pour les caractéristiques décrites ci-dessus. Ainsi les conclusions de la section III.B nous amèneront à mettre en doute la pertinence de ces résultats.

II.C.2. Influence de l'individu

En outre, la majorité des caractéristiques physiologiques est sujette à des variations inter-individuelles, c'est-à-dire à l'influence de la biologie ou du style de vie de la personne. Par exemple, la puissance HF du rythme cardiaque est connue pour diminuer au cours du vieillissement (Karlsson et al., 2012). De même, la prise de médicaments peut influencer les caractéristiques HRV (Acharya et al., 2006). La consommation d'alcool ou de tabac peut augmenter l'activité sympathique et/ou diminuer l'activité parasympathique sur le coeur. D'après Boucsein (2012), les caractéristiques électrodermales dépendent aussi directement de l'individu. Le genre et l'origine démographique, notamment, qui influencent la densité des glandes sudoripares. De même, l'âge et le degré d'exposition au soleil modifient les propriétés électriques de l'épiderme.

Ces propriétés individuelles participent largement à l'effet du sujet, un concept introduit au Chapitre Deux - pour désigner les différences interindividuelles dans la réaction physiologique au stress. Si on considère que cet effet est prévisible pour un individu donné, des méthodes standard existent pour le corriger : la mesure d'une ligne de base ou la création d'un modèle pour chaque individu. Ces stratégies sont détaillées en Annexe 3. Enfin, bien que les différences *inter-individus* soient un phénomène bien connu dans la littérature en détection du stress, les différences *inter-tâches* dans la réponse de chaque caractéristique sont relativement peu explorées. Cette problématique est abordée du point de vue de l'apprentissage, dans la section III.B.

III. Apprentissage d'un Détecteur de Stress

Après l'extraction des caractéristiques, la construction d'un modèle de détection du stress passe par l'utilisation de méthodes issues de l'apprentissage automatique. La méthodologie dans ce domaine est riche et ses applications s'étendent au-delà de nos problématiques. Il est inutile d'inventorier ici tous les algorithmes qui peuvent aboutir à un modèle de détection du stress. Cependant, le choix d'une méthode d'apprentissage peut avoir des implications profondes sur les performances et la validité de ce modèle. On discutera dans cette partie des différentes étapes de l'élaboration d'un modèle de détection du stress : la sélection des caractéristiques, le choix de l'algorithme et la méthode de validation croisée.

III.A. Apprentissage Supervisé

III.A.1. Usage de la supervision

On peut apprendre un modèle de manière supervisée ou non-supervisée. L'apprentissage *supervisé* consiste à faire en sorte que l'estimée (la sortie réelle de l'algorithme) s'approche d'une *vérité de terrain*. Cette approche suppose que le concept-cible (ici, le stress) est connu et observable. A l'inverse, l'apprentissage *non-supervisé* ne nécessite pas de vérité de terrain. On laisse l'algorithme décider lui-même quelle sera la sortie du modèle, à partir de propriétés statistiques ou de similitudes dans les données d'entrées.

L'approche non-supervisée est particulièrement utilisée sur les grands jeux de données (*big data*). Par exemple, en bio-informatique, pour grouper entre *clusters* des caractéristiques de l'information génétique (Hastie et al., 2009). Pour l'évaluation du stress, cette approche disposera d'un certain potentiel lorsqu'une grande quantité d'information physiologique multi-mesures sera disponible dans les *clouds* gérés par les fabricants d'objets connectés. Samsung (2017, 2018), Apple (2018, 2019), pour ne citer qu'eux, ont lancé leur propre gamme d'objets connectés (voir section I) ; avec l'objectif affiché de détecter le stress des consommateurs. A l'heure actuelle, cependant, l'approche non-supervisée se révèle peu efficace sur le petit jeu de données accessible au moyen de l'expérimentation scientifique (Can et al., 2019). Par ailleurs, la proximité de l'état de stress avec ses facteurs confondants (activité physique, effort mental) suggère que ceux-ci ne sont pas séparables sans supervision au moyen de caractéristiques standard.

Si certaines exceptions demeurent, une large majorité des études académiques pour la détection du stress a donc eu recours aux techniques supervisées (Sharma & Gedeon, 2012). A notre connaissance, Healey et Picard (2005) furent les premières à utiliser l'apprentissage pour concevoir un modèle de détection du stress ; et à mesurer ses performances par validation croisée. De précédentes tentatives ont utilisé la logique floue (Sin et al., 1999; Sul et al., 2002), mais ces modèles dépendent de lois définies à l'avance par le chercheur et n'ont pas fait l'objet d'une mesure de performances.

La Figure 3.11 ci-dessous apporte des précisions à la Figure 3.1 présentée en début de partie pour schématiser un protocole d'apprentissage complet. L'axe III.B distingue deux choix en particulier : celui de la technique d'apprentissage et celui de la méthode de validation croisée. En amont de ces étapes, la sélection de caractéristique (axe III.A) permet de réduire l'ensemble des variables utiles pour détecter le stress.

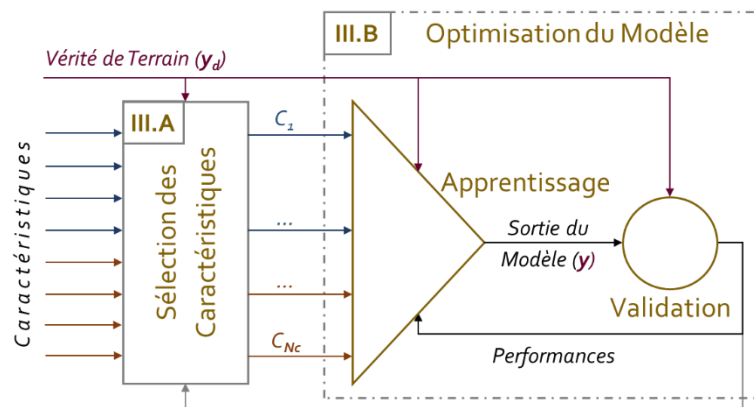


Figure 3.11 : Schéma de méthodologie générale en apprentissage supervisé

III.A.2. Sélection des caractéristiques

Les bénéfices de la sélection de caractéristiques sont multiples : augmenter les performances de l'algorithme en diminuant le risque de surapprentissage, réduire le nombre de mesures nécessaires, ou obtenir une meilleure compréhension des mesures liées au concept-cible. Kohavi et Sommerfield (1995) ont distingué les techniques de sélection de caractéristiques en deux catégories : les approches de type *filtre* et les approches de type *wrapper* (dénomination anglaise standard).

- Le type *filtre* classe les variables sur la base de leur « pertinence » relativement au concept-cible. On utilise des indicateurs statistiques pour trouver les plus sensibles à la sortie désirée.
- Le type *wrapper* classe les variables sur la base de leur « utilité » à l'intérieur d'un sous-ensemble de caractéristiques. Les étapes III.A et III.B de la Figure 3.11 sont itérées entre elles pour sélectionner le groupe de variables qui obtiendra les meilleures performances pendant l'apprentissage.

Certaines techniques, comme l'analyse en composantes principales (ACP) ou en composantes indépendantes (ACI), combinent les caractéristiques pour former des bases orthogonales dans cet espace. D'autres, dites *intégrées*, réalisent la sélection des caractéristiques en même temps que l'apprentissage (Guyon & Elisseff, 2003). Cela dit, la distinction *filter* / *wrapper* illustre une subtilité fondamentale dans les méthodes de l'apprentissage automatique.

Les méthodes de type *wrapper* ont une approche en *boîte noire* : elles ne permettent pas d'expliquer la sélection d'un ensemble de caractéristiques plutôt qu'un autre (Guyon & Elisseff, 2003). En revanche, elles permettent d'adapter cet ensemble à la méthode d'apprentissage utilisée, ce qui peut augmenter les performances du modèle final. Les techniques de type *filtre* ne prennent pas cette variable en compte, ce qui peut limiter les performances du modèle final. En retour, ce type d'approche en *boîte blanche* décrit le lien entre chaque caractéristique et le concept-cible. Le critère de sélection des caractéristiques est connu. Ainsi, l'expérience peut être reproduite à l'identique sur une base de données quelconque. Par la suite, les résultats sont interprétables et peuvent être comparés à la littérature existante.

III.A.3. Algorithmes d'apprentissage

a) Principes fondamentaux

Une fois sélectionnées, la fusion des caractéristiques appelle le choix d'un algorithme d'apprentissage. La sortie du modèle peut avoir deux échelles : discrète ou continue. Lorsque la sortie est continue, on parle de *régression*. Lorsqu'elle est discrète, on parle de *classification*. La classification fournit un étalon commode à la mesure des performances, chaque sortie pouvant être qualifiée de *vraie* ou de *fausse* par comparaison à la vérité de terrain. On peut ainsi calculer un pourcentage de « bonnes » classifications : on obtient la *précision* du modèle, une mesure standard de sa performance.

Lorsque le problème possède deux classes, on parle de *détection*. En modélisation du stress, c'est l'objectif privilégié des protocoles d'apprentissage. D'une part, parce qu'il est plus facile d'apprendre deux états (Stress et Non-Stress) qu'une multitude (classification) ou qu'une infinité (régression). D'autre part, parce qu'elle facilite la comparaison avec d'autres études existantes. Par exemple, on peut calculer la « précision » d'un algorithme régressif en plaçant un seuil sur la sortie continue (qui devient une sortie à deux classes). Et enfin, parce que la sortie désirée peut être fixée par les conditions expérimentales : les tâches stressantes et la condition référence fixent les classes Stress et Non-Stress.

Ainsi, la majorité des méthodes d'apprentissage recensées par Sharma et Gedeon (2012) pour la modélisation du stress est composée de *détecteurs* ; et les *conditions expérimentales* produisent une forme consensuelle de vérité de terrain. Le choix de l'algorithme dépend d'hypothèses sur la « bonne » manière de combiner les différentes caractéristiques, ou sur la distribution des données dans l'espace des caractéristiques. Selon la méthode, la structure du modèle final sera plus ou moins accessible à l'analyse, la contrepartie étant un modèle plus ajusté aux données d'apprentissage.

Le Tableau 3.4 ci-dessous propose une liste d'algorithmes *off-the-shelf* proposés dans la majorité des problèmes d'apprentissage. Les algorithmes de la colonne de gauche permettent l'exploration d'une base de données (*data mining*), au prix d'hypothèses fortes qui peuvent restreindre les performances du modèle final. Les algorithmes de la colonne de droite sont optimisés pour prédire la sortie désirée. On retrouve la distinction déjà évoquée entre une approche en *boîte blanche* et une approche en *boîte-noire*.

Tableau 3.4 : Algorithmes d'apprentissage communs

Modèles plus interprétables	Modèles plus ajustés aux données
Analyse Discriminante (linéaire, quadratique)	Algorithme des Plus Proches Voisins (K-NN)
Algorithme Bayésien Naïf	Séparateur à Vastes Marges (SVM)
Arbres de Décision	Forêts Aléatoires
Régression Logistique	Réseaux de Neurones

b) Frontière de décision

Dans le problème de *détection*, on caractérise un modèle par la limite de décision qui sépare les deux classes (ex. : Stress et Non-Stress) dans l'espace des caractéristiques. Ainsi, on peut illustrer la distinction boîte-noire / boîte-blanche dans la frontière de décision de deux algorithmes en particulier : l'analyse discriminante linéaire et la classification par plus proches voisins. Comme chaque exemple du Tableau 3.4, ces techniques ont été utilisées pour détecter le stress ou l'effort mental (Cinaz et al.,

2013; Healey & Picard, 2005; Mozos et al., 2016). La Figure 3.12 (ci-dessous) reprend une simulation de Hastie et collaborateurs (2009) pour représenter la limite de décision entre deux classes (« orange » et « bleu ») dans un espace à deux caractéristiques.

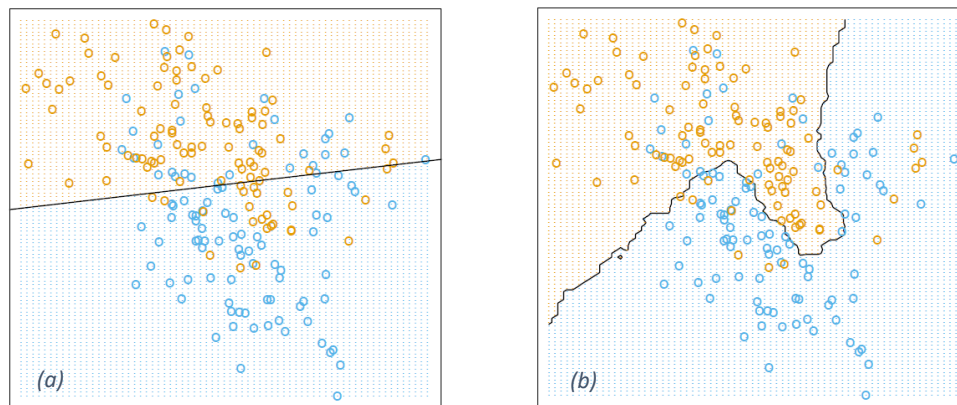


Figure 3.12 – Frontières de décision dans le cas de la détection par analyse discriminante (a) et par les 15 plus proches voisins (b), dans un espace à deux caractéristiques. D'après Hastie et collaborateurs (2009).

En analyse discriminante (*boîte-blanche*, graphique (a)), la limite de décision est facilement paramétrable : il s'agit ici d'une ligne droite dans l'espace 2D des caractéristiques. Les paramètres de cette droite sont peu sensibles aux données sélectionnées pour l'apprentissage, pourvu qu'elles soient représentatives de chaque classe. En revanche, l'hypothèse qu'une droite suffit à séparer ces deux classes pose une contrainte forte sur les performances du modèle final.

A l'inverse, la méthode des plus proches voisins (*boîte-noire*, graphique (b)) fait peu d'hypothèses sur la limite de décision. Ce degré de liberté supplémentaire lui octroie souvent de bonnes performances. En revanche, cette limite de décision n'est pas paramétrable : dans un espace à plus de 2 caractéristiques, il est difficile de mesurer dans quelle mesure chacune d'elles contribue à la sortie du modèle. Enfin, la limite de décision dépend plus fortement des données choisies pour l'apprentissage.

III.A.4. Evaluation des performances : la validation croisée

Dans la majorité des études sur l'estimation du stress, la *performance* d'un classifieur qualifie la capacité de l'algorithme à reconnaître le stress et se mesure par la proportion d'exemples bien classés dans leur catégorie. Nous conserverons cette définition dans la suite de ce manuscrit.

a) Principe général

Lorsqu'un modèle appris sur un certain jeu de données montre de mauvaises performances dans un autre contexte, on parle de *surapprentissage*. De manière générale, plus le nombre de paramètres d'un modèle est élevé par rapport au nombre d'exemples, plus le risque de surapprentissage est élevé. La validation croisée fournit une mesure des capacités de *généralisation* d'un modèle pour un nombre donné de paramètres, lorsqu'on dispose d'un nombre limité d'exemples pour apprendre le modèle.

La méthode consiste à diviser la base de données en deux sous-ensembles : une base A (apprentissage) et une base V (validation). Le modèle est appris sur la base A et ses performances sont mesurées sur la base V. L'opération est répétée autant de fois qu'il est possible de trouver une nouvelle base V dont les exemples n'ont pas encore servi à la validation. Les capacités de généralisation du

modèle (estimées par validation croisée) correspondent à la performance moyenne de l'algorithme sur toutes les bases V . Ce processus est illustré dans sur la Figure 3.13 ci-dessous¹.

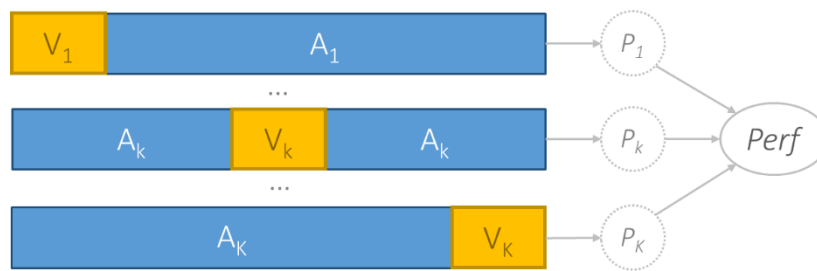


Figure 3.13 : Validation croisée : divisions successives de la base d'apprentissage.

A_k : Base d'apprentissage N^k . V_k : Base de validation N^k . P_k : Performance mesurée N^k . Perf : Performance finale.

Il existe différentes variantes dans la manière de choisir les bases de validation. L'approche *K-fold* (dénomination anglaise standard) est la plus courante : elle consiste à choisir la base V de manière aléatoire parmi les exemples n'ayant pas encore servi à la validation (Hastie et al., 2009). Cette méthode a l'avantage de pouvoir s'appliquer de manière automatique à n'importe quel problème d'apprentissage. Lorsque l'opération est répétée K fois, les données de validation représentent alors une fraction $1/K$ de l'ensemble de départ et les données d'apprentissage une fraction $(K-1)/K$. Lorsqu'un seul exemple constitue la base V à chaque étape de validation, on parle « d'un contre tous », ou *leave-one-out* (dénomination anglaise standard). Cette méthode est utile lorsque peu d'exemples sont disponibles pour constituer la base d'apprentissage.

b) Intérêt du *leave-one-out* pour la prise en compte de l'effet du sujet ou de la tâche

Lorsqu'on valide un modèle de stress sur de nouveaux exemples au hasard, sa « capacité de généralisation » contrôle la variance générale des données dans chaque classe. Mais dans le cas d'un modèle de stress, cette variance ne provient pas seulement du bruit de chaque mesure. Elle dépend aussi de causes externes, profondément liées aux conditions de la mesure. Deux de ces causes ont été détaillées à la fin du Chapitre Deux - : *l'effet de la tâche* et de *l'effet du sujet*. Ceux-ci posent une problématique essentielle pour l'efficacité du modèle en vie réelle : dans quelle mesure celui-ci peut-il être généralisé sur une personne ou une activité inconnues ? Les méthodes citées plus haut ne permettent pas de mesurer une telle capacité, car elles incluent des exemples de chaque tâche et de chaque individu dans la base d'apprentissage.

En revanche, il est possible d'isoler judicieusement les bases de validation (Sakri et al., 2018). Lorsqu'on dispose d'une base de données enregistrée sur N participants, on peut entraîner l'algorithme sur les données de $N-1$ d'entre eux, puis valider le modèle sur les données du N -ième. L'opération est répétée N fois, de sorte que chaque participant soit isolé comme base de validation. Cette variante de l'approche *leave-one-out* (**LOSO**, pour *leave-one-subject-out*) contrôle ainsi l'influence du facteur Sujet sur les performances du modèle. Lorsque les données disponibles sont issues d'une campagne expérimentale, on pourrait faire de même pour chaque tâche du protocole

¹ Remarque : En plus de mesurer les capacités de généralisation du modèle, la validation croisée permet d'ajuster ses *hyperparamètres* lorsque ceux-ci existent ; comme le nombre de plus proches voisins dans l'algorithme du même nom. On peut aussi s'en servir pour choisir l'algorithme d'apprentissage ou le type de modèle le plus adaptée aux données. Lorsque le modèle est ajusté par validation croisée sur une base de données X , ses performances finales doivent être mesurées sur une base Y (la base de *test*) indépendante de X .

(**LOTO**, pour *leave-one-task-out*). Ces configurations garantissent, dans une certaine mesure, la *validité externe* des performances revendiquées par le modèle de stress. Ce terme désigne la reproductibilité des résultats pour des individus ou des situations autres que ceux considérés lors de la construction du modèle (Calder et al., 1982).

III.B. Comparaison de Modèles Publiés dans la Littérature

Dans l'immense majorité, les protocoles d'apprentissage d'un modèle de stress se basent sur une campagne de mesures réalisées en conditions contrôlées (voir Chapitre Deux - IV). Si les possibilités d'apprentissage sont nombreuses, la majorité des algorithmes reposent ainsi sur des tâches standardisées issues d'un paradigme expérimental commun. Comme nous l'avons précisé au Chapitre Deux - IV plusieurs facteurs dits *confondants* peuvent être pris en compte dans la mise en place du protocole expérimental : l'effet de l'activité (motrice ou mentale), l'effet de la tâche et l'effet du sujet. La prise en compte de ces facteurs peut rendre la détection du stress plus difficile. En effet, il sera bien plus facile de concevoir un détecteur de stress performant quand celui-ci est appris et évalué à partir de données provenant d'un unique sujet, ne réalisant qu'une unique tâche stressante, que l'on oppose à une période de repos. Dans ce paragraphe, nous allons analyser et comparer les algorithmes de la littérature en regard de ces facteurs confondants. Ont-ils été pris en compte par les auteurs ? Quels impacts ont-ils sur les performances mesurées ?

III.B.1. Méthodologie de comparaison des modèles

a) Critères de sélection des modèles dans la littérature

On a comparé la performance en classification de modèles de stress élaborés dans un cadre similaire aux travaux de ce manuscrit. Ce cadre peut être défini par trois critères de sélection.

1. Le modèle cherche à estimer du *stress mental* dans les situations où l'activité physique est faible.
2. Le modèle s'appuie exclusivement sur les *signaux physiologiques* présentés en section I, à l'exception de l'électroencéphalographie. (L'inclusion de signaux d'accéléromètre a été tolérée, en raison de la place croissante accordée à ce type de mesures en condition ambulatoire et de la disponibilité de ce signal dans la majorité des capteurs portables.)
3. Les performances mesurées sont celles d'un *problème de détection*. Les résultats sont associés à une stratégie de validation ; et non issus de la phase d'apprentissage.

Au total, quinze modèles ont passé le filtre de ces trois critères. Parmi les études considérées, certaines proposent plusieurs définitions du stress (Mohino-Herranz et al., 2015). Dans ce contexte, on a sélectionné le paradigme expérimental le plus proche de ceux décrits au Chapitre Deux - IV, à savoir les situations pouvant être vécues comme difficiles par les participants et pertinentes pour leur bien-être. D'autres études ont testé parallèlement plusieurs algorithmes d'apprentissage (Setz et al., 2010; Zhai & Barreto, 2006b). Dans ce contexte, on a sélectionné celui qui donnait les meilleures performances. Une autre étude a optimisé son modèle pour plusieurs types de validation croisée (Sun et al., 2012). Dans ce contexte, on a choisi le meilleur modèle pour chaque type de validation.

Pour finir, le score de performance retenu pour chaque modèle est le nombre d'exemples correctement classés dans leur catégorie (Stress, Non-Stress). Cette information était toujours disponible dans les études considérées. Certains auteurs ont cherché à détecter le stress dans un problème à trois classes : Non-Stress, Stress Modéré, Stress Fort (Gjoreski et al., 2016; Healey & Picard,

2005). Dans ce contexte, on a calculé la performance pour un problème à deux classes (Stress VS Non-Stress), à partir de la matrice de confusion proposée.

b) Critères de comparaison des modèles retenus

Six critères ont été retenus pour comparer les méthodologies d'apprentissage des modèles sélectionnés (voir plus bas). Comme on a pu l'établir au Chapitre Deux - IV, la prise en compte de *l'effet d'activité* est un critère essentiel pour distinguer le stress d'une activité non-stressante (ex. : effort mental faible). Cet objectif doit être prévu dans le protocole expérimental d'une part, qui doit inclure une tâche contrôle ; et dans le processus d'apprentissage d'autre part, qui doit incorporer les données de cette tâche contrôle dans la classe *non-stress*. Le respect de cette procédure correspond à notre critère C1.

Dans les sections précédentes, on a aussi rappelé l'importance de contrôler *l'effet de la tâche* et *l'effet du sujet* sur la mesure des performances. La prise en compte de ces effets dans le protocole expérimental d'une part, et dans la validation croisée d'autre part, correspond aux critères C2 à C3bis proposés dans cette analyse de la littérature. Pour finir, on cherche à mesurer l'impact réel d'un apprentissage en boîte-blanche sur l'efficacité d'un détecteur de stress. L'utilisation d'un modèle interprétable correspond au critère C0. Ainsi, les six critères retenus pour analyser le protocole d'apprentissage de chaque modèle considéré dans cette analyse sont résumés ci-dessous.

- *C0 : Modèle en boîte-blanche.* La décision prise par le détecteur de stress est-elle facilement interprétable ?
- *C1 : Contrôle de l'activité.* Le modèle peut-il distinguer le stress d'une activité (motrice ou mentale) non-stressante ?
- *C2 : Entraînement multi-tâche.* Le modèle est-il appris sur des tâches stressantes distinctes ?
- *C2_{bis} : Validation multi-tâche.* Les performances sont-elles mesurées sur une tâche différente de celles utilisées lors de l'apprentissage ?
- *C3 : Entraînement multi-sujet.* Le modèle est-il appris sur des participants distincts ?
- *C3_{bis} : Validation multi-sujet.* Les performances sont-elles mesurées sur un participant différent de ceux utilisés lors de l'apprentissage ?

Concrètement, le critère C0 a été validé lorsque l'algorithme d'apprentissage correspondait à ceux mentionnés la colonne de gauche du Tableau 3.4. Les critères C2_{bis} et C3_{bis} ont été respectivement validés lorsque les performances du modèle ont été respectivement mesurées par validation croisée de type *LOTO (leave-one-task-out)* et *LOSO (leave-one-subject-out)*. Dans chaque étude, on a aussi relevé les tâches utilisées pour induire le stress : la nature de celles-ci peut expliquer en partie la performance de l'algorithme pour la détection du stress mental.

c) Illustration : analyse comparative de trois modèles

Pour illustrer notre méthode, le Tableau 3.5 ci-dessous propose d'analyser trois modèles de stress.

Healey et Picard (2005) ont correctement classifié 97% des fenêtres de temps sur leurs trois niveaux de stress respectifs : non-stress (repos), stress faible (conduite sur autoroute) et stress fort (conduite en centre-ville). Ramené à un problème de détection (Repos versus Stress), leur modèle obtenait 100% de bonnes classifications avec une validation croisée de type *k-fold* sur l'ensemble de leurs sujets (critère C3 valide, C3bis invalide). Dans cette étude, cependant, la condition de *non-stress* était instanciée par une tâche de repos (critère C1 invalide).

Tableau 3.5 : Echantillon de modèles pour la détection du stress et validation de trois critères d'observation.

<i>Etude</i>	<i>(Healey & Picard, 2005)</i>	<i>(Setz et al., 2010)</i>	<i>(Sandulescu et al., 2015)</i>
Stresseur(s) utilisé(s)	Conduite en voiture	Tâche arithmétique	Tâche arithm. + TSST
Mesure(s) incluse(s)	ECG, AED, Resp, EMG	AED	AED, PPG
Algorithme d'apprentissage	Analyse Discriminante	Analyse Discriminante	Sép. à Vastes Marges
Précision obtenue	100,0%	82,8%	80,9%
C0. Modèle interprétable	1	1	0
C1. Utilisation d'une tâche contrôle	0	1	1
C2. Entraînement multi-tâche	0	0	1
C2bis. Validation multi-tâche (LOTO)	0	0	0
C3. Entraînement multi-sujet	1	1	0
C3bis. Validation multi-sujet (LOSO)	0	1	0

Mesures – AED : Activité électrodermale. ECG : Electrocardiographie. EMG : Electromyographie. PPG : Photopléthysmographie. Resp. : Signal Respiratoire. Tâches – CPT : Cold Pressor Test. TSST : Trier Social Stress Test.

De leur côté, Setz et collaborateurs (2010) ont tenté de distinguer le stress de la charge mentale au moyen d'une tâche arithmétique en conditions Stress et Contrôle (critère C1 valide). Après validation croisée de type *LOSO* (critères C3 et C3bis valides), leur algorithme final revendique 83% de bonnes classifications. Enfin, Sandulescu et collaborateurs (2015) ont construit un modèle de stress pour chacun de leurs participants (critères C3 et C3bis invalides) ; tandis que leur protocole expérimental proposait plusieurs types de stresseurs (critère C2 valide). En utilisant une même tâche Contrôle pour condition de référence, leur protocole de validation a mesuré une précision de 81%.

Chacun des modèles analysés ici valide à sa manière nos six critères d'observation ; et propose son propre score de performance. Pour les 15 modèles retenus dans cette section, ces informations et la validation de nos six critères sont affichées en Annexe 4. Dans les paragraphes qui suivent, on se pose la question suivante : dans quelle mesure la « rigueur » du processus d'apprentissage (c'est-à-dire la validation des six critères précédents) influence-t-elle la performance mesurée par validation croisée ?

III.B.2. Influence de la prise en compte des facteurs confondants

Cinq de nos critères concernent la prise en compte de trois facteurs confondants : l'effet de l'activité (C1), l'effet de la tâche (C2 + C2_{bis}), et l'effet du sujet (C3 + C3_{bis}). On souhaite illustrer le lien entre la prise en compte de ces facteurs confondants d'une part, et la performance mesurée pour chaque modèle d'autre par, sur l'ensemble des 15 modèles retenus dans la section précédente.

a) Calcul d'un score d'apprentissage

En première approche, on peut synthétiser le respect de nos cinq critères avec un *score d'apprentissage* qui pondère équitablement le contrôle des trois facteurs confondants (C1 pondéré deux fois plus que les autres critères étant le seul à prendre compte l'effet d'activité). Pour chaque modèle, le calcul de ce score est donné par la formule ci-dessous (la valeur de chaque critère vaut 1 si ce critère est validé par le modèle en question ; et 0 sinon).

$$\text{Score} = 2 * C1 + C2 + C2_{bis} + C3 + C3_{bis} \quad (3-3)$$

Ce score d'apprentissage a été calculé pour les 15 modèles retenus. Normalement compris entre 0 et 6, les scores varient en pratique de 1 à 5. Seuls quatre modèles obtiennent un score supérieur à 3. La Figure 3.14 ci-contre affiche le score d'apprentissage en fonction de la performance mesurée.

Sur ce graphique, la ligne continue est issue d'une régression linéaire sur l'ensemble des points (qui correspondent chacun à un modèle). D'après son coefficient de détermination (R^2), cette droite explique 54% de la variance observée parmi les performances revendiquées dans notre corpus de modèles. La relation négative entre score et performance est significative ($p < 0,01$), avec un coefficient de corrélation de $-0,74$. Cette analyse montre une influence *générale* du contrôle des facteurs confondants sur le pourcentage d'exemples bien classés par le modèle final. Plus le score est élevé, plus la stratégie d'apprentissage a intégré de facteurs confondante ; et plus la performance semble diminuer.

b) Effet de chaque critère d'apprentissage

A l'aide du score d'apprentissage, on a montré que la validation de *plusieurs* critères fait baisser la performance mesurée par validation croisée. On s'intéresse maintenant à l'effet de *chaque* critère sur cette performance mesurée.

Dans ce but, le Tableau 3.6 nous permet d'observer plus en détail le lien entre les critères et les performances pour les 15 modèles retenus pour l'analyse. Chaque ligne du tableau représente un des six critères. Pour chaque critère, la première colonne précise le nombre de modèles qui le valident (sur l'ensemble des modèles qui peuvent le valider). On rappelle en effet qu'une validation croisée multi-tâche (type LOTO) est impossible dans un protocole mono-tâche ; et qu'une validation croisée multi-sujets (type LOSO) est incohérente dans un modèle mono-sujet. Ainsi, seuls les modèles validant les critères C2/C3 peuvent aussi valider les critères C2bis/C3bis.

Tableau 3.6 : Précision des modèles de stress en fonction de chaque critère d'observation (moyenne \pm écart-type)

Critères	Nb Modèles : Critère Valide	Précision : Critère Valide	Précision : Critère Invalide
C0. Modèle interprétable	5 / 15	86 \pm 10 %	84 \pm 10 %
C1. Utilisation d'une tâche Contrôle	6 / 15	76 \pm 7 %	90 \pm 7 %
C2. Entraînement multi-tâche	9 / 15	81 \pm 10 %	89 \pm 8 %
C2bis. Validation multi-tâche (LOTO)	0 / 9	-	-
C3. Entraînement multi-sujet	12 / 15	85 \pm 11 %	82 \pm 8 %
C3bis. Validation multi-sujet (LOSO)	7 / 12	82 \pm 9 %	89 \pm 12 %
<i>Tous les modèles</i>	15 / 15	84,6 \pm 8,0 %	

Les deux dernières colonnes présentent la performance (moyenne \pm écart-type) obtenue par les modèles qui valident le critère et par ceux que ne le valident pas. A nouveau, les modèles ne validant

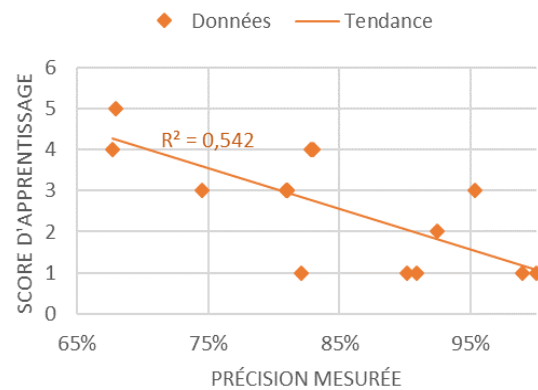


Figure 3.14 : Lien empirique entre la performance des modèles et le contrôle de facteurs confondants

pas les critères C2_{bis} ou C3_{bis} ont été sélectionnés parmi ceux qui validaient C2 ou C3. Ce résultat est disponible pour chaque critère, sauf pour la validation croisée multi-tâche (critère C2_{bis}) qui n'a été proposée dans aucune des 9 études ayant validé le critère C2.

Conformément à nos attentes, la prise en compte de l'effet de l'activité dans le protocole expérimental (C1) montre un coût élevé sur les performances mesurées (-14% en moyenne, soit deux fois l'écart-type, entre les deux dernières colonnes du Tableau 3.6). Une tâche stressante est plus difficile à discriminer d'une tâche contrôle que d'une tâche de repos. L'entraînement multi-tâche (C2) et la validation multi-sujet LOSO (C3_{bis}) semblent aussi diminuer les performances mesurées (entre les deux colonnes : -8% et -7% en moyenne, des chiffres proches des écarts-types). En revanche, le Tableau 3.6 ne montre pas l'impact attendu pour le critère C3 : les performances moyennes sont de 85% pour un entraînement multi-sujet et de 82% pour un entraînement mono-sujet, une différence bien inférieure aux écarts-types dans chacun des groupes.

Cela dit, il convient de rester prudent quant aux conclusions que l'on pourrait tirer de ce tableau. On peut noter, par exemple, que seules trois études n'ont pas validé le critère C3 en proposant un modèle *mono-utilisateur* (c'est-à-dire, entraîné et validé indépendamment sur chaque participant). Dans chacun des articles, les auteurs affirment que cette stratégie était nécessaire pour obtenir des modèles performants. Dans notre analyse (voir Annexe 4), les 3 modèles qui ne *valident pas* le critère C3 valident cependant le critère C2 (entraînement multi-tâche) ; et 2 sur 3 valident le critère C1 (utilisation d'une tâche contrôle). Or, parmi les modèles qui *valident* le critère C3, la validation des critères C1 et C2 est bien plus rare (*resp.* : 4 et 6 modèles sur 12). Ainsi, la prise en compte de *l'effet de l'activité* et de *l'effet de la tâche* dans les modèles mono-sujets (C3 invalide) peut expliquer la faible différence de performance observée entre les deux colonnes de la ligne « C3 » du Tableau 3.6.

III.B.3. Influence du type de modèle utilisé : boîte-noire ou boîte-blanche

Pour finir, le Tableau 3.6 ne montre aucune différence de performance entre les modèles de type *boîte-noire* et les modèles de type *boîte-blanche* (+2% en moyenne pour les modèles qui valident C0, une différence bien inférieure aux écarts-types de chaque colonne). Bien que les modèles de type *boîte-noire* soient deux fois plus nombreux que les modèles de type *boîte-blanche* (10 contre 5), la validation des autres critères (C1, etc.) est identique en proportion pour les deux types de modèles². Ainsi, la prise en compte des facteurs confondants n'explique pas l'absence d'effet observée dans le Tableau 3.6. Cela dit, d'autres différences méthodologiques subsistent entre les modèles retenus dans cette analyse (la nature et l'intensité des stressseurs, les mesures et les caractéristiques disponibles...) et nous empêchent de conclure que les deux types de modèles ont des performances équivalentes.

Parmi les études analysées dans le corpus, cependant, certaines ont eu recours à une *sélection de modèles*. Les résultats publiés permettent de comparer plusieurs algorithmes d'apprentissage sur un même jeu de données, en gommant les différences méthodologiques observées plus haut. Par exemple, Zhai et Barreto (2006b) ont confronté trois algorithmes dans une validation croisée de type *k-fold*. Leur séparateur à vastes marges (boîte-noire) a obtenu une précision de 90,1% ; l'arbre de décision (boîte-blanche) a obtenu une précision de 88,0% ; et l'algorithme bayésien naïf (boîte-

² Parmi les modèles de type boîte-blanche : 2/5 valident le critère C1 (contre 4/10 parmi les boîtes noires), 3/5 valident le critère C2 (contre 6/10 parmi les boîtes noires), 2/5 valident le critère C3bis (contre 5/10 parmi les boîtes noires). Voir Annexe 4.

blanche) a obtenu une précision de 78,7%. Les auteurs ont finalement sélectionné le séparateur à vastes marges (boîte-noire).

De leur côté, Sun et collaborateurs (2012) ont comparé ces trois mêmes algorithmes pour deux types de validation croisée : *k-fold* et *LOSO*. Contrairement à Zhai et Barreto (2006b), leur validation *k-fold* a désigné l'arbre de décision (boîte-blanche) comme technique la plus efficace. Parallèlement, la validation *LOSO* a sélectionné le séparateur à vastes marges (boîte-noire). Du côté de Setz et collaborateurs (2010), cependant, la validation *LOSO* a mesuré des performances supérieures pour l'analyse discriminante (boîte-blanche), par rapport à quatre versions différentes du séparateur à vastes marges. Enfin, Gjoreski et collaborateurs (2017) ont sélectionné le séparateur à vastes marges (boîte-noire) dans un problème de détection à trois classes avec validation *LOSO*.

Sur la base de cette analyse, il est à nouveau difficile d'établir la supériorité des modèles de type *boîte-noire* sur les modèles de type *boîte-blanche*. Les différences de résultat entre ces quatre études peuvent s'expliquer par les divergences fortes en termes de protocole expérimental, de mesures utilisées et de sélection des caractéristiques³. Par ailleurs, un modèle de stress est souvent sélectionné à quelques points de performance près, sur des bases de données toujours restreintes (généralement 20 participants), à l'image des résultats de Zhai et Barreto (2006b) reportés ci-dessus. Ainsi, les différences observées dans une même étude entre différents modèles de stress sont souvent peu significatives. Elles le sont d'une part, au regard du nombre d'exemples disponibles pour calculer la performance ; et d'autre part, au regard de l'influence du protocole expérimental sur la mesure de cette performance (voir Tableau 3.6).

III.B.4. Synthèse

Contrairement à d'autres sujets d'étude en apprentissage, il est délicat de comparer des détecteurs de stress sur la base des performances mesurées par validation croisée. En effet, les paragraphes précédents ont permis de montrer que cette performance était fortement déterminée par la prise en compte des facteurs confondants définis au Chapitre Deux - IV : *l'effet de l'activité, l'effet de la tâche et l'effet du sujet*. Ces résultats suggèrent d'une part, qu'il « difficile » pour les modèles physiologiques de distinguer le stress d'un effort mental non-stressant. Ils suggèrent d'autre part, qu'il est « difficile » pour ces modèles de généraliser dans un contexte (sujet, tâche) différent de la base d'apprentissage.

De même, l'étude précédente a permis de montrer que le traitement des facteurs confondants était très hétérogène dans la littérature en détection du stress. Cette incohérence méthodologique compromet fortement la synthèse comparative proposée en II.C, qui suggérait un classement des caractéristiques selon leur réponse au stress. Ainsi, ce travail de sélection des caractéristiques doit être repensé dans l'objectif explicite de contrôler ces facteurs confondants – en particulier *l'activité*.

³ Zhai et Barreto (2006b) ont utilisé des mesures électrodermales, du rythme cardiaque, de la température cutanée et du diamètre de la pupille ; et ont testé leurs modèles sur un même ensemble de caractéristiques issues de ces quatre mesures. Par comparaison, Setz et collaborateurs (2010) ont utilisé seulement l'activité électrodermale ; et ont sélectionné leurs caractéristiques indépendamment pour chaque modèle, avec un algorithme de type *wrapper*. Gjoreski et collaborateurs (2017) ont inclus directement une soixantaine de caractéristiques issues de l'AED, de la PPG, de la température de la peau dans l'apprentissage de leurs modèles. Les trois expériences citées étaient mono-tâches et réalisées en position assise ; tandis que celle de Sun et collaborateurs (2012) était multi-tâche et contenait de la marche légère. Deux de ces quatre expériences n'ont pas utilisé de condition contrôle (Sun et al., 2012; Zhai & Barreto, 2006b).

Enfin, l'avantage des modèles de type *boîte-noire* (peu interprétables, mais plus ajustés aux données) n'est pas évident à l'examen des résultats publiés dans la littérature. Cette conclusion s'impose malgré le fait que la plupart de ces modèles peuvent ajuster leurs paramètres pendant la validation croisée, augmentant ainsi la performance mesurée. Dans les paragraphes qui suivent, nous serons amenés à questionner le choix de ces algorithmes pour des applications sur le terrain.

IV. Obstacles à l'Utilisation d'un Modèle en Vie Réelle

Dans cette quatrième section, on identifie les difficultés liées à l'application d'un modèle de stress dans la vie quotidienne. De notre point de vue, ces difficultés s'articulent autour de deux notions-clés : la validité des modèles et la qualité des mesures en vie réelle. Ces deux notions sont définies dans les sections IV.A et IV.B. Les éléments développés ici nous serviront à analyser les études ayant traité de la détection du stress en vie réelle dans la section IV.C.

IV.A. Validité d'un Modèle de Stress

IV.A.1. Définition

Les résultats de notre comparaison de modèles soulèvent la question suivante : dans quelle mesure un détecteur revendiqué comme *performant* fait-il un modèle *valide* du stress mental ? Comme dans la section précédente, on peut chercher à « mesurer » la validité d'un modèle de stress à l'aide de plusieurs critères prédéfinis. Broek et collaborateurs (2009) ont traité cette problématique dans le domaine de l'informatique affective, en proposant certains « prérequis » à observer lorsqu'on modélise un phénomène mental (ici : le *stress*) à partir de mesures physiologiques. Dans notre cas :

- i. La *validation du concept* consiste à développer un cadre théorique précis autour de la notion de « stress ». Durant cette étape préliminaire, il faut expliciter le lien entre le stress, ses corrélats physiologiques et leurs facteurs confondants.
- ii. La *validité du contenu* concerne le degré de similitude entre le niveau de stress et les caractéristiques physiologiques. Elle diminue avec l'influence des facteurs confondants.
- iii. La *validité des critères* concerne le degré de similitude entre le niveau de stress et la vérité de terrain. Pour garantir ce lien, on peut utiliser plusieurs vérités de terrain (questionnaire, conditions expérimentales standard, etc.) ; puis vérifier que ces mesures sont bien corrélées entre elles.
- iv. La *validité écologique* concerne la capacité du modèle à estimer le stress dans la vie réelle.

Les notions (i), (ii) et (iii) ne sont pas nouvelles : elles font écho à plus d'un demi-siècle de développements théoriques en sciences humaines (Cronbach & Meehl, 1955; G. T. Smith, 2005). Formellement, elles déterminent la « validité du concept » (ici le stress) dont on cherche à construire une « mesure objective » (ici basée sur un modèle physiologique). De même, la « validité écologique » d'un détecteur de stress est devenue pertinente depuis que l'utilisation de capteurs portatifs rend possible son application dans la vie quotidienne.

IV.A.2. Illustration

A titre d'exemple, Broek et collaborateurs (2009) ont analysé la validité du détecteur de stress proposé par Healey et Picard (2005). Leurs conclusions sont reportées dans le Tableau 3.7. Pour

mémoire, l'étude en question avait utilisé trois conditions expérimentales (*resp.* repos, conduite sur autoroute et conduite en ville) pour produire trois niveaux de stress distincts (*resp.* nul, faible et fort). Le détecteur de stress proposé revendiquait 97% d'exemples bien classés dans un problème à deux classes et une validation croisée de type *k-fold*.

Tableau 3.7 : Validité du modèle de Healey et Picard (2005) selon Broek et collaborateurs (2009)

Critères	Conclusion	Justification
(i) Validation du concept	Mauvaise	Les auteurs ont identifié le stress au déséquilibre entre l'activité nerveuse sympathique et parasympathique ; mais elles n'ont pas décrit le lien entre cet équilibre et leurs mesures périphériques.
(ii) Validité du contenu	Limitée	Les signaux physiologiques sont confondus par le mouvement, qui ne fait l'objet d'aucun contrôle dans cette étude.
(iii) Validité des critères	Bonne	Les auteurs ont utilisé une évaluation subjective (échelles visuelles) pour valider les niveaux de stress supposés par le protocole.
(iv) Validité écologique	Excellente	Les auteurs ont cherché à évaluer le stress du conducteur et ont enregistré leurs participants en situation réelle de conduite.

Certains éléments développés dans ce chapitre nous permettent de compléter le point de vue de Broek et collaborateurs (2009) sur ce détecteur de stress. En particulier, on a déjà critiqué la *validité du contenu* de ce détecteur dans la section III.B, à travers la prise en compte des facteurs confondants.

D'une part, le détecteur en question a été appris et testé sur une seule tâche de conduite : la validité de ce modèle (ou même des caractéristiques mises en évidence) dans une autre situation est donc discutable. D'autre part, l'utilisation d'une validation croisée de type *k-fold* ne permet pas réellement de prendre compte *l'effet du sujet*. Enfin, le concept de *stress* (tel qu'on l'a défini au Chapitre Deux - III.C) n'est pas réductible au déséquilibre sympatho-vagal (voir Tableau 3.7, ligne 1) ; car les indicateurs de ce déséquilibre sont aussi visibles en situation d'effort mental non-stressant. L'absence de prise en compte d'un *effet de l'activité* nous empêche de dire si le modèle proposé détecte le stress ou l'effort lié à la conduite.

IV.A.3. Synthèse

Comme dans le dans le Tableau 3.7, les notions présentées dans cette section peuvent servir de métrique pour évaluer la qualité d'un détecteur de stress dans la vie quotidienne. Par exemple, la *validité du contenu* du détecteur concerne le contrôle des facteurs confondants dans le processus d'apprentissage. Comme on l'a vu dans la section III.B, cette question a des implications concrètes sur la performance du détecteur dans un contexte différent de la base d'apprentissage. De même, la *validité écologique* du détecteur correspond à sa performance sur des données de terrain

A ce sujet, L'exemple de Healey et Picard nous a permis d'observer un paradoxe dans la conception d'un détecteur de stress. D'une part, la *validité écologique* du détecteur ne peut être évaluée qu'en effectuant des enregistrements sur le terrain. D'autre part, la *validité du contenu* du détecteur doit être garantie par un protocole expérimental qui permet le contrôle de facteurs confondants.

Par conséquent, un détecteur de stress doit être validé sur deux bases de données distinctes : (i) des données issues de tâches expérimentales standard et (ii) des données issues de la vie quotidienne. Cette application « en vie réelle » d'un modèle « de laboratoire » est la démarche adoptée par un certain nombre d'études sur la détection en vie réelle, que nous détaillons dans la section IV.C.

IV.B. Qualité des Mesures dans la Vie Quotidienne

Lorsque le contexte de mesure n'est pas contrôlé par l'expérimentateur (port des capteurs dans la vie quotidienne), les signaux physiologiques recueillis sont sensiblement différents de ceux obtenus en laboratoire. La qualité fluctuante des enregistrements physiologiques en condition ambulatoire est un problème bien connu en sciences expérimentales.

IV.B.1. Altération du signal physiologique

Des mesures effectuées avec des capteurs portés par la personne présentent des *artefacts* ; ces composantes parasites brèves rencontrées, par exemple, lorsqu'une électrode se décolle de la peau. En vie réelle, le mouvement est la principale source d'artefacts. En ce qui concerne l'activité électrodermale, Boucsein (2012) a pu identifier trois types d'artefacts liés au mouvement⁴. Sur les signaux de rythme cardiaque, ceux-ci se traduisent généralement par des battements cardiaques manqués ou surdétectés à partir de la mesure ECG ou PPG brute (Acharya et al., 2006). De la même manière, toutes les mesures physiologiques retenues en section I.C sont sujettes à des altérations fortes en lien avec le mouvement.

Or, l'estimation automatique du stress en vie réelle nécessite de pouvoir dire si une fenêtre de signal est propre (ou non) au calcul de caractéristiques physiologiques. Lors de l'élaboration d'un détecteur de stress, on peut avoir recours à une inspection visuelle des signaux recueillis afin de retirer les parties artéfactées de chaque signal (Wijsman, 2014). Mais une telle approche est chronophage et ne peut être lors de l'application du modèle dans la vie quotidienne.

Aussi, le retrait automatique des artefacts liés au mouvement représente un domaine actif de la recherche en traitement du signal physiologique ; notamment sur le signal PPG (Periyasamy et al., 2017) et sur le signal ECG (Sweeney et al., 2013). Certaines études ont aussi tenté d'identifier des zones fortement altérées du signal de respiration ou de température de la peau (Plarre et al., 2010).

Même en l'absence d'artefacts, cela dit, la qualité d'une mesure physiologique peut être temporairement insuffisante pour en extraire une information fiable. C'est le cas notamment avec l'estimation du rythme cardiaque. Plusieurs études ont abordé ce sujet dans la littérature, en présentant un indice de qualité sur le signal PPG brut (Q. Li & Clifford, 2012; Orphanidou et al., 2015; Papini et al., 2017) ; ou même sur le signal ECG brut (Behar et al., 2012). A nouveau, ces indices peuvent être utilisées pour identifier les zones où le signal est altéré.

Pour garantir la performance de leurs algorithmes, les études précédentes ont eu massivement recours à l'apprentissage automatique. Ainsi, il n'existe pas de méthode de référence pour mesurer la qualité d'un signal physiologique quelconque. De même, lorsqu'une mesure de ce type est disponible, il n'existe pas de méthode de référence pour l'exploiter. Ainsi, les stratégies de gestion de la qualité du signal sont très hétérogènes dans la littérature. La section IV.C détaillera comment les tentatives

⁴ (i) Un changement dans la surface/pression de contact entre l'électrode et la peau, qui se traduit sur le signal par un changement brusque, persistant, du niveau moyen. (ii) Suite à une pression extérieure sur l'électrode, une modification brève et locale des propriétés mécaniques de la peau se traduit dans le signal AED par un pic bref et de forte amplitude (*onde d'Ebbecke*). (iii) Un décollement momentané de l'électrode se traduit par une remise à zéro du signal.

de détection du stress dans la vie quotidienne ont recueilli puis utilisé de l'information sur la qualité des mesures de leurs capteurs portables.

IV.B.2. Fiabilité des capteurs portables

Les capteurs portables utilisent des mesures alternatives à l'instrumentation standard, ce qui peut modifier certaines propriétés des signaux physiologiques obtenus.

L'électrode humide, par exemple, ne peut être utilisée hors du cadre médical ; car sa pose est délicate et elle nécessite un remplacement fréquent. Pour la mesure électrodermale, un capteur adapté à la vie réelle embarquera donc des électrodes sèches. Cette solution n'est pas sans impact sur l'impédance à l'interface de l'électrode : celle-ci dépend de la quantité de sueur qui sert d'électrolyte naturel au dispositif (Boucsein, 2012). Par ailleurs, certains dispositifs portables mesurent l'impédance avec un courant alternatif (c'est le cas d'Empatica E4). Dans ce cas, il faudrait tenir compte d'effets capacitifs non corrigés. Pour finir, l'électrode sèche est susceptible de se déplacer sur la peau (puisque contrairement aux électrodes utilisées en laboratoire, elle n'est pas fixée au moyen d'un adhésif). Ceci la rend très sensible aux artéfacts liés au mouvement.

En ce qui concerne la mesure cardiaque, un capteur portable embarquera généralement des mesures PPG au lieu de mesures ECG. Ces deux mesures ne sont pas exactement synchrones : il existe un temps de transit (*pulse transit time*) entre le cœur et le lieu de mesure, qui dépend notamment de la pression sanguine (Y. Li et al., 2014; Thomas et al., 2016). Les variations de ce temps de transit sont susceptibles d'affecter le signal de variabilité cardiaque mesuré par PPG : celui-ci serait similaire à l'ECG en période de repos, tout en montrant des différences significatives en situation de stress physique ou mental modéré (Schäfer & Vagedes, 2013).

Ces petites différences entre un signal issu d'un capteur portable et celui de l'instrumentation standard, peuvent engendrer des problèmes de portabilité des caractéristiques physiologiques d'un environnement de mesure à l'autre.

Ollander et collaborateurs (2016) ont cherché à quantifier ce phénomène, à l'aide d'un protocole expérimental réalisé en contexte de laboratoire, impliquant une instrumentation standard et un bracelet de mesures portable (Empatica E4). Les auteurs ont constaté une dissemblance nette entre les signaux AED recueillis à partir de leur dispositif de laboratoire (électrodes humides positionnées sur les doigts) et ceux d'Empatica E4 (électrodes sèches positionnées au poignet). Concernant le signal cardiaque, le nombre de battements manqués était bien plus grand sur les enregistrements d'E4 (issus de la PPG) que sur ceux de l'instrumentation standard (issus de l'ECG). Le signal PPG serait ainsi bien plus sensible aux artéfacts même avec peu de mouvement. En revanche, certaines caractéristiques temporelles comme la moyenne ou l'écart-type du rythme cardiaque ont pu être estimées correctement à partir du signal PPG. A notre connaissance, cependant, aucune étude de ce type n'a été menée sur des données enregistrées en vie réelle.

IV.C. Analyse des Tentatives de Détection du Stress dans la Vie Quotidienne

Bien qu'ils soient difficiles à mettre en œuvre, des protocoles d'estimation du stress dans la vie quotidienne apparaissent depuis peu dans la littérature. Les données récoltées peuvent servir soit à valider, soit à consolider un modèle de stress élaboré sur des données de laboratoire.

IV.C.1. Modèles publiés dans la littérature

Dans une expérience de laboratoire comprenant trois tâches standard (TSST, arithmétique et CPT, sans tâche contrôle), Hovsepian et collaborateurs (2015) ont appris un modèle par validation LOSO et mesuré une performance de 95% sur une cohorte de sujets différente de la base d'apprentissage. Lors d'une campagne de mesures sur le terrain, une version améliorée du même modèle a bien classé 72% des périodes de stress et de non-stress (sans activité physique) reportées par les participants.

Dans une expérience de laboratoire comprenant une seule tâche arithmétique (avec condition contrôle), Gjoreski et collaborateurs (2016) ont obtenu une précision de 83% en validation LOSO. Dans leur campagne de mesure en vie réelle, ce modèle de détection a obtenu une précision de 76%. Par la suite, les auteurs ont combiné ce modèle physiologique « de laboratoire » à un détecteur d'activité physique (développé dans une étude sans lien avec la détection du stress) ; et à de l'information de contexte (niveaux antérieurs d'activité physique, heure du jour, jour travaillé/chômé). Après un nouvel apprentissage sur les mesures en vie réelle, le détecteur final revendiquait une précision de 92% en validation LOSO. Les auteurs ont conclu sur la nécessité d'inclure des informations de contexte pour une meilleure détection du stress en vie réelle.

Dans une autre étude publiée un an plus tard (2017), ces mêmes auteurs ont utilisée la même approche en ajoutant une sélection de modèles pour chacun des deux détecteurs. Cette démarche a sélectionné un support à vastes marges pour le détecteur sans contexte et un arbre de décision pour le détecteur avec contexte. Les auteurs revendiquent 70% d'exemples Stress bien classés sur les données de terrain, le verdict « stress » étant juste dans 95% des cas.

Enfin, une publication très récente revendique 82% de bien classés dans un contexte ambulatoire (Han et al., 2020). En retirant manuellement les exemples contenant de l'activité, ces performances étaient de 100%. De notre point de vue, cependant, la méthode proposée dans cette étude est problématique sur plusieurs points. En particulier, la quantité de données de terrain disponibles était très insuffisante comparée aux autres études (340 minutes sur 1 sujet ici, contre +1000 heures sur 20 sujets pour Hovsepian et coll. (2015) et 55 jours sur 5 sujets pour Gjoreski et coll. (2017, 2016)). Ainsi, nous n'analyserons pas la performance revendiquée par ces auteurs dans la suite de cette section.

Les modèles comparables sont résumés dans le Tableau 3.8. Plus d'informations concernant la conception de ces modèles sur les données de laboratoire sont disponibles en Annexe 4. Pour pouvoir comparer aux autres la performance revendiquée Gjoreski et collaborateurs (2017), on a calculé le pourcentage de bien classés à partir des matrices de confusion dans un problème à deux classes. Lorsque la mention « données de test » suit une mesure de performance, cela signifie que les exemples ayant servi à cette mesure n'ont pas servi pendant l'apprentissage du modèle par validation croisée.

Tableau 3.8 : Tentatives de détection du stress dans la vie quotidienne

Étude	(Hovsepian et al., 2015)	(Gjoreski et al., 2016)	(Gjoreski et al., 2016)	(Gjoreski et al., 2017)
Mesures incluses	ECG, Respiration	AED, PPG, Temp. Peau, Accélération	AED, PPG, Temp. Peau, Accélération, Données de Contexte	
Tâches Stress	TSST, Tâche arithmétique, CPT	Tâche arithmétique		
Tâches Non-Stress	Repos	Contrôle		
Algorithme d'apprentissage	SVM + Réseau Bayésien	Forêt Aléatoire	Forêt Aléatoire (physio) + SVM (activité & contexte)	SVM (physio) + Arbre de Décision (activité & contexte)
Performance en Laboratoire	95% (Données de test)	83% (Validation LOSO)	83% (Validation LOSO)	85% (Validation LOSO)
Performance en Vie Réelle	72% (Données de test)	76% (Données de test)	92% (Validation LOSO)	99% (Validation LOSO)

Mesures – AED : Activité électrodermale. ECG : Electrocardiographie. EMG : Electromyographie. PPG : Photopléthysmographie. Resp. : Signal Respiratoire. Tâches – CPT : Cold Pressor Test. TSST : Trier Social Stress Test. Algorithmes – SVM : Séparateur à Vastes Marges.

Comme on pouvait s’y attendre, l’utilisation directe d’un modèle de laboratoire dans un contexte ambulatoire s’accompagne d’une baisse de la performance (colonnes 1 et 2). Pour Hovsepian et collaborateurs (2015), cette baisse est conséquente (-23%) ; mais on peut supposer que la forte précision obtenue en laboratoire (95%) était liée à l’absence de condition contrôle dans le protocole d’apprentissage. Par ailleurs, ce dernier modèle a été appliqué tel quel sur les données de la vie réelle ; ainsi, on peut supposer que la présence de bruit (même sans activité physique) a aussi contribué à la diminution des performances.

Les résultats de Gjoreski et collaborateurs (2016) sont similaires à ceux de Hovsepian et collaborateurs (2015) lorsque leur détecteur « de laboratoire » est appliqué *tel quel* dans la vie quotidienne. En revanche, l’utilisation de données de contexte semble avoir augmenté la performance mesurée sur les données de terrain (+16%). Dans le second article des mêmes auteurs (2017), la sélection de modèles sur ces mêmes données semble même proposer une classification quasi-parfaite en validation LOSO (99%).

Les travaux rapportés dans le Tableau 3.8 représentent une avancée significative pour la conception d’une mesure objective du stress mental. L’acquisition et le traitement de données expérimentales dans un milieu non contrôlé est un problème complexe et nécessite une préparation rigoureuse. Et en affichant un taux d’erreur aussi faible, les derniers travaux de Gjoreski et collaborateurs (2017) semblent avoir presque résolu le problème de la conception d’un détecteur de stress en vie réelle. Cependant, les performances revendiquées par ces modèles sont critiquables sur plusieurs aspects que nous détaillons dans le paragraphe suivant. En particulier, nous chercherons à montrer les trois points suivants.

- i) La performance revendiquée par les « meilleurs » modèles du Tableau 3.8 ne suffit pas à mesurer leur capacité à détecter le stress dans la vie quotidienne.
- ii) On peut questionner la validité de ces détecteurs de stress.
- iii) L’utilisation d’un modèle de type *boîte noire* compromet cette validité.

- iv) Les stratégies mises en place pour gérer la qualité des données ne suffisent pas à garantir la robustesse du modèle sur des données de la vie réelle.

IV.C.2. Sur la performance revendiquée par les détecteurs en vie réelle

Une analyse pertinente de la performance d'un modèle en vie réelle se trouve dans les résultats fournis par Gjoreski et collaborateurs (2016). La Figure 3.15 ci-contre affiche la matrice de confusion rapportée par les auteurs. Dans la vie quotidienne, la prévalence du *non-stress* dépasse largement celle du *stress*. Ainsi, la précision de 92% revendiquée par Gjoreski et collaborateurs (2016) repose sur 813 exemples de Non-Stress contre 114 exemples de Stress. La Figure 3.15 montre que le modèle « avec contexte » classe les exemples de la classe Stress avec une précision faible (55% : 5^e ligne, dernière colonne) ; et que le score final de performance ne tient pas compte de ce déséquilibre entre les classes. Si l'on considère la précision comme l'équilibre entre le taux de vrais positifs (55%) et le taux de vrais négatifs (97%), alors la performance de ce modèle est de 76% « avec contexte » et de 70% « sans contexte ». De ce point de vue, la marge de progression n'est plus aussi importante que celle revendiquée dans l'abstract de l'article.

	No Context		With context	
	0	12	0	12
No Str. (0)	638	175	790	23
Stress (12)	44	70	51	63
Rec. (%)	78	61	97	55
Prec. (%)	94	29	94	73
F1. (%)	85	39	96	63
Acc. (%)	76		92	

Figure 3.15 : Matrice de confusion des modèles proposés par Gjoreski et collaborateurs (2016) sur des mesures en vie réelle.

Modèles sans (à gauche) et avec (à droite) un détecteur de contexte. 0, 1 et 2 : niveaux de stress

Dans l'étude publiée un an plus tard, Gjoreski et collaborateurs (2017) ne donnent pas de pourcentage de bien classés et revendiquent simplement 70% de vrais positifs. Les auteurs évitent ainsi l'écueil précédemment décrit sur la mesure des performances, contrairement aux autres articles publiés sur le même sujet (Han et al., 2020; Hovsepian et al., 2015). En revanche, on peut dégager une différence essentielle entre le modèle « sans contexte » et les modèles « avec contexte » proposés par Gjoreski et collaborateurs (2017, 2016). Le modèle « sans contexte » (colonne 2 du Tableau 3.8) n'a bénéficié d'aucun apprentissage supplémentaire par rapport aux données de laboratoire ; tandis que les modèles « avec contexte » (colonnes 3 & 4 du Tableau 3.8) ont été appris sur les données de terrain.

Ainsi, les performances affichées dans le Tableau 3.8 ne sont pas toutes comparables entre elles ; et reflètent pauvrement la capacité du modèle à détecter du stress dans la vie réelle. La prévalence du *non-stress* en vie réelle empêche de tirer des conclusions pertinentes des performances affichées dans les colonnes 1, 2 et 3. L'absence de données de *test* dans les colonnes 3 et 4 empêche ces auteurs d'apporter une réponse forte sur la capacité de leur modèle à détecter du stress dans d'autres contextes que ceux contenus dans leur base de données.

IV.C.3. Sur la validité des détecteurs en vie réelle

La performance d'un détecteur de stress sur des données de terrain est une mesure de sa *validité écologique*. Mais comme on l'a vu dans le paragraphe précédent, la performance revendiquée par les détecteurs du Tableau 3.8 ne mesure pas fidèlement la capacité de ces modèles à fonctionner dans un contexte quelconque de la vie quotidienne. En revanche, la prise en compte de facteurs confondants durant le processus d'apprentissage augmente les capacités de généralisation des détecteurs à d'autres contextes. On peut donc considérer un second critère pour évaluer les détecteurs du Tableau 3.8 : la *validité du contenu*.

Dans les colonnes 1, 3 et 4 du Tableau 3.8, le couplage du détecteur de stress avec un détecteur d'activité physique préserve ces modèles de l'influence de ce facteur confondant. De même, les trois modèles de laboratoire ont été élaborés sur la base d'une validation croisée de type *LOSO*. Ainsi, l'activité physique et l'effet du sujet ont été pris en compte dans ces études ayant cherché à détecter le stress dans la vie quotidienne. En revanche, d'autres facteurs ont fait l'objet d'un traitement inégal, à l'image des études recensées en section III.B.

Dans l'étude de Hovsepien et collaborateurs (2015), l'apprentissage du détecteur n'a pas bénéficié d'une condition contrôle. De la même manière, l'apprentissage du modèle physiologique dans l'étude de Gjoreski et collaborateurs (2017, 2016) n'a pas bénéficié d'un protocole expérimental multitâche. Ainsi, il est possible que ces modèles soient inefficaces dans contextes différents de la base d'apprentissage. Ces carences ont pu participer à la chute de performance observée entre les données de laboratoire et les données de terrain dans les colonnes 1 et 2 du Tableau 3.8.

Dans les colonnes 3 et 4, les modèles ont été appris et validés sur un grand nombre de données de la vie quotidienne. Cependant, cette validation *LOSO* repose sur peu d'individus (5) ; et la mesure des performances peut être biaisée par la prépondérance d'un seul participant, qui a fourni l'essentiel des données (48 jours sur 55). Par conséquent, il est difficile de dire si les détecteurs proposés pourront généraliser à des sujets différents de ceux utilisés dans la base d'apprentissage.

Pour finir, on peut questionner la manière dont les données de contexte ont été insérées dans le processus d'apprentissage. Ces informations ont été introduites comme caractéristiques, au même titre que la sortie du modèle physiologique, dans un super-détecteur entraîné et validé sur les mesures en vie réelle. Pour un tel algorithme, une stratégie prometteuse était d'utiliser les données de contexte pour étiqueter comme Non-Stress les situations où le modèle physiologique détectait *à tort* un état de Stress. Ce comportement est confirmé sur la Figure 3.15 : le détecteur de contexte a diminué efficacement le nombre de faux positifs, qui est passé de 175 à 23.

Bien que cette stratégie semble payante, le modèle utilisé en vie réelle n'est plus le même que celui qui a été testé en laboratoire. Cela implique de vérifier si les données supplémentaires (heure du jour, jour travaillé/chômé, niveau des mesures physiologiques) n'ont pas introduit de nouveaux facteurs confondants dans le modèle de stress. Si notre raisonnement précédent est correct, ces facteurs pourraient biaiser le modèle vers la production de faux négatifs ; ce qui pourrait expliquer les faibles taux de vrais positifs (70%, 55%) observés dans les deux articles de Gjoreski et collaborateurs (2017, 2016). Cependant, la présence d'éventuels facteurs confondants dans le détecteur « avec contexte » n'est jamais évoquée dans les articles en question.

IV.C.4. Sur l'inconvénient des modèles de type boîte-noire

L'exemple précédent illustre la nécessité de vérifier la validité « du contenu » des modèles au moyen d'une approche en *boîte-blanche*. En effet, l'utilisation d'informations de contexte (comme la présence d'activité physique) semble indispensable pour préserver le modèle de facteurs confondants dans un milieu non contrôlé. Mais dans l'étude de Gjoreski et collaborateurs (2016), l'inclusion de ces données dans modèle de type *boîte-noire* (séparateur à vastes marges) empêche ces auteurs de répondre à la question suivante : dans quelle mesure le modèle détecte-t-il toujours du *stress* ?

Pour s'en convaincre, il suffit d'imaginer un modèle de stress qui ferait un tirage aléatoire (un lancer de pièce) sur l'état de stress ; et dont l'issue serait corrigée par un bon détecteur des contextes

non-stressants (voir plus bas). On obtiendrait alors des résultats similaires à ceux affichés dans la Figure 3.15 : un taux de vrais négatifs élevé (97%), avec un taux de vrais positifs proche du hasard (55%). Comme dans la colonne 3 du Tableau 3.8, on pourrait revendiquer une performance de 92% sans avoir besoin de mesures physiologiques pour détecter l'état de stress. Le modèle physiologique ayant été appris et validé sur peu de participants, une stratégie efficace pour l'algorithme d'apprentissage serait de repérer les habitudes de ces participants de sorte à obtenir un taux de vrais négatifs élevé (par exemple, avec des assertions du type « pas de stress entre 0h et 8h du matin »). Avec une telle stratégie, les performances du détecteur dans des contextes différents de la base d'apprentissage (sujets, habitudes) seraient alors médiocres.

Pour Gjoreski et collaborateurs (2016), la seule manière de vérifier que les données de contexte n'ont pas introduit de facteur confondant dans le détecteur de stress serait d'interpréter les paramètres du modèle. Or, ceci est impossible avec un séparateur à vastes marges.

Le deuxième détecteur proposé par ces auteurs (2017) utilise une approche « boîte-grise ». La sortie d'un modèle boîte-noire (séparateur à vastes marges pour détecter le stress physiologique) est insérée dans un modèle boîte-blanche (arbre de décision pour inclure les informations de contexte). Cependant, les auteurs ne décrivent pas le contenu du modèle boîte-blanche. Les données fournies ne permettent pas d'invalider l'hypothèse précédente : les données de contexte ont pu biaiser le détecteur vers la production de faux négatifs et diminuer ainsi taux de vrais positifs (70%).

Cette hypothèse n'est pas la seule explication possible aux résultats : par exemple, certains échantillons Stress ont pu s'accompagner d'une réponse physiologique insuffisante pour être bien classés. Mais on peut aussi supposer que l'apprentissage monotâche du détecteur physiologique ne lui a pas permis de reconnaître d'autres situations de stress de la vie quotidienne. A nouveau, une réponse précise à ces questions nécessiterait d'interpréter les paramètres du détecteur de stress. Ceci est impossible avec le séparateur à vastes marges. Ainsi, un apprentissage en *boîte-noire* ne permet pas au concepteur de garantir la validité du détecteur de stress en vie réelle.

IV.C.5. Sur les stratégies de gestion de la qualité des données

Pour finir, les études considérées dans cette section ont traité différemment la question des artefacts dans les signaux enregistrés. Hovsepian et collaborateurs (2015), ont appliqué un algorithme de détection automatique des artefacts, développé parallèlement à leur étude. Cet algorithme s'appuyait sur les propriétés des signaux ECG et de Respiration pour identifier les épisodes où le capteur (une ceinture thoracique) se déplaçait sur la peau (Plarre et al., 2010). Les intervalles temporels sur lesquels un signal présentait un artefact ont été retirées de la base de données, afin de préserver le détecteur de stress d'une grande partie des données artefactées.

Dans les deux articles de Gjoreski et collaborateurs, un seul (2017) fait mention d'une stratégie de gestion de la qualité des données. Celle-ci concerne seulement la mesure PPG. Les auteurs n'ont extrait des caractéristiques de variabilité cardiaque (HRV) que lorsque le signal d'intervalle inter-battments (IBI) présentait plus de 32 échantillons successifs ; afin de s'assurer les calculs ne soient pas biaisés par des échantillons manquants. En effet, le bracelet Empatica E4 ne renvoie un signal IBI que lorsque le signal PPG présente une qualité suffisante (Empatica, 2019). Sur une fenêtre de temps quelconque (10min dans l'algorithme final), la valeur retenue pour chaque caractéristique HRV correspondait à une moyenne des valeurs extraites sur 32 échantillons IBI successifs (*i.e.*, séparés de moins de 2 secondes).

Dans les deux études, la gestion des artefacts laissait deux options. Soit les auteurs acceptaient la présence de mesures dégradées dans le calcul des caractéristiques : c'est le cas pour Gjoreski et collaborateurs (2017, 2016) en ce qui concerne l'activité électrodermale. Soit ils n'acceptaient pas ces mesures dégradées : c'est le cas pour les deux études en ce qui concerne les signaux de rythme cardiaque. Dans le premier cas, on risque de calculer la sortie du modèle sur des caractéristiques peu *fiables*. Dans le second cas, la présence d'un artefact sur une seule mesure implique de retirer de l'analyse toute la période concernée, au risque d'un problème de *disponibilité* des données physiologiques. Aucun des trois articles ne fournit d'information sur cette quantité de données indisponibles à la détection du stress. L'article de Hovsepien et collaborateurs (2015) laisse entendre que ce phénomène n'était pas rare.

La nécessité de retirer du calcul des données partiellement bruitées vient du fait que les modèles standard ont besoin de toutes leurs caractéristiques pour aboutir à une décision finale sur l'état de stress. De même, un modèle standard suppose que la valeur des caractéristiques est fiable. Mais dans un modèle de stress multi-mesures, la quantité d'artefacts rencontrés est d'autant plus importante que les mesures physiologiques sont nombreuses. Ainsi des perspectives de recherche se trouvent dans la conception d'un modèle de stress *robuste* aux données de la vie réelle, c'est-à-dire performant même lorsque les conditions de mesure sont dégradées.

V. Positionnement de la Thèse

V.A. Principes et Limites d'une Méthode Standard pour l'Estimation du Stress en Vie Réelle

V.A.1. Détection du stress à partir de signaux physiologiques : un bref historique

Parmi les études présentées dans ce chapitre, deux nous semblent résumer l'histoire moderne de la modélisation du stress. Avec l'idée de détecter le stress du conducteur, Healey et Picard (2005) ont recueilli puis analysé les données de plusieurs personnes en situation de repos (non-stress), de conduite sur autoroute (stress faible) et de conduite en ville (stress fort). Un processus impliquant toutes les étapes de l'apprentissage automatique a permis de distinguer les trois conditions expérimentales avec 97% d'exemples bien classés (100% si l'on fusionnait les deux classes de stress). Avec cette expérience débute une décennie de tentatives de détection du stress, dans un contexte plus ou moins contrôlé, impliquant un protocole expérimental plus ou moins complexe et aboutissant à un modèle plus ou moins interprétable du stress mental.

Depuis le milieu des années 2010, la perspective d'identifier les épisodes de stress dans la vie quotidienne se concrétise avec l'essor des capteurs physiologiques portables. Les corrélats du stress sont désormais connus : on les mesure au travers de caractéristiques désormais standard, incorporées par défaut dans les algorithmes d'apprentissage. Les signaux physiologiques les plus utilisés dans ce cadre sont le rythme cardiaque et l'activité électrodermale. Ils sont mesurables en vie réelle au moyen de dispositifs connectés, dont certains (comme le bracelet Empatica E4) font désormais référence dans les études de terrain. Récemment, Gjoreski et collaborateurs (2017) ont proposé un détecteur de

stress qui combine la sortie d'un modèle physiologique et celle d'un détecteur d'activité, avec des données de contexte, dans un détecteur unique appris et validé sur des données de terrain. Sur ces mêmes données, le détecteur en question est parvenu à identifier 70% des épisodes de stress et 99% des épisodes de non-stress.

Actuellement, ces résultats sont les plus concluants parmi les études ayant cherché à détecter le stress dans un environnement non contrôlé. Bien qu'ils représentent un pas significatif vers l'estimation du stress dans la vie quotidienne, certains choix méthodologiques sont critiquables au vu des éléments développés dans ce chapitre.

V.A.2. Performance, validité et robustesse d'un modèle de stress

Comme une majorité de détecteurs dans la littérature, le modèle précédent a été appris et validé de sorte à maximiser un score de *performance* sur les données de terrain. Cet impératif de performance a justifié l'utilisation d'un modèle de type *boîte-noire* pour détecter le stress sur la base des mesures physiologiques.

Ce type de modèle a l'avantage de proposer une frontière de décision peu contrainte entre les classes Stress et Non-Stress. Durant la validation croisée, cette frontière peut être ajustée de sorte à maximiser la performance obtenue sur les données disponibles. Cela dit, notre analyse de l'état de l'art n'a pas trouvé d'élément qui montre que le recours à un modèle de type *boîte-noire* soit plus efficace que le recours à un modèle de type *boîte-blanche* pour détecter le stress en conditions contrôlées. On peut ainsi supposer qu'une frontière de décision complexe n'est pas nécessaire pour identifier la réponse physiologique au stress dans l'espace des caractéristiques connues. En revanche, un apprentissage en *boîte-noire* ne permet pas d'interpréter les paramètres du détecteur de stress. Dans un contexte ambulatoire, cette ignorance est problématique dans la mesure où la réponse physiologique au stress est sujette à des *facteurs confondants* puissants.

Dans le Chapitre Deux - III, notre recherche d'une définition du stress mental nous a permis d'identifier certains facteurs confondants physiques (*e.g.* : l'activité physique, l'heure du jour), qui peuvent être pris en compte à l'aide de données non physiologiques (*e.g.* accéléromètre, horloge numérique). Mais aussi des facteurs confondants psychologiques (*e.g.* : l'effort mental faible, le type de stressor), qui ne sont pas directement mesurables et doivent être intégrés dans le modèle physiologique utilisé pour détecter le stress.

La prise en compte de ces facteurs permet de garantir le bon comportement du modèle dans un environnement différent de la base d'apprentissage ; par exemple, sur des tâches qui n'ont pas déjà été réalisées par le sujet, ce qui est indispensable pour un modèle prévu pour fonctionner en vie réelle. Comme on l'a constaté en section III.B, cette augmentation des capacités de généralisation s'accompagne d'une diminution de la performance mesurée par validation croisée. Ainsi, la mesure de performance à elle seule ne suffit pas à produire un modèle *valide* du stress mental. On augmente la validité d'un modèle en tenant compte de facteurs confondants dans le processus d'apprentissage. Comme on l'a vu en section IV.C, cette validité est difficile à garantir avec un modèle *boîte noire*.

Qui plus est, la mesure ambulatoire est sujette à de nombreux artéfacts qui diminuent la performance des modèles élaborés en laboratoire. La présence de ces artéfacts introduit du bruit dans les caractéristiques physiologiques. Ce bruit peut poser un problème de *fiabilité* ou de *disponibilité* des caractéristiques, selon que l'on accepte ou que l'on rejette la valeur bruitée. La plupart des modèles

de stress reposant sur des propriétés fines du signal physiologique (comme la variabilité cardiaque haute fréquence), une erreur dans le calcul de ces caractéristiques peut compromettre l'ensemble du processus de décision. Dans ce contexte, la littérature manque d'un modèle *robuste* qui inclut une stratégie de gestion de la qualité des données dans ce processus de décision. A nouveau, cette propriété est difficile à garantir au moyen d'un modèle de type boîte-noire.

Ainsi, notre démarche de conception d'un détecteur de stress devra dépasser la simple recherche de performance et proposer une réponse construite aux deux questions suivantes :

- Dans quelle mesure pouvons-nous garantir que notre détection du stress n'est pas soumise à des facteurs confondants ?
- Dans quelle mesure pouvons-nous garantir la performance du modèle face à des mesures de terrain bruitées, voire indisponibles ?

La réponse à la première question détermine la *validité* du modèle de stress. La réponse à la deuxième question détermine sa *robustesse*. A notre connaissance, aucune approche de la littérature n'affiche ces deux ambitions simultanément.

V.B. Questions de Recherche

Nous cherchons à proposer un modèle valide du stress mental, qui puisse résister aux aléas de la mesure en condition ambulatoire. Dans cette démarche, nous orienterons nos recherches vers des approches en boîte-blanche, depuis la sélection des caractéristiques jusqu'au processus de fusion des données, de manière à construire des modèles dont les décisions soient interprétables. Cette transparence permettra de valider, voire d'adapter nos méthodes sur des bases de données différentes de la nôtre.

V.B.1. Sélection de caractéristiques sensibles à la charge affective

Dans l'état actuel des connaissances sur le stress mental, le recours à une campagne expérimentale en milieu contrôlé est toujours nécessaire pour cibler la charge affective dans la réponse physiologique associée. Pour permettre une recherche efficace des marqueurs du stress, le protocole expérimental doit permettre de contrôler les trois facteurs confondants introduits au Chapitre Deux - IV.C : l'effet de l'activité, l'effet de la tâche et l'effet du sujet.

Au demeurant, il n'existe pas de base de données libre d'accès qui permettrait de traiter la détection du stress sous cet angle particulier. C'est dans cette perspective qu'une campagne de mesures en laboratoire a été réalisée collaborativement par le LPNC, le GIPSA-Lab et le CEA-Leti – et tout juste achevée au commencement de ces travaux de thèse. Le Chapitre Quatre - détaille le déroulement de la campagne expérimentale, ainsi que l'extraction des caractéristiques qui ont servi de support aux Chapitre Cinq - et Six.

Les facteurs confondants du stress font l'objet d'un traitement très hétérogène dans la littérature. Dans la section III.B, notre analyse des protocoles expérimentaux nous a conduits à mettre en doute la validité des caractéristiques fréquemment incluses dans les modèles de stress.

Dans ce contexte, le Chapitre Cinq - présente un travail de sélection des caractéristiques basé sur l'estimation de la charge affective en situation de stress. Dans ce travail, on considère qu'une caractéristique est *sensible* au stress si elle présente une charge affective en situation de stress. Cela

revient à dire qu'elle réagit davantage en situation de stress qu'en situation d'activité. De même, on considère qu'une caractéristique est *spécifique* au stress si elle réagit peu aux facteurs confondants.

Notre recherche de caractéristiques sensibles et spécifiques au stress nous a conduits à rejeter la plupart des caractéristiques fréquemment citées dans la littérature. Cette étude montre la nécessité de contrôler *l'effet d'activité* lors de la conception d'un modèle de stress. A l'issue de la sélection, quatre caractéristiques ont été retenues pour être incluses dans un nouveau modèle de stress.

V.B.2. Modèle adaptatif du stress émotionnel

La conception de ce modèle est l'objet du Chapitre Six - ..

En vie réelle, une caractéristique peut se montrer brusquement indisponible pour l'estimation de l'état de stress, du fait des artefacts perturbant fréquemment le signal physiologique. Pour un modèle standard utilisant peu de caractéristiques, ce phénomène peut compromettre totalement l'estimation du stress sur une période prolongée. La qualité des données représente ainsi un verrou majeur dans la conception *boîte-blanche* d'un modèle de stress.

De notre point de vue, la robustesse d'un modèle à des données bruitées peut être améliorée de deux manières :

- D'une part, en limitant la contribution de chaque caractéristique sur la sortie du modèle. De cette manière, une caractéristique artéfactée aura moins d'impact dans le processus de détection du stress.
- D'autre part, en retirant simplement du calcul, chaque caractéristique rendue durablement indisponible par un signal physiologique de faible qualité. De cette manière, le modèle pourra continuer de fonctionner sur la base des caractéristiques restantes.

Ces objectifs appellent à changer de paradigme dans la conception du modèle de stress. Plutôt que de définir une frontière optimale dans l'espace des caractéristiques, nous proposons un modèle qui fusionne différents indices mono-caractéristiques du stress. L'idée principale est d'utiliser des règles génériques pour fusionner ces indices (par exemple une simple moyenne) afin de pouvoir modifier en temps réel l'ensemble de caractéristiques sur lequel on détermine l'état de stress.

Le Chapitre Six - présente une preuve de concept en mesurant la performance de notre modèle sur des données volontairement bruitées.

V.B.3. Algorithme d'estimation en ligne de la qualité d'un signal cardiaque

La mise en œuvre d'un algorithme de détection du stress capable de s'adapter en temps réel à la faible qualité des mesures physiologiques nécessite le développement d'un indicateur de qualité. Pour traiter cette problématique, on s'est concentré sur les signaux cardiaques. Le Chapitre Sept - propose une méthode pour évaluer en ligne la qualité d'un signal de rythme cardiaque, ainsi qu'un critère pour déterminer si une fenêtre de temps présente un signal « sans défaut ». On s'appuiera pour cela sur une nouvelle base de données recueillie dans le cadre de nos travaux de thèse, et qui se déroule dans la vie quotidienne des participants.

Nos résultats sont à la base de l'élaboration d'un critère de *disponibilité* en vie réelle pour chaque caractéristique cardiaque sélectionnée dans le Chapitre Cinq - et utilisée dans le modèle du Chapitre Six - . Un tel critère peut permettre de modifier en temps réel l'ensemble de caractéristiques sur lequel

est calculé notre indice de stress. Cette particularité adaptative peut constituer un avantage décisif sur les modèles standards pour l'estimation du stress en condition ambulatoire.

A RETENIR

- Durant l'apprentissage d'un détecteur de stress, la prise en compte de facteurs confondants fait l'objet d'un traitement inégal, voire incomplet dans la littérature.
- De même, il n'existe pas de stratégie standard pour gérer la qualité fluctuante des mesures physiologiques en vie réelle.
- Nos travaux de thèse apportent trois contributions principales à l'état de l'art :
 - Une sélection de caractéristiques basée sur la prise en compte de facteurs confondants, dont l'effet de l'activité,
 - Un détecteur boîte-blanc du stress qui résiste à des données bruitées et qui s'adapte à un ensemble variable de caractéristiques,
 - Un indicateur de la qualité du signal de rythme cardiaque qui permet de proposer un critère de disponibilité sur les caractéristiques associées à ce signal.

Chapitre Quatre - Base de Données pour l'Étude du Stress en Conditions Contrôlées

Au commencement de toute recherche en sciences affectives se trouve une base de données expérimentale dédiée à l'étude du concept cible. Le Chapitre Deux - IV nous a donné l'occasion de lister les principales méthodes qui permettent l'enregistrement de mesures en lien avec le stress mental. La plupart d'entre elles sont appliquées dans ce chapitre.

Les paragraphes qui suivent retracent la création de la base de données qui servira de support aux analyses du Chapitre Cinq - et du Chapitre Six - . Parmi les étapes qui ont marqué ce processus, notre contribution se situe principalement dans l'extraction des caractéristiques physiologiques. Si nous avons pu assister à certaines sessions expérimentales, la majorité des données étaient déjà recueillies et dépouillées au début de ces travaux de thèse. Les quelques pages qui suivent se concentreront donc sur l'essentiel : présenter les conditions expérimentales et les dispositifs de mesures qui ont donné naissance à la base de données que nous appellerons : Laboratoire.

La section I de ce chapitre détaille le protocole sur lequel s'est appuyée la campagne expérimentale. La section II détaille les mesures utilisés dans la suite du manuscrit, depuis l'acquisition des signaux jusqu'au calcul des caractéristiques (dont l'implémentation est proposée en Annexe 5).

I. Protocole et Campagne d'Expériences

I.A. Cadre Expérimental et Recrutement des Participants

La campagne expérimentale s'est déroulée entre l'été 2015 et l'été 2016 dans les locaux de la plateforme SCREEN de la Maison des Sciences et de l'Homme, à Saint Martin d'Hères (38400). Le protocole a fait l'objet d'une validation éthique auprès du CERNI (Comité d'Évaluation des Recherches Non Interventionnels, N° 2015-05-12-67) de l'université Grenoble Alpes et a été réalisé selon les principes d'Helsinki.

Vingt participants jeunes ont été recrutés par diffusion d'annonce, puis sélectionnés selon des critères usuels pour les protocoles impliquant le langage. Ceux-ci parlaient couramment Français. Sur déclaration, leur vue et leur audition étaient normales ou corrigées à la normale. Ils ne présentaient aucun antécédent de trouble psychologique ou neurologique actuel ou passé, et ne suivaient aucun traitement médicamenteux susceptible de moduler leurs processus cognitifs et émotionnels.

Lors de la première session, les participants ont signé un consentement éclairé et ont été informés sur leur droit à se retirer à tout moment. En revanche, aucun participant n'a été informé du but réel de la campagne, et ce jusqu'à la fin des expériences (ceci afin d'éviter un stress par anticipation des tâches et les biais expérimentaux qui en découlent). Le motif officiel concernait la recherche de marqueurs de la charge mentale ; et dans cette version, les performances de chaque participant seraient systématiquement comparées à celles des autres. Lors d'un débriefing à la fin des quatre

sessions expérimentales, chaque participant a reçu une indemnité financière et a été informé des objectifs réels de l'étude.

I.B. Tâches Expérimentales

Le protocole s'articule autour des quatre tâches standard pour l'étude du stress, dont trois ont été présentées dans le Chapitre Deux - IV.B (Page 27).

- La tâche **TSST** (*Trier Social Stress Test*) qui consiste en une simulation d'entretien oral, telle que présentée par Kirschbaum, Pirke, & Hellhammer (1993).
- La tâche arithmétique **TA**, qui correspond à la tâche classiquement réalisée dans la littérature à la suite de la tâche TSST.
- Le **SECPT** (*Socially Evaluated Cold Pressor Test*) a été intégré dans le protocole tel que présenté par Schwabe, Haddad, & Schachinger (2008), afin d'étudier l'effet particulier d'une tâche à coping passif.
- La tâche d'attention visuelle **D2** a été introduite afin de mesurer l'effet *stressant* de la pression temporelle (Brickenkamp & Zillmer, 1998).

Cette dernière tâche est similaire au test de Stroop, dont l'efficacité pour induire du stress ne fait pas consensus dans la littérature. La recherche sur le stress en situation d'attention sélective pourrait s'enrichir d'une tâche qui n'a pas encore été utilisée dans cet objectif, et la tâche D2 propose à la fois un stressor (temps + évaluation sociale) et de la charge mentale.

Ces quatre tâches mobilisent les ressources cognitives de la personne et peuvent simuler des conditions usuelles de stress. Elles font appel à des compétences complexes comme la génération de parole ou le calcul mental. Chacune d'elles était déclinée dans une version *Stress (S)* et une version *Contrôle (C)* :

1. La version Stress était calibrée pour maximiser le stress de l'individu.
2. La version Contrôle était calibrée pour minimiser le stress de l'individu, dans un cadre qui le contraint à engager les mêmes processus cognitifs et moteurs que dans la version Stress.

Le Tableau 4.1 (plus bas) résume le déroulement des huit conditions expérimentales proposées.

Tableau 4.1: Résumé des tâches de la campagne expérimentale en laboratoire

	Tâche d'attention visuelle D2	<i>Socially Evaluated Cold Pressor Test</i>	<i>Trier Social Stress Test</i>	Tâche Arithmétique
VERSION STRESS	D2_s	SECPT_s	TSST_s	TA_s
Principe	Identifier une association de symboles (la lettre « d » suivie de deux apostrophes placées dans des configurations différentes), au sein d'un texte constitué de symboles similaires. <i>Temps limite</i> : 20 secondes par ligne, sur 14 lignes composées de 47 caractères.	Laisser sa main dominante dans un bain d'eau très froide (0°C à 4°C) pendant 3 minutes. La réaction du participant serait comparée à celle des autres.	Se mettre dans la peau de candidat au job de ses rêves. Un discours de 5 minutes pour convaincre le jury, sans notes, après 5 minutes de préparation. Si le <i>pitch</i> était trop court, le participant se heurtait à un long silence, suivi d'une relance, puis de questions de la part de l'expérimentateur.	A partir de 1022, compter à l'envers de 13 en 13. En cas d'erreur, on invitait sèchement le participant à recommencer. L'exercice devait être terminé le plus rapidement possible (performance comparée aux autres participants), ou bien s'arrêtait au bout de 5 minutes.
Environnement	Durant l'exercice, les participants étaient chronométrés par l'expérimentateur et surveillés par au moins une autre personne présentée comme experte des comportements non-verbaux.	Environnement identique à D2 _s . De plus, la réaction de chaque participant était filmée par une caméra mise en évidence sur la table d'expérience.	Environnement identique à SECPT _s .	Environnement identique à SECPT _s .
Principaux Stresseurs	Pression sociale, Pression temporelle, Attention sélective	Pression sociale, Souffrance physique	Pression sociale, Introspection, Charge mentale	Pression sociale, Pression temporelle, Effort mental
VERSION CONTROLE	D2_c	SECPT_c	TSST_c	TA_c
Principe & Environnement	La tâche D2 effectuée seul, sans surveillance et sans temps limité.	Laisser sa main dominante pendant 3 minutes dans un bain d'eau tempérée (37°C).	Lire un texte à voix haute, seul et sans surveillance, pendant 5 minutes.	Compter à l'envers de 100 à 0 (par pas de 1), puis recommencer. Seul et sans surveillance pendant 5 minutes.

I.C. Déroulement Pratique de l'Expérience

Pour réaliser chacune de ces tâches, le protocole se divisait en quatre sessions expérimentales. Chaque session était effectuée un jour différent (mais sur des créneaux horaires similaires d'un jour à l'autre), dans un ordre aléatoire. La répartition des tâches entre les 4 sessions expérimentales a été réalisée de sorte à séparer dans le temps les tâches Contrôle (**D2_c**, **SECPT_c**, **TSST_c**, **TA_c**) des tâches Stress (**D2_s**, **SECPT_s**, **TSST_s**, **TA_s**), ainsi que les tâches Stress entre elles. Ceci afin de permettre une récupération physiologique complète entre ces différentes conditions.

Seule exception à ces règles : la séquence TSST-TA. Suivant la méthode de Kirschbaum, Pirke, & Hellhammer, la tâche *entretien* (**TSST_s**) et la tâche *arithmétique* (**TA_s**) étaient toujours effectuées à la suite et dans cet ordre. En effet, le succès de la tâche TA_s repose en partie sur le risque que le

participant commette une erreur de calcul. Lors de la session Stress, la charge affective du participant à l'issue de la partie TSST_S augmentait sa probabilité d'erreur durant la partie TA_S, élicitant ainsi du stress supplémentaire. Par souci de similarité, l'enchaînement TSST-TA a été conservé tel quel dans la session Contrôle.

Le protocole expérimental a donc consisté en une session avec les tâches TSST-TA, une session avec la tâche D2, une session avec la tâche SECPT et une session avec toutes les tâches contrôles. L'ordre des sessions a été randomisé entre les participants. Enfin, chaque tâche était immédiatement précédée d'une période de Repos de 5 minutes : en amont de la version Stress (R_S) ; et en amont de la version Contrôle (R_C). La Figure 4.1 ci-dessous résume ainsi l'enchaînement des tâches du point de vue d'un participant quelconque.

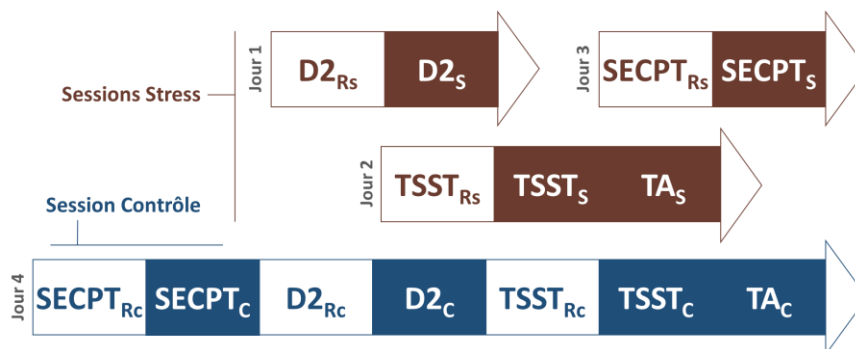


Figure 4.1 : Sessions expérimentales du point de vue d'un participant

I.D. Evaluations Psychologiques

A l'issue de chacune des 14 conditions expérimentales visibles sur la Figure 4.1, (**D2_{S-C-Rs-Rc}** ; **SECPT_{S-C-Rs-Rc}** ; **TSST_{S-C-Rs-Rc}** ; **TA_{S-C}**), chaque participant remplissait deux questionnaires destinés à évaluer sa charge affective. Ces questionnaires ont été présentés en section Chapitre Deux - IV.A (Page 27).

- Le questionnaire STAI-Y-A en version abrégée (20 items), qui évalue le niveau d'anxiété du participant à l'instant t ;
- Un ensemble de 5 échelles visuelles analogues (EVA) portant sur chacun des sentiments suivants : Calme/Apaisé, Joyeux/Heureux, Effrayé/Apeuré, Confus/Perturbé et Stressé (de *nul* à *très élevé*).

Bien qu'ils ne soient pas analysés dans le cadre de ce chapitre, les quatre items autres que « Stressé » permettaient à la fois de ne pas révéler au participant le concept cible du protocole expérimental et de mieux caractériser l'état émotionnel du participant.

II. Enregistrements Physiologiques

II.A. Signaux Recueillis

II.A.1. Disposition des capteurs

Chaque participant était enregistré avec le même dispositif de mesures physiologiques, constitué d'une instrumentation standard de laboratoire. Le configuration complète du dispositif de mesures est

illustrée dans la Figure 4.2. L'instrumentation comprenait une mesure d'Electrocardiographie (ECG), de conductance de la peau (SC, pour *Skin Conductance*), et de Respiration. Le signal ECG a été recueilli à l'aide d'un système à trois électrodes (sur le torse dans la configuration présentée sur la Figure 4.2), et amplifiée par un dispositif AD Instruments FE132. La conductance de la peau a été mesurée entre deux électrodes Ag/AgCl (AD Instruments MLT 117F). Celles-ci étaient positionnées sur les phalanges médianes de l'index et du majeur de la main non-dominante, et connectées à un amplificateur AD Instruments FE116. Le signal respiratoire a été mesuré par pléthysmographie respiratoire à variation d'inductance, au moyen d'une ceinture AD Instruments TN1132/ST. Tous ces signaux ont été échantillonnés à 1kHz, synchronisés par le système d'acquisition PowerLab 16/35, affichés et recueillis à l'aide du logiciel d'acquisition LabChart.

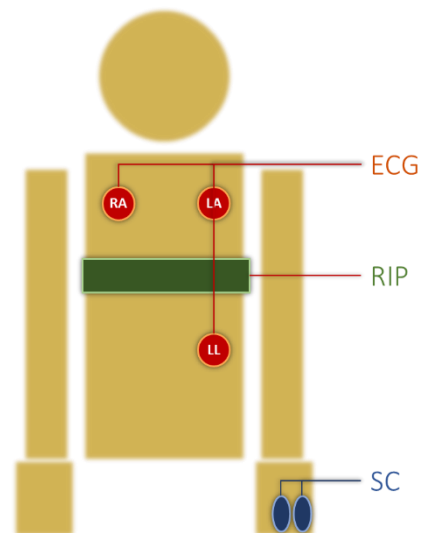


Figure 4.2 : Configuration du dispositif de mesures physiologiques.

ECG – électrocardiogramme – RIP : ceinture respiratoire – SC : conductance de la peau – E4 : bracelet de mesures portable

II.A.2. Données disponibles

Les quatre tâches standard $i \in \{D2, SECPT, TA, TSST\}$ ont comprises chacune les quatre conditions expérimentales $j \in \{Rc \text{ – repos avant la tâche contrôle, } Rs \text{ – repos avant la tâche stressante, } C \text{ – tâche contrôle, } S \text{ – tâche stressante}\}$; à l'exception des Repos pour la tâche TA qui a suivi systématiquement le TSST conformément à la littérature (voir Figure 4.1). Cela dit, le court intervalle (5 minutes) séparant le Repos ($TSST_{Rc/Rs}$) de la tâche $TA_{C/S}$ nous permet de supposer qu'une ligne de base valide pour le TSST est aussi valide pour la TA. Suivant cette hypothèse, on a doublé les enregistrements de $TSST_{Rc,Rs}$ pour obtenir une base de donnée complète à 4 tâches * 4 conditions. Ainsi, chaque mesure qui a été recueillie à chaque session expérimentale comprend au total $4*4*20$ enregistrements (20 étant le nombre total de sujets enregistrés). C'est notamment le cas pour les capteurs d'ECG et de SC. Pour des raisons de disponibilité du matériel, cependant, le dispositif de mesures respiratoires n'a pu équiper que 11 sujets sur 20 (dans toutes les tâches et toutes les conditions expérimentales).

En ce qui concerne les évaluations psychologiques, on dispose également de $4*4*20$ mesures pour les échelles visuelles analogiques. Cependant, certaines données sont manquantes sur les questionnaires STAI qui ont été ignorés par certains participants durant les sections contrôle. En fonction des conditions expérimentales étudiées, il faut ainsi s'attendre à un nombre variable d'exemples pour cette évaluation psychologique (voir Chapitre Cinq -).

II.B. Caractéristiques Extraites

Au total, 32 caractéristiques physiologiques ont été extraites des signaux de la base de données décrite dans cette partie. Ces caractéristiques sont listées dans la dernière colonne du Tableau 4.2 ci-dessous, avec leur mode de calcul et les signaux correspondants rappelés dans les colonnes adjacentes. La terminologie en Anglais a été préférée ici pour faciliter la comparaison avec la littérature, comme dans (Vila, Godin, Charbonnier, et al., 2019). Pour alléger la lecture de ce chapitre, le détail des calculs et le lien éventuel de ces caractéristiques avec le stress est proposé en Annexe 5.

Tableau 4.2 : Résumé des caractéristiques retenues dans le cadre de ce chapitre

Mesure	Signal d'intérêt	Calcul	Référence	
Rythme Cardiaque	Signal	Moyenne, Ecart-type	hrMean, hrStd	
	Différences Successives	Racine des carrés moyens (RMS)	hrRmssd	
Intervalle Inter-Battements	Densité Spectrale de Puissance	Puissance totale ou normalisée dans les fréquences Basses: [0.04; 0.15]Hz et Hautes: [0.15; 0.5]Hz	ibiLF, ibiLFnorm, ibiHF, ibiHFnorm	
		Rapport (ibiLF / ibiHF)	ibiLF/HF	
	Différences Successives	Racine des carrés moyens (RMS)	ibiRmssd	
Respiration	Signal	Intervalle, Ecart-type	respRange, respStd	
	Autocorrélation	Aire sous la courbe de la 1 ^{ère} seconde des retards	respAutocorr	
	Densité Spectrale de Puissance	Fréquence à l'amplitude maximale	respRate	
AED tonique	Signal	Moyenne	sclMean	
	Dérivée Première	Moyenne, RMS	sclDerMean, sclDerRMS	
Réponses électro-dermales	Courbe	Aire totale sous la courbe des réponses	scrArea	
	Amplitudes	Médiane et 85 ^e centile, Ecart-type	scrAmpP50, scrAmpP85, scrAmpStd	
	Temps de montée	Moyenne, Ecart-type	scrRiseMean, scrRiseStd	
	Temps de récupération	Moyenne, Ecart-type	scrRecMean, scrRecStd	
	Fréquence globale	Nombre de réponses sur la fenêtre		scrFreq
		Nombre de SCRs classés par intervalle d'amplitude: [0.02; 0.1[μ S, [0.1; 0.3[μ S, [0.3; 0.8[μ S et \geq 0.8 μ S		scrWeak, scrMild, scrStrong, scrGreat
	Rythme instantané	Médiane et 85 ^e centile, Ecart-type	scrrP50, scrrP85, scrrStd	

Parmi ces caractéristiques, certaines sont fréquemment rencontrées dans l'état de l'art (e.g. : hrMean, ibiLF, ibiHF, ibiRmssd, ibiLF/HF, sclMean, scrFreq). Certaines ont été citées dans la littérature récente sur la modélisation du stress (e.g. : hrStd, hrRmssd, sclDerRMS, scrAmpP50, scrrP85, respRate). Enfin, certaines ont été introduites dans un but exploratoire, sur la base de la littérature en physiologie du stress (e.g. : scrRiseMean, scrRecMean, scrWeak), ou bien de travaux antérieurs non publiés sur cette base de données (e.g. : respAutocorr).

Ces caractéristiques ont été calculées sur l'ensemble des enregistrements de la base de données (sur les 4 tâches tâches {D2, SCEPT, TSST, TA}, les 4 conditions {Rc, Rs, C, S} et les 20 participants). Bien que de même ordre de grandeur, les durées des tâches n'étaient pas égales entre elles. La plus courte était la SCEPT_s, qui durait théoriquement 3 minutes, mais certains enregistrements avaient une durée inférieure (environ 1 minute, les participants pouvant se retirer à n'importe quel moment du fait que l'expérience pouvait se montrer douloureuse). Pour cette raison, toutes nos caractéristiques ont été calculées sur des fenêtres successives de 60 secondes, sans recouvrement. De cette manière, chaque enregistrement $k \in \{D2, SCEPT, TSST, TA\} \times \{Rc, Rs, C, S\} \times \{1..20\}$ disposait de 1 à 5 échantillons $f_k(n)$, $n \in \{1..5\}$ calculés indépendamment pour chacune des 32 caractéristiques.

Dans le cadre de ce manuscrit, la base de données ainsi constituée a servi de support à la sélection de caractéristiques sensibles à la charge affective (Chapitre Cinq -) et à la conception de modèles physiologiques du stress émotionnel (Chapitre Six -). De même, le calcul des caractéristiques cardiaques proposé en Annexe 5 sera réemployé pour estimer la qualité des mesures extraites d'un capteur portable dans la vie quotidienne (Chapitre Sept -).

A RETENIR

- Pour les deux chapitres à venir, on dispose d'une base de données constituée de 32 caractéristiques calculées sur 11 à 20 sujets (selon la mesure), pour 4 tâches expérimentales (D2, SECPT, TSST, TA) enregistrées dans 4 conditions différentes (1 Stress, 1 Contrôle et 2 Repos).

Chapitre Cinq - Sélection de Caractéristiques Sensibles à la Charge Affective

Dans le Chapitre Trois - , on a recensé différentes caractéristiques « standard » : ces corrélats (supposément) liés au stress ont été fréquemment introduits dans les modèles de l'état de l'art. Mais le manque de cohérence méthodologique observé dans la littérature nous a conduits à questionner le lien réel entre ces caractéristiques et le *stress* (tel qu'on l'entend dans ce manuscrit). En effet, le fait qu'une caractéristique se montre réactive en situation de stress ne suffit pas à montrer que le *stress* est bien la cause de la réaction mesurée. Comme on l'a montré au Chapitre Deux - , ce lien causal ne peut être établi qu'au regard de certains facteurs confondants : *l'effet de l'activité*, *l'effet de la tâche* et *l'effet du sujet*. A notre connaissance, ces facteurs n'ont jamais été simultanément pris en compte dans un processus de sélection de caractéristiques.

Dans ce chapitre, nous proposons une méthode de sélection de type « filtre » qui mesure la *charge affective* d'une caractéristique en situation de stress, c'est-à-dire les variations dans cette caractéristique qui se distinguent de la réponse réflexe du corps à une situation d'activité. Cette « part du stress » dans la réponse physiologique doit être *universelle*, c'est-à-dire observable quelle que soit la personne ; et *unique*, c'est-à-dire observable quelle que soit le contexte stressant. Dans la suite de ce chapitre, le terme « effet » sera réservé à l'observation d'une réponse physiologique à l'échelle de plusieurs tâches et de plusieurs sujets. On s'aidera pour cela de tests statistiques non-paramétriques.

Notre analyse s'appuie sur la base de données présentée au Chapitre Quatre - II.B : 32 caractéristiques calculées sur 4 tâches, dans 4 conditions, pour 11 à 20 participants selon la mesure. La méthode de sélection proposée repose sur quatre étapes clés.

- (1) Proposer un *modèle* de la charge affective en situation de stress, en fonction des facteurs confondants que l'on peut contrôler dans notre base de données.
- (2) *Estimer* la « part » de chacun de ces facteurs (ici : *stress*, *activité* et *environnement*) dans la valeur d'une caractéristique en formulant certaines hypothèses sur nos conditions expérimentales.
- (3) Sélectionner les *données* qui répondent à ces hypothèses, en s'aidant d'une vérité de terrain (ici : les évaluations psychologiques sur le niveau de stress).
- (4) Sélectionner les caractéristiques à partir d'un certain nombre de *critères*, sur la valeur des estimateurs parmi les données conservées.

Pour formuler nos critères de sélection, on a considéré qu'une caractéristique était *sensible* au stress lorsque sa « part du stress » était cohérente entre les sujets et les tâches. Cette même caractéristique serait *spécifique* au stress si elle se montrait peu sensible aux facteurs confondants. La sélection de caractéristiques *sensibles* au stress nous a conduit à rejeter la plupart des caractéristiques déjà citées dans la littérature ; et aucune des caractéristiques proposées dans ce manuscrit n'a satisfait notre critère de *spécificité*.

Avant de détailler nos calculs, la section I de ce chapitre propose un modèle additif dans lequel nous exprimons la « part » du stress et celle des facteurs confondants dans la valeur d'une caractéristique quelconque. La section II développe la méthode de sélection résumée ci-dessus. La section III présente nos résultats et leur discussion ; et la section IV propose une synthèse de nos observations dans le cadre de cette sélection.

I. Modélisation de l'Effet du Stress et des Facteurs Confondants sur une Caractéristique

I.A. Sources d'Influence Observables sur une Caractéristique Physiologique

Dans les paragraphes qui suivent, on définit les concepts que l'on cherche à estimer au travers des calculs de la section II. Pour cela, nous allons confronter notre protocole expérimental aux conclusions du Chapitre Deux - IV sur la mesure expérimentale du stress, afin de déterminer les phénomènes qu'il permet d'observer.

I.A.1. Notre protocole expérimental définit implicitement le stress

Comme nous l'avons vu au Chapitre Deux - III.B, une situation de stress correspond à un événement vécu comme *pertinent* et *difficile* par la personne. Dans cette situation, on désigne l'ensemble des bouleversements physiologiques à l'œuvre sous le terme de *charge allostatique*. De manière générale, les bouleversements physiologiques liés au stress possèdent au moins quatre causes distinctes.

- *L'homéostasie spontanée* adapte le corps à des facteurs tels que la température extérieure, les rythmes circadiens, l'activité des fonctions végétatives comme la digestion.
- *L'activité physique* engendre une réponse physiologique rapide et forte afin d'assurer les besoins énergétiques de l'organisme.
- *La charge mentale* est une réponse réflexe liée au recrutement des fonctions cognitives nécessaires au traitement de la situation.
- Enfin, *la charge affective* est une réponse décorrélée des besoins de l'organisme, liée à l'interprétation de la personne sur l'événement. Elle est le marqueur d'un état de stress.

Lorsqu'un modèle est construit sur une approche empirique, il est essentiel de savoir quels facteurs sont déterminants dans les mesures disponibles. En sciences affectives, ce problème n'est pas trivial : il est à l'origine de certaines incohérences autour de la notion de *stress* dans la littérature (voir Chapitre Deux -). Le but de cette section est de prévenir ces incohérences, en clarifiant trois éléments clés.

- a) Les facteurs confondants que l'on peut contrôler ;
- b) Les propriétés du « stress » que l'on peut observer ;
- c) Les conditions dans lesquelles nos résultats seront valides.

En raisonnant ainsi, on élucide le problème de la *définition implicite* que notre protocole expérimental donne au stress. Les paragraphes qui suivent définissent dans l'ordre ces trois éléments.

I.A.2. Deux facteurs confondants sont pris en compte : l'environnement et l'activité

La condition Repos permet de mesurer une *ligne de base* à chaque session expérimentale. Lorsqu'un participant réalise cette tâche, on le suppose libre de toute charge affective, cognitive ou physique. En principe, cette « ligne de base » correspond donc à l'homéostasie spontanée. En pratique,

une caractéristique physiologique peut dépendre de bien d'autres facteurs – comme la manière dont le capteur a été installé préalablement à l'expérience. Comme ces contributions ne peuvent être mesurées séparément, nous les regrouperons dans un facteur confondant unique : la **contribution de l'environnement**. Ici *l'environnement* désigne tout ce qui est indépendant des tâches expérimentales : depuis les conditions extérieures (température, humidité, etc.) jusqu'à la position des capteurs, en passant par le contexte « somatique » de la personne (position du corps, phase digestive, etc.). Cette contribution est imprévisible car elle dépend de facteurs non prévus par le protocole expérimental.

Nos tâches Contrôle ont, quant à elles, deux objectifs : (i) rester aussi neutres que possible sur le plan affectif, tout en (ii) reproduisant les processus cognitifs et moteurs à l'œuvre durant la tâche Stress correspondante. De cette manière, on cherche à contrôler la charge mentale et l'activité motrice que les tâches expérimentales imposent au participant. Comme ces deux facteurs ne peuvent être contrôlés séparément, nous les regrouperons dans un facteur confondant unique : la **contribution de l'activité**. Ici, le terme *activité* inclut toutes les fonctions cognitives et motrices engagées par les tâches réalisées – comme le langage (parler, écrire, ...), la mémoire, l'attention. Cette contribution est prévisible car elle dépend de facteurs imposés par le protocole expérimental.

I.A.3. Un concept-cible étendu : la contribution de la pression

Reste l'ensemble des phénomènes que l'on ne cherche pas à observer dans les conditions Repos et Contrôle, mais seulement dans la condition Stress. Durant cette période, ce qui vient s'ajouter à *l'environnement* et à *l'activité*, c'est la *pression* que le sujet ressent face aux stressors qu'on lui oppose. Lorsque les stressors sont *efficaces*, cette *pression* est l'effet d'un jugement social, d'un calcul difficile, d'un sentiment de douleur ou d'une contrainte temporelle. Dans ces conditions, on peut identifier la *pression* au stress émotionnel. La « part du stress » dans la réponse physiologique est alors celle d'un état d'éveil, parfois associé à un sentiment d'anxiété ou de détermination, et celle d'un effort mental supplémentaire par rapport à la condition Contrôle. En effet, nos quatre tâches Stress possèdent les attributs d'une situation stressante : elles sont *difficiles* et *pertinentes*, tandis que les tâches Contrôle ont été conçues pour ne poser aucune difficulté aux participants.

Mais le *stress* est un concept bien plus difficile à contrôler que *l'activité* ou *l'environnement*. Contrairement à la température extérieure, à l'heure du jour ou aux processus cognitifs et moteurs (qui peuvent être imposés par le protocole), le stress n'est pas qu'un facteur exogène : il dépend du *vécu* et de la personnalité des participants. Lorsque nos stressors sont inefficaces sur un participant, la « part des stressors » dans la réponse physiologique est celle d'une charge mentale et affective non-stressante. Ainsi, il nous appartiendra de montrer que nos tâches expérimentales étaient réellement perçues comme stressantes par nos participants. De la même manière, l'absence de stressor spécifique ne garantit pas l'absence de charge affective (ni même de stress) durant les conditions Repos et Contrôle. Afin de travailler sur un concept valide, on appellera **contribution de la pression** toute charge allostatique qui n'est liée ni à *l'activité* ni à *l'environnement*. En principe, cette contribution est d'origine affective.

I.A.4. Synthèse

Notre protocole expérimental ne permet pas de mesurer l'effet de certains phénomènes sans lien avec un stress émotionnel, mais susceptibles d'engendrer une réponse physiologique similaire. Ces phénomènes peuvent correspondre, par exemple, à une activité physique, à un sentiment positif

intense (ex. joie, libido), à une pathologie actuelle ou récente, à une consommation d’excitants. La liste n’est pas exhaustive.

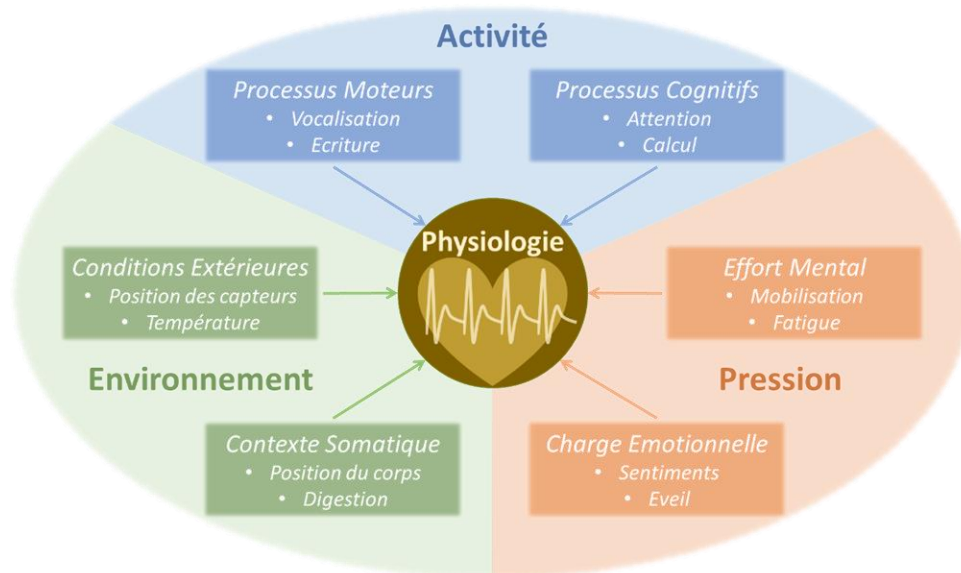


Figure 5.1 : Sources d’influence possibles sur une caractéristique physiologique

En toute rigueur, les marqueurs du stress qui seront mis en évidence dans ce chapitre ne seront efficaces qu’en l’absence de ces phénomènes. Ces conditions restreintes ont été volontairement imposées aux participants afin de cibler plus facilement la réponse physiologique au stress. C’est à ce prix que l’on peut concevoir cette réponse comme une synthèse des trois contributions (*activité, environnement, pression*) que l’on a définies dans cette section. Dans le cadre ciblé par notre protocole expérimental, on peut ainsi considérer qu’un nombre limité de facteurs détermine totalement la valeur d’une caractéristique physiologique quelconque. Ces facteurs sont résumés dans la Figure 5.1.

I.B. Hypothèses sur la Variation d’une Caractéristique en Fonction des Contributions

Pour expliquer les calculs proposés en section II, on peut formaliser le modèle précédent en exprimant nos trois contributions dans nos trois conditions expérimentales : Repos, Contrôle et Stress. Pour cela, on doit formuler certaines hypothèses sur le comportement d’une caractéristique dans notre base de données. C’est l’objet des paragraphes qui suivent.

I.B.1. La charge allostatique est la synthèse de trois contributions indépendantes

Soit une caractéristique f (de l’anglais : *feature*), calculée sur une fenêtre temporelle τ . En situation de stress, on peut modéliser la valeur de f à l’instant τ à partir des trois contributions définies dans la section précédente. On considère pour cela l’expression suivante :

$$f_{\tau} - f_0 = \varepsilon_{\tau} + \alpha_{\tau} + \pi_{\tau} \tag{5-1}$$

La signification de chacune de ces composantes est donnée dans le Tableau 5.1 ci-dessous.

Tableau 5.1 : Composantes du modèle

Composante	Signification
f_{τ}	Niveau de f à l'instant τ
f_0	Niveau de base Valeur fictive de f correspondant à un équilibre homéostatique parfait
ε_{τ}	Contribution de l' <i>environnement</i> à l'instant τ Composante spontanée liée au contexte de la mesure
α_{τ}	Contribution de l' <i>activité</i> à l'instant τ Effet des processus cognitifs et moteurs liés à la tâche en cours
π_{τ}	Contribution de la <i>pression</i> à l'instant τ Charge affective issue de la perception de la tâche par la personne

Le terme : $f_{\tau} - f_0$ représente la charge allostatique sur la caractéristique f . L'objectif de la section II sera de classer chacune des caractéristiques f en fonction de la « part du stress » dans cette réponse (la composante π), par rapport la « part des facteurs confondants » (les composantes ε et α). Le choix d'un modèle additif permet de justifier, entre autres, le calcul des « estimateurs » qui nous permettront de mesurer ces contributions.

En décomposant ainsi la charge allostatique, on suppose que nos facteurs *environnement*, *activité* et *pression* agissent sur la caractéristique f de manière indépendante. En d'autres termes : les sources des composantes ε , α et π (contexte extérieur et somatique, processus moteurs et cognitifs, état d'éveil et sentiments...) n'interfèrent pas entre elles et s'excluent mutuellement. On pourra vérifier cette hypothèse d'indépendance dans l'Annexe 8, à partir des résultats montrés dans la section III.B.

Bien qu'on les ait définies pour isoler la « part du stress », on suppose ici que la présence de ces contributions ne dépend pas de l'état de stress. Les composantes ε , α et π peuvent être non-nulles dans chaque condition expérimentale. L'expression (5-1) est valable pour n'importe quelle fenêtre de temps τ à l'intérieur d'une session de mesures ; et il existe autant de valeurs de ε , α et π que de valeurs de f dans la base de données.

I.B.2. Une contribution reste stable durant une condition expérimentale

Afin de proposer un modèle plus simple, on formule l'hypothèse suivante : au sein d'une *même* condition expérimentale (ex. : TSST_{Rs}, la condition de repos avant la version stressante du TSST), les contributions ε , α et π varient « peu » par comparaison aux différences observées *entre* deux conditions d'une même tâche (ex : TSST_{Rs} → TSST_S).

Une telle propriété nous permet de conserver un échantillon *représentatif* pour chaque condition expérimentale. Pour chaque caractéristique f , on disposera ainsi d'une mesure $f_{i,j,k}$ par tâche $i \in \{D2, SCEPT, TSST, TA\}$, par condition $j \in \{Rc, Rs, C, S\}$ et par participant $k \in \{1..20\}$. Pour limiter l'influence d'éventuels artéfacts, on a sélectionné la valeur médiane de f comme valeur représentative de chaque enregistrement (i,j,k) . A nouveau, on suppose que cet échantillon unique est déterminé par trois composantes : $\varepsilon_{i,j,k}$, $\alpha_{i,j,k}$ et $\pi_{i,j,k}$, c'est-à-dire les contributions représentatives de l'*environnement*, de l'*activité* et de la *pression* sur cet enregistrement.

Cette hypothèse est nécessaire pour exprimer le modèle de l'équation (5-1) dans chaque condition expérimentale. Elle ne sera pas remise en question dans la suite de cette étude.

I.B.3. Certaines contributions sont identiques entre certaines conditions

Soit une tâche expérimentale et un participant donnés ($i \in \{D2, SECPT, TA, TSST\}$, $k \in \{1..20\}$). Par commodité, on notera en minuscules : x_j la contribution x représentative de la condition $j \in \{rs, c, rc, s\}$. Avec ces notations, notre protocole expérimental nous permet d'établir un lien d'égalité entre certaines contributions de l'équation (5-1).

- La version Contrôle de chaque tâche a été conçue pour engager les mêmes processus cognitifs et moteurs que la version Stress. Ainsi : $\alpha_c \approx \alpha_s = \alpha$.
- Les versions Stress et Contrôle de chaque tâche étaient effectuées immédiatement à la suite de leurs Repos respectifs, sur un intervalle court et dans le même contexte. Ainsi, on peut considérer que la contribution de l'environnement (ε) n'a pas varié entre chaque Repos et la tâche qui suit. Ce qui donne d'une part : $\varepsilon_{rc} \approx \varepsilon_c$ et d'autre part : $\varepsilon_{rs} \approx \varepsilon_s$.

Ces hypothèses sont nécessaires pour que la condition Repos permette de contrôler ε ; et pour que la condition Contrôle permette de contrôler α . Elles ne seront pas remises en question dans la suite de cette étude.

I.B.4. Certaines contributions sont négligeables dans certaines conditions

Pour finir, le protocole a été conçu pour minimiser certaines contributions selon deux scénarios.

a) Scénario idéal

- D'une part, la contribution de l'*activité* (α) et celle de la *pression* (π) sont minimales durant les conditions de Repos (R_c, R_s). Le participant restait sans activité, sans distraction et ne pouvait anticiper la tâche à venir. Ainsi, théoriquement : $\alpha_{rc,rs} \approx 0$ et $\pi_{rc}, \pi_{rs} \approx 0$.
- D'autre part, la contribution de la *pression* (π) est minimale dans les conditions Contrôle (C). Celles-ci étaient effectuées sans présence de stresser et demandaient un effort mental minimal. Ainsi, théoriquement : $\pi_c \approx 0$.

Ces hypothèses permettent de simplifier l'expression du modèle (5-1) dans les conditions Repos et Contrôle. Elles seront utiles pour comprendre la cible de nos estimateurs dans la section II.

b) Scénario réaliste

Etant donné que cela est imposé par le protocole expérimental, on peut raisonnablement compter sur l'absence d'*activité* durant les périodes de Repos : $\alpha_{rc,rs} \approx 0$.

En revanche, notre protocole ne permet pas de garantir l'absence de *charge affective* sur ces mêmes périodes. Un participant peut éprouver, par exemple, une émotion ou de la fatigue résiduelles d'une activité antérieure à la tâche de repos. De même, la version Contrôle de nos quatre tâches est susceptible de générer des affects non-stressants, comme l'apaisement ou l'ennui. On a aussi supposé que l'absence de stress n'était pas garantie hors des conditions Stress. Dans un scénario plus réaliste que le précédent, on peut ainsi considérer que les contributions de la *pression* ne sont pas négligeables en conditions de non-stress : $\pi_{rs}, \pi_{rc}, \pi_c \neq 0$.

Cette dernière hypothèse permet de prendre en compte certains résultats qui seront obtenus dans la section III. Comme le modèle par contributions, elle sera vérifiée dans l'Annexe 8.

I.B.5. Synthèse

L'équation (5-2) ci-dessous exprime la charge allostatique sur la caractéristique f dans chaque condition $j \in \{rc, rs, c, s\}$, pour une tâche i et un participant k donnés. Les composantes additionnelles du scénario « réaliste » par rapport au scénario « idéal » sont identifiées en **gris**.

$$\begin{cases} f_{rc} - f_0 = \varepsilon_{rc} & + \pi_{rc} \\ f_{rs} - f_0 = \varepsilon_{rs} & + \pi_{rs} \\ f_c - f_0 = \varepsilon_{rc} + \alpha & + \pi_c \\ f_s - f_0 = \varepsilon_{rs} + \alpha & + \pi_s \end{cases} \quad (5-2)$$

Dans le cas général, la contribution π exprime l'effet de tout ce qui est lié au vécu de la personne. Lorsque la tâche Stress est efficace, π_s exprime ainsi l'effet du stress. Cette « part du stress » dans les variations de f est identifiée en **rouge** dans l'équation (5-2). Si cette contribution est *grande*, alors la caractéristique f est *sensible* au stress émotionnel. Si cette contribution est *prédominante* sur les deux autres (ε , α), alors la caractéristique f est *spécifique* au stress émotionnel. Pour une sélection de type *filtre*, on pourrait simplement classer nos caractéristiques selon leur sensibilité au stress. Mais une caractéristique *spécifique* sera bien plus utile pour détecter le stress en vie réelle, d'où l'intérêt de mesurer aussi l'effet des facteurs confondants. Ce raisonnement est à la base des méthodes développées dans la section suivante.

II. Méthodes pour la Sélection de Caractéristiques Sensibles et Spécifiques au Stress

A partir des mesures disponibles pour une même tâche et un même participant, on peut *estimer* la part du stress et des facteurs confondants dans l'évolution d'une caractéristique : ce sera l'objet de la section II.A. A l'échelle de plusieurs tâches et plusieurs participants, on peut vérifier si ces contributions sont *uniques* (elles ne dépendent pas des tâches stressantes effectuées) et *universelles* (elles sont observables pour tous les sujets). Pour cela, nous avons besoin de critères statistiques adaptés aux mesures physiologiques : ceux-ci sont présentés en section II.B. Notre stratégie de sélection des caractéristiques est finalement développée en section II.C.

Comme indiqué dans la section précédente, on travaillera sur la valeur médiane f_{ijk} de chaque période d'enregistrement (une tâche $i \in \{D2, SECPT, TA, TSST\}$, une condition $j \in \{Rc, Rs, C, S\}$, un participant $k \in \{1..20\}$). Une telle réduction des données (à $4 \times 4 \times 20$ échantillons par caractéristique) présente deux avantages : (i) diminuer la présence d'artéfacts dans notre base de données ; et (ii) focaliser nos calculs sur l'évolution de f au travers des conditions expérimentales.

II.A. Estimation de la Part du Stress et des Facteurs Confondants

II.A.1. Scénario « idéal » : part du stress et part des facteurs confondants

Soit une tâche i et un participant k donnés. A partir des quatre mesures f_j correspondantes, on peut calculer cinq *estimateurs* qui ont chacun pour cible une contribution de l'équation (5-2). Le calcul de

ces « méta-caractéristiques » est présenté dans le Tableau 5.2 (ci-dessous). Dans la Figure 5.2 (plus bas), des mesures f_j fictives sont utilisées afin d'illustrer la position relative de ces estimateurs dans l'espace de la caractéristique f .

Chaque estimateur porte le nom de sa « contribution cible ». Par exemple, l'estimateur d'Activité (**A**) cherche à évaluer la contribution de l'activité (α), c'est-à-dire l'effet des processus cognitifs et moteurs sur f . Dans notre « scénario idéal » (voir I.B.4), cette contribution correspond exactement à la variation de f entre les conditions Repos et Contrôle. D'après l'équation (5-2) : $A = f_c - f_{rc} = \alpha$.

Tableau 5.2 : Calcul des cinq estimateurs

Effet	Symbole & Calcul	Description
Tâche	$T = f_s - f_{rs}$	Variation de f induite par la condition Stress
Activité	$A = f_c - f_{rc}$	Variation de f induite par la condition Contrôle
Contrainte	$K = f_s - f_c$	Différence entre les conditions Stress et Contrôle
Contexte	$E = f_{rs} - f_{rc}$	Différence entre les deux Repos
Pression	$P = T - A$ $P = K - E$	Différence entre la variation induite par la condition Stress et la variation induite par la condition Contrôle

De même, on peut chercher à estimer l'effet **T** de la Tâche en condition Stress. La présence de cet effet est nécessaire pour que f puisse être utilisée dans un estimateur de stress ; car une absence d'effet signifierait que f est identique en situation de stress et de repos. Dans le scénario idéal, d'après l'équation (5-2), cet effet s'exprime : $T = \alpha + \pi_s$. Il regroupe ainsi les contributions de l'activité (α) et de la pression (π).

De son côté, l'estimateur **E** s'exprime : $E = \varepsilon_{rs} - \varepsilon_{rc}$. Il mesure « l'effet du contexte » sur la contribution de l'environnement, c'est-à-dire la variation spontanée de la caractéristique entre deux sessions expérimentales. C'est cette variation qui fait la nature « confondante » du facteur environnement, car elle peut masquer la « part du stress » dans les variations de la caractéristique f entre deux sessions de mesures différentes. On nommera désormais ε_e cette « contribution du contexte », facteur confondant estimé par **E**.

$$\varepsilon_e = \varepsilon_{rs} - \varepsilon_{rc}$$

Un tel effet est à prévoir si l'on fait directement la différence entre les valeurs f_s et f_c ; car elles ont été enregistrées sur deux jours différents. Si l'on appelle « contrainte » le simple ajout de stressseurs à une tâche Contrôle pour obtenir sa version Stress, on peut exprimer l'effet de la Contrainte (**K**) comme la somme des contributions de la pression et du contexte : $K = f_s - f_c = \pi_s + \varepsilon_e$.

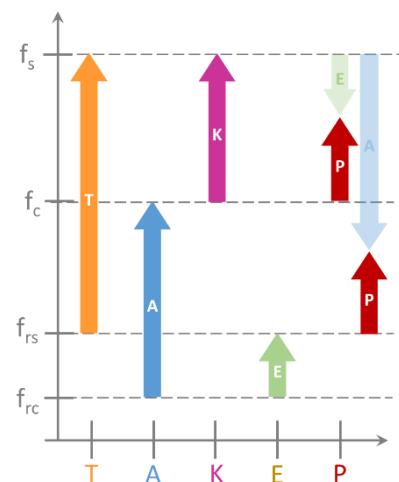


Figure 5.2 : Position relative des estimateurs

Pour finir, on peut mesurer directement la contribution du stress (π_s) avec l'estimateur Pression : $\mathbf{P} = (f_s - f_{rs}) - (f_c - f_{rc}) = \pi_s$. On peut ainsi exprimer les cinq estimateurs du Tableau 5.2 à partir de seulement trois composantes : π_s , α et ε_e .

$$\begin{cases} \mathbf{T} = \alpha + \pi_s \\ \mathbf{A} = \alpha \\ \mathbf{K} = \pi_s + \varepsilon_e \\ \mathbf{E} = \varepsilon_e \\ \mathbf{P} = \pi_s \end{cases} \quad (5-3)$$

Dans ce scénario « idéal », il suffit que la tâche stressante soit efficace pour que la contribution π_s (en rouge) représente la « part du stress » dans la valeur de la caractéristique f ; tandis que α et ε_e représentent la « part des facteurs confondants ». A partir de l'équation (5-3), on peut alors traduire et compléter les propriétés recherchées pour f en fonction de la valeur de ces estimateurs.

Si la valeur de \mathbf{T} est « grande », alors f est *réactive en situation de stress* pour la tâche i et le participant k . Cette condition est nécessaire, mais pas suffisante pour montrer que f répond au stress *émotionnel*; car \mathbf{T} inclut la contribution de l'*activité* (α). Une « grande » valeur de \mathbf{P} suffit alors pour montrer que f est *sensible au stress émotionnel* pour la tâche i et le participant k ; car \mathbf{P} estime directement la « part du stress » (π_s) dans la variation mesurée par \mathbf{T} . Pour finir, \mathbf{A} mesure directement la contribution de l'*activité* (α); et \mathbf{E} mesure la contribution du *contexte* (ε_e). Si les valeur de \mathbf{A} et \mathbf{E} sont « petites », alors on peut considérer que f est aussi *spécifique au stress émotionnel*.

II.A.2. Scénario « réaliste » : approximations et contributions parasites

Dans l'hypothèse où la charge affective des participants ne peut pas être contrôlée pendant les périodes de non-stress (Rc, Rs, C), l'équation (5-3) ci-dessus comprend alors des termes *pression* supplémentaires qui sont représentés en gris dans l'équation (5-4) ci-dessous.

$$\begin{cases} \mathbf{T} = \alpha + (\pi_s - \pi_{rs}) & = \alpha + \pi_t \\ \mathbf{A} = \alpha + (\pi_c - \pi_{rc}) & = \alpha + \pi_a \\ \mathbf{K} = (\pi_s - \pi_c) + \varepsilon_e & = \pi_k + \varepsilon_e \\ \mathbf{E} = \varepsilon_e + (\pi_{rs} - \pi_{rc}) & = \varepsilon_e + \pi_e \\ \mathbf{P} = (\pi_s - \pi_c) + (\pi_{rs} - \pi_{rc}) & = \pi_k - \pi_e \end{cases} \quad (5-4)$$

La dernière colonne fait apparaître des termes *pression* composés (π_t , π_k , π_a , π_e), que l'on peut interpréter de la même manière que les estimateurs \mathbf{T} , \mathbf{K} , \mathbf{A} , \mathbf{E} .

- $\pi_t = \pi_s - \pi_{rs}$ marque un « effet de la tâche » sur la contribution de la *pression*. Lorsque celle-ci est *négligeable* en condition Repos par rapport à la condition Stress ($\pi_{rs} \ll \pi_s$), on a : $\pi_t \approx \pi_s$.
- $\pi_k = \pi_s - \pi_c$ marque un « effet de la contrainte » sur la contribution de la *pression*. Lorsque celle-ci est *négligeable* en condition Contrôle par rapport à la condition Stress ($\pi_c \ll \pi_s$), on a : $\pi_k \approx \pi_s$.
- $\pi_a = \pi_c - \pi_{rc}$ marque un « effet de l'activité » sur la contribution de la *pression*. Lorsque celle-ci est *similaire* entre les conditions Repos et Contrôle ($\pi_c \approx \pi_{rc}$), on a : $\pi_a \approx 0$ et $\mathbf{A} \approx \alpha$.
- $\pi_e = \pi_{rs} - \pi_{rc}$ marque un « effet du contexte » sur la contribution de la *pression*. Lorsque celle-ci est *similaire* entre les deux Repos ($\pi_{rs} \approx \pi_{rc}$), on a : $\pi_e \approx 0$ et $\mathbf{E} \approx \varepsilon_e$. De même : $\mathbf{P} \approx \pi_k$.

Dans ce scénario, les termes π_t et π_k sont des approximations de la « part du stress », tandis que les termes π_a et π_e peuvent être qualifiés de « parasites ». \mathbf{K} et \mathbf{P} apparaissent ainsi comme deux estimateurs *imparfaits* de la « part du stress »; \mathbf{A} comme un estimateur *imparfait* de la « part de

l'activité » ; et **E** comme un estimateur *imparfait* de la « part du contexte » telle qu'on l'a définie précédemment. On rejoint le scénario « idéal » lorsque nos 4 tâches (Rc, Rs, C, S) ont eu les effets attendus sur nos participants, c'est-à-dire lorsqu'on vérifie la relation suivante : $\pi_{rs} \approx \pi_{rc} \approx \pi_c \ll \pi_s$, où π_s est liée au stress émotionnel.

II.A.3. Synthèse et perspectives

Dans le scénario « idéal », on suppose qu'une « grande » valeur de **P** suffit à montrer que *f* est sensible au stress émotionnel. Dans le scénario « réaliste », cette propriété sur **P** n'est valide qu'à deux conditions : pour la tâche *i*, la charge affective du participant *k* est *équivalente* entre les deux Repos ($\pi_{rs} \approx \pi_{rc}$) et *négligeable* en condition Contrôle ($\pi_c \ll \pi_s$). Nos évaluations subjectives du niveau de stress peuvent fournir des éléments de réponse pour vérifier ces hypothèses. Cependant, elles ne peuvent prouver l'absence de sentiment non-stressants (*e.g.*, apaisement, ennui), qui pourraient biaiser la mesure de nos contributions cibles : la « part du stress » et celle des facteurs confondants.

Pour que la valeur de nos estimateurs ait un sens, il nous appartiendra de montrer que le scénario « idéal » est *vraisemblable* dans nos données expérimentales.

- En sélectionnant d'une part, les tâches expérimentales (D2, TSST...) qui ont montré les *effets attendus* sur nos participants (*i.e.* : stress fort en condition Stress, stress faible sinon).
- En validant d'autre part, la valeur d'un estimateur par un autre (*i.e.*, **P** par **K**), afin de montrer qu'on mesure bien sa *contribution cible* et non un effet parasite.

Dans l'évolution d'une caractéristique *f*, on peut montrer que la part du stress (π_s) ou celle des facteurs confondants (ε_e, α) est « universelle » si on la mesure à l'échelle d'un *groupe d'individus*. De même, cette part est « unique » si on la mesure à l'échelle d'un *groupe de tâches*. Ces propriétés se traduisent par la présence ou l'absence d'un *effet* dans les valeurs d'un estimateur, lorsque ces valeurs sont calculées pour l'ensemble des participants et des tâches disponibles. Cet *effet* doit être mesuré à l'aide d'indicateurs statistiques : c'est l'objet de la section suivante.

II.B. Critères de Sélection des Caractéristiques

II.B.1. Contraintes liées à notre base de données

Soit une caractéristique *f*, pour laquelle on souhaite mesurer l'*effet* observé pour un estimateur quelconque $\mathbf{X} \in \{\mathbf{T}, \mathbf{A}, \mathbf{K}, \mathbf{E}, \mathbf{P}\}$, à l'échelle inter-individuelle. Les valeurs de **X** ont été calculées sur l'ensemble des données *disponibles*. Or, le nombre d'exemples disponibles pour la mesure de **X** varie en fonction de la caractéristique et de l'ensemble de tâches choisi. Par exemple, les caractéristiques respiratoires concerneront au maximum 11 participants (voir Chapitre Quatre - II.A). On dispose ainsi d'un échantillon de données $\{\mathbf{X}_n\}_{n \in \{1 \dots N\}}$ dont la taille **N** peut varier entre quelques exemples et N_{\max} ($N_{\max}=20$ lorsqu'on ne considère qu'une tâche, $N_{\max}=80$ pour quatre tâches).

Par ailleurs, il faut considérer que la valeur de **X** est imprévisible et ne répond à aucune loi de probabilité. Ceci est d'autant plus vrai que **X** peut être estimée pour n'importe quelle caractéristique physiologique (énergie d'un signal, nombre de pics, etc.). Ainsi ses valeurs ne peuvent pas toujours prendre une distribution gaussienne. Enfin, il faut s'attendre à la présence de valeurs aberrantes dans nos échantillons de données, pouvant découler d'un éventuel problème de mesure.

Pour s'affranchir de ces contraintes (distribution des données variable, valeurs aberrantes), on ne travaillera pas sur l'*amplitude*, mais sur la *cohérence* des valeurs de \mathbf{X} dans notre échantillon de données. Dans les sections suivantes, on appellera « cohérence » l'homogénéité des signes dans les valeurs de \mathbf{X} (qui est une variable normalement centrée sur Zéro). Si la majorité des valeurs ont même signe, alors ces variations sont cohérentes entre elles. Cette notion est illustrée dans la Figure 5.3.

II.B.2. Mesurer un effet

Le test des rangs signés de Wilcoxon vérifie si la médiane d'un échantillon de valeurs est différente de Zéro. C'est ce test que l'on va utiliser pour déterminer la présence d'un effet. Soit un échantillon de données $\{x_i\}$: on cherche à montrer la *présence d'un effet* à l'intérieur de cet échantillon. Cela revient à rejeter l'hypothèse nulle H_0 , qui considère que la médiane de $\{x_i\}$ est nulle au profit d'une hypothèse alternative H_A , qui considère que la médiane de $\{x_i\}$ n'est pas nulle. Ce test sera appliqué aux estimateurs $\mathbf{X} \in \{\mathbf{T}, \mathbf{A}, \mathbf{K}, \mathbf{E}, \mathbf{P}\}$, pour chaque caractéristique f .

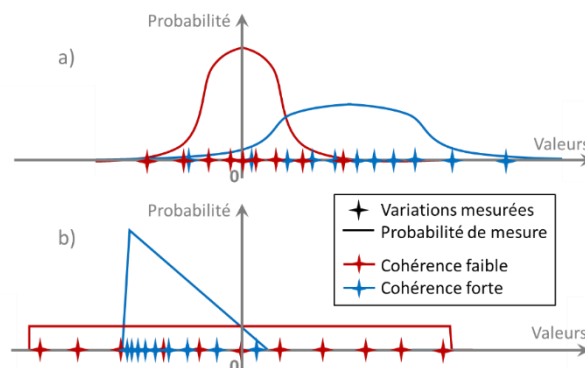


Figure 5.3 : Variations fortement cohérentes (en bleu) ou faiblement cohérentes (en rouge), de distribution quasi-gaussienne (a) ou quelconque (b)

Le test des rangs signés de Wilcoxon est un test statistique standard que l'on peut appliquer aux données dont on ne connaît pas la densité de probabilité. Il a l'avantage d'être robuste aux échantillons de petite taille ($N \geq 5$) et à la présence de valeurs aberrantes. La procédure du test est largement décrite dans la littérature scientifique (Gibbons & Chakraborti, 2003, p. 197-205). Initialement développée pour montrer la *présence d'un effet*, ce test peut être adapté pour montrer l'*absence d'effet* selon une procédure de test d'équivalence (Lakens et al., 2018). Pour ne pas développer ici-même des éléments de statistiques, un résumé de ces procédures (et la détermination de certains paramètres) est donné en Annexe 6. A partir du test de Wilcoxon, on peut ainsi produire trois indicateurs qui serviront de base à la sélection des caractéristiques.

- Une probabilité ^{eff}p : c'est la p-valeur obtenue à partir du test bilatéral classique. Elle représente le « risque de se tromper » si on affirme qu'un effet est *présent* dans l'échantillon.
- Une probabilité ^{abs}p : c'est la p-valeur obtenue à partir du test d'équivalence. Elle représente le « risque de se tromper » si on affirme que l'effet est *absent* de l'échantillon.
- Une mesure ϑ : c'est la *taille d'effet* mesuré dans l'échantillon. Elle mesure le déséquilibre des valeurs de l'échantillon \mathbf{X} autour de Zéro.

Contrairement à ^{eff}p et à ^{abs}p , la valeur de ϑ ne dépend pas de la taille \mathbf{N} de l'échantillon. Ces propriétés de la taille d'effet en font une mesure de la *cohérence* des variations d'une caractéristique f quelconque à l'échelle interindividuelle. Si on considère l'ensemble des valeurs de f disponibles pour une tâche donnée, on peut appliquer le calcul de ϑ pour chacun des estimateurs $\mathbf{T}, \mathbf{A}, \mathbf{K}, \mathbf{E}$ et \mathbf{P} . Ainsi ϑ mesure l'effet Tâche⁵ (ϑ_T), l'effet Activité (ϑ_A), l'effet Contrainte (ϑ_K), l'effet Contexte (ϑ_E) et l'effet

⁵ Dans ce chapitre, l'effet Tâche (ϑ_T) désigne l'effet de l'exécution d'une ou plusieurs tâches stressantes sur la valeur d'une caractéristique, à l'échelle interindividuelle. A ne pas confondre avec l'*effet de la tâche* : un facteur confondant mis en évidence au Chapitre Deux - IV, cité comme tel dans les autres chapitres (*i.e.* en minuscules italiques), qui désigne la différence observée entre plusieurs tâches sur la réponse physiologique au stress.

Pression (ϑ_p). Pour chacun de ces estimateurs, on pourra comparer la *taille d'effet* de toutes les caractéristiques, quel que soit le nombre d'exemples disponibles pour chacune d'elles.

II.B.3. Démontrer la présence ou l'absence d'un effet

a) Critère sur ϑ

En pratique, « $\vartheta \approx 0$ » indique une absence d'effet, tandis que « $\vartheta \gg 0$ » caractérise un effet positif et « $\vartheta \ll 0$ » caractérise un effet négatif. Il est d'usage de considérer un *effet modéré* lorsque la valeur absolue de ϑ dépasse 0,3 (Fritz et al., 2012). En-dessous de ce seuil, l'effet est donc *faible*. On peut ainsi postuler l'existence d'un seuil $\vartheta_{seuil} = 0.3$ tel que :

- a) On considère la *présence* d'un effet si $\vartheta > \vartheta_{seuil}$.
- b) On considère l'*absence* d'un effet si $\vartheta < \vartheta_{seuil}$.

Ce critère est indépendant de la présence de valeurs aberrantes et de la taille de l'échantillon. Mais sa valeur seuil reste arbitraire ; et la question de son interprétation lorsqu'on cherche à démontrer que le stress a (ou n'a pas) un *impact* sur la caractéristique f n'est pas entièrement résolue. Seules les probabilités ^{eff}p et ^{abs}p peuvent apporter une réponse forte à cette question.

b) Critère sur p

Selon les principes du test statistique : si ^{eff}p (ou respectivement : ^{abs}p) est inférieure à un certain seuil p_{seuil} , on pourra conclure à la présence d'effet (ou inversement : une absence d'effet pour ^{abs}p), avec un risque d'erreur égal à p_{seuil} . Dans la littérature, il est d'usage de fixer : $p_{seuil} = 0,05$. Mais pour notre base de données, cette approche standard présente un défaut : la probabilité de passer sous p_{seuil} augmente sensiblement avec le nombre d'exemples disponibles. Ainsi, on aura bien plus de chances de trouver un effet du stress sur le rythme cardiaque que sur le rythme respiratoire (avec respectivement 20 et 11 sujets disponibles). Pour corriger ce problème, on peut ajuster p_{seuil} de sorte à conserver une *puissance statistique* suffisante pour différentes tailles d'échantillon. Pour une justification plus précise de ce choix méthodologique, se reporter en Annexe 6.

Cet ajustement est nécessaire, mais pas idéal pour compenser fidèlement l'effet des variations de N. Pour conserver une référence parfaitement indépendante de la taille d'échantillon (et donc de la caractéristique f), on aura donc aussi recours au critère proposé plus haut sur la taille d'effet ϑ . Comme le critère sur p est plus *conservatif* que le critère sur ϑ , le second aura plus de chances que le premier de révéler la présence (ou l'absence) d'un effet dans nos échantillons de données.

c) Synthèse

On a ainsi défini un **critère fort** et un **critère faible** pour déterminer l'*absence* ou la *présence* d'un *effet* dans les valeurs de nos estimateurs Activité (A), Tâche (T), Contexte (E), Contrainte (K) ou Pression (P). Ces critères sont résumés dans le Tableau 5.3 ci-dessous, avec pour paramètres : $p_{seuil}=0,1$ pour $N \leq 15$; $p_{seuil}=0,05$ pour $15 < N \leq 25$; $p_{seuil}=0,01$ pour $25 < N \leq 45$; et $p_{seuil}=0,001$ pour $N > 45$; et $\vartheta_{seuil}=0.3$.

Tableau 5.3 : Critères statistiques sur la présence d'un effet dans un échantillon de données

	Présence d'effet	Absence d'effet
Critère fort	*** : $^{eff}p < p_{seuil}$	ooo : $^{abs}p < p_{seuil}$
Critère faible	* : $ \vartheta > \vartheta_{seuil}$	o : $ \vartheta < \vartheta_{seuil}$

L'ensemble de ces critères est utilisé dans notre stratégie de sélection des caractéristiques, que l'on développe dans la section suivante.

II.C. Stratégie de Sélection des Caractéristiques

II.C.1. Validation du protocole : sélection des tâches

a) Principe

Avant tout calcul sur les caractéristiques physiologiques, on cherche à vérifier que nos tâches expérimentales ont bien produit les effets attendus sur nos participants. Soit une tâche $i \in \{\text{TSST, TA, D2, SECPT}\}$. Si l'on nomme π_j ($j \in \{rc, rs, c, s\}$) la contribution de la *pression* pour la tâche i dans ses quatre versions, on aimerait montrer la propriété suivante : $\pi_{rs} \approx \pi_{rc} \approx \pi_c \ll \pi_s$ pour la majorité des participants à l'expérience. Cette analyse s'appuiera donc sur les tests statistiques qui ont été développés en section II.B. Comme l'estimation de $(\pi_{rs}, \pi_{rc}, \pi_c)$ est impossible sur le plan physiologique, on s'appuiera sur les évaluations psychologiques subjectives retenues par le protocole.

Parmi ces évaluations, l'échelle visuelle analogique de stress (**EVAS**) et le questionnaire STAI-Y-A (**STAI**) permettent de mesurer le niveau de stress de nos participants dans les différentes conditions de notre base de données. En revanche, il n'existe pas de mesure pertinente de la *pression* lorsque celle-ci est non-stressante.

Lorsque les niveaux de stress rapportés sont bien supérieurs en condition Stress que dans les autres conditions, les hypothèses : $\pi_{rs} \ll \pi_s$ et $\pi_c \ll \pi_s$, sont vraisemblables car la différence de *pression* peut être assimilée à cette variation du niveau de stress. En revanche, la comparaison est moins évidente entre deux périodes non-stressantes. De notre point de vue, l'hypothèse : $\pi_{rc} \approx \pi_c$ est vraisemblable si le niveau de stress n'a pas évolué entre les conditions Repos et Contrôle ; car ces deux tâches se suivent immédiatement et la tâche Contrôle est normalement neutre sur le plan émotionnel. Cependant, les mesures EVAS et STAI ne sont pas pertinentes pour montrer l'équivalence de *pression* entre les Repos ($\pi_{rc} \approx \pi_{rs}$) ; car lorsque le participant reste inactif, son état émotionnel dépend en partie du contexte. Or, les deux conditions Repos se déroulent sur deux jours différents.

Dans la suite de ce chapitre, une tâche expérimentale $i \in \{\text{TSST, TA, D2, SECPT}\}$ sera considérée comme *valide* si elle vérifie les trois conditions suivantes.

- a) Les niveaux de stress rapportés sont supérieurs en condition Stress qu'en condition Repos.
- b) Les niveaux de stress rapportés sont supérieurs en condition Stress qu'en condition Contrôle.
- c) Les niveaux de stress rapportés sont équivalents entre les conditions Repos et Contrôle.

b) Critères de sélection

La comparaison des niveaux de stress entre plusieurs conditions d'une même tâche peut être réalisée par les estimateurs Activité (**A**), Tâche (**T**), Contrainte (**K**) et Pression (**P**), appliqués aux valeurs disponibles pour EVAS et STAI.

En effet, soit une tâche quelconque $i \in \{\text{D2, SECPT, TA, TSST}\}$; et ces estimateurs calculés sur l'ensemble des 20 sujets disponibles pour l'évaluation $\psi \in \{\text{EVAS, STAI}\}$. Des valeurs positives de l'estimateur Tâche (**T**) indiquent que le niveau de stress était globalement plus élevé durant la condition Stress que durant la condition Repos (critère (a)). Ainsi, la validation du critère fort ($^{eff}p < p_{seuil}$) associé à une taille d'effet positive ($\vartheta > 0$) démontre un *effet stressant* sur nos participants. De même,

on vérifie que la condition Stress est plus stressante que la condition Contrôle en montrant que les valeurs de l'estimateur Contrainte (**K**) sont globalement positives (critère (b)). A nouveau, la validation du critère fort ($^{eff}p < p_{seuil}$) associé à une taille d'effet positive ($\vartheta > 0$) démontre cet effet stressant.

Pour finir, des valeurs de l'estimateur Activité (**A**) centrées sur Zéro indiquent que la condition Contrôle n'a globalement pas d'effet stressant par rapport au Repos (critère (c)). La validation du critère fort ($^{abs}p < p_{seuil}$) démontre cette *absence* d'effet stressant. Cela dit, *l'absence d'effet* sur l'estimateur Activité (**A**) ne veut pas dire que les variations de $\psi \in \{\text{EVAS, STAI}\}$ sont *faibles* entre Repos et Contrôle ; mais qu'elles sont *incohérentes* entre elles. Pour consolider le critère (c), on peut vérifier que la variation Repos-Contrôle dans les niveaux de stress (estimateur **A**) est bien inférieure à la variation Repos-Stress (estimateur **T**). Cela revient à montrer que les valeurs de l'estimateur Pression (**P = T - A**) sont majoritairement positives. La validation du critère fort ($^{eff}p < p_{seuil}$) associé à une taille d'effet positive ($\vartheta > 0$) démontre cette absence relative de variation du niveau de stress en condition Contrôle par rapport à la condition Stress.

A partir des évaluations subjectives, on peut ainsi valider notre protocole expérimental au moyen des sept critères exprimés dans l'équation (5-5) ci-dessous. Pour augmenter la validité de ces critères, ceux-ci doivent être vérifiés à la fois pour les EVAS et STAI.

$$i \in \mathcal{V} \Leftrightarrow \begin{cases} \theta_{T,K,P}(i) > 0 \\ ^{eff}p_{T,K,P}(i) < p_{seuil} \\ ^{abs}p_A(i) < p_{seuil} \end{cases} \quad (5-5)$$

II.C.2. Caractérisation de la réponse au stress

a) Caractéristiques sensibles au stress

Supposons maintenant qu'il existe un ensemble \mathcal{V} de tâches qui ont été sélectionnées à partir des critères ci-dessus. Soit une caractéristique f , pour laquelle on aura calculé les estimateurs Activité (**A**), Tâche (**T**), Contexte (**E**), Contrainte (**K**) et Pression (**P**) sur l'ensemble des participants disponibles et des tâches valides. D'après les conclusions de la section II.A, on peut émettre les propriétés suivantes :

- Si f présente de « fortes » valeurs de **T**, alors f est *réactive en situation de stress*.
- Si de plus, f présente de « fortes » valeurs de **K** et **P**, alors f est *sensible au stress émotionnel*.
- Si de plus, f présente de « faibles » valeurs de **A** et **E**, alors f est *spécifique au stress émotionnel*.

A l'échelle inter-individuelle, dire si les valeurs d'un estimateur sont « fortes » ou « faibles » revient à déterminer la *présence* ou *l'absence* d'un effet ; à partir des critères développés dans la section II.B. On parlera de *l'effet Tâche*, *l'effet Activité*, *l'effet Contrainte*, *l'effet Contexte* ou *l'effet Pression* sur la caractéristique f . En fonction des effets observés, la hiérarchie des attributs pour la caractéristique f est résumée dans le diagramme de la Figure 5.4. On sélectionnera en priorité les caractéristiques *sensibles* au stress émotionnel. On utilisera la *spécificité* comme un critère secondaire pour comparer les caractéristiques *sensibles* entre elles.

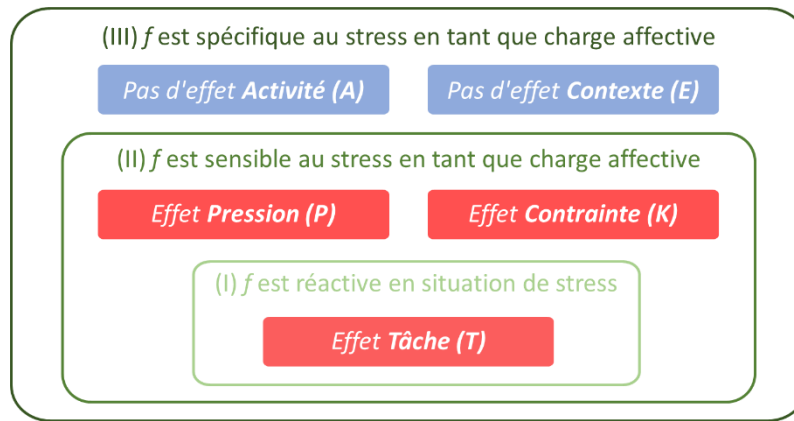


Figure 5.4 : Sensibilité et spécificité au stress d'une caractéristique f selon les 5 estimateurs T, A, K, E, P

En raison de la disparité du nombre d'exemples disponibles pour une caractéristique (voir section II.B), ou de la présence éventuelle d'effets parasites (voir section II.A), il est possible que l'estimateur Contrainte (K) et l'estimateur Pression (P) ne puissent pas satisfaire tous les deux le critère fort sur la présence d'un effet ($^{eff}p < p_{seuil}$). Afin de ne pas rejeter une caractéristique pour de mauvaises raisons, on instaure la règle suivante : il y a présence d'effet si au moins l'un des deux estimateurs valide le critère fort ($^{eff}p < p_{seuil}$) et que l'autre respecte au moins le critère faible ($|\vartheta| > \vartheta_{seuil}$) pour un estimateur, à condition que l'autre valide le critère fort ($^{eff}p < p_{seuil}$).

Ainsi, si l'on appelle \mathcal{F} l'ensemble des caractéristiques sélectionnées comme pertinentes pour estimer le stress :

$$f \in \mathcal{F} \Leftrightarrow \begin{cases} |\theta_P| > \theta_{seuil} \\ ^{eff}p_K(f) < p_{seuil} \\ ^{eff}p_T(f) < p_{seuil} \end{cases}, \quad \text{ou :} \quad \begin{cases} |\theta_K| > \theta_{seuil} \\ ^{eff}p_P(f) < p_{seuil} \\ ^{eff}p_T(f) < p_{seuil} \end{cases} \quad (5-6)$$

b) Cohérence inter-tâches

Lorsque les données de plusieurs tâches sont mises en commun, on maximise la quantité d'informations utiles à la détection d'un effet. Dans un premier temps, on cherchera donc des caractéristiques sensibles au stress pour l'ensemble des tâches *valides*, c'est-à-dire celles qui ont produit les effets attendus sur nos participants et que l'on a mises en évidence durant l'étape précédente. De cette manière, on met en évidence des caractéristiques pour lesquelles la « part du stress » (π_s) est bien *universelle*, c'est-à-dire qu'elle a lieu chez la plupart de nos participants. Implicitement, on a aussi supposé que cette contribution était *univoque*, c'est-à-dire qu'elle a toujours lieu dans le même sens (positif ou négatif) quel que soit le participant. Mais il faut encore montrer l'*unicité* de cette réponse, c'est-à-dire que celle-ci reste observable dans la majorité des situations stressantes. Pour cela, on a vérifié que l'effet recherché mesuré sur l'ensemble des tâches de l'ensemble \mathcal{V} reste observable pour *chaque* tâche de l'ensemble \mathcal{V} .

III. Résultats et Discussion

III.A. Validation du Protocole Expérimental

III.A.1. Effets des conditions expérimentales sur les EVAS et scores STAI

Le Tableau 5.4 ci-dessous présente les résultats de chaque tâche pour la mesure EVAS. Le Tableau 5.5 présente les mêmes résultats pour la mesure STAI. Les tâches sont affichées en ligne, les estimateurs sont affichés en colonne. Chaque estimateur X possède N_x exemples disponibles pour le calcul des trois indicateurs statistiques ($^{eff}p_x$, $^{abs}p_x$ et ϑ_x). Pour une tâche et un estimateur donnés, un résultat est constitué de trois éléments : (i) le critère (fort ou faible, présence ou absence d'effet) validé par les indicateurs statistiques, suivi (ii) de la taille d'effet ϑ_x et (ii) du nombre d'exemples disponibles N_x . La première ligne de chaque tableau marque le résultat « attendu » pour chaque tâche, c'est-à-dire le critère qui doit être respecté pour l'estimateur X suivi du signe de la taille d'effet ϑ_x .

Tableau 5.4 : Statistiques sur les estimateurs pour la sélection des tâches – Echelles Visuelles Analogiques de Stress (EVAS)

Dans chaque cellule : Critère validé | Taille d'effet | [Nombre d'exemples].

'***': forte présence d'effet – '*__': faible présence d'effet – 'o__': faible absence d'effet – 'ooo': forte absence d'effet.

EVAS	Tâche (T)	Contrainte (K)	Pression (P)	Activité (A)	Contexte (E)
Attendu	*** $\vartheta_T > 0$ [N_T]	*** $\vartheta_K > 0$ [N_K]	*** $\vartheta_P > 0$ [N_P]	ooo ϑ_A [N_A]	__ ϑ_E [N_E]
D2	*** +0,47 [20]	*** +0,53 [20]	o__ +0,30 [20]	*** +0,65 [20]	*__ +0,43 [20]
SECPT	o__ +0,29 [20]	*** +0,58 [20]	*** +0,60 [20]	*** -0,51 [20]	ooo +0,12 [20]
TSST	*** +0,69 [20]	*** +0,83 [20]	*** +0,68 [20]	ooo -0,17 [20]	*** +0,48 [20]
TA	*** +0,68 [20]	*** +0,82 [20]	*** +0,67 [20]	ooo -0,16 [20]	*** +0,48 [20]

Tableau 5.5 : Statistiques sur les estimateurs pour la sélection des tâches – Stait-Trait Anxiety Inventory (STAI-Y-A)

Dans chaque cellule : Critère validé | Taille d'effet | [Nombre d'exemples].

'***': forte présence d'effet – '*__': faible présence d'effet – 'o__': faible absence d'effet – 'ooo': forte absence d'effet.

STAI	Tâche (T)	Contrainte (K)	Pression (P)	Activité (A)	Contexte (E)
Attendu	*** $\vartheta_T > 0$ [N_T]	*** $\vartheta_K > 0$ [N_K]	*** $\vartheta_P > 0$ [N_P]	ooo ϑ_A [N_A]	__ ϑ_E [N_E]
D2	*** +0,56 [20]	*** +0,66 [20]	o__ +0,29 [20]	ooo +0,17 [20]	*__ +0,41 [20]
SECPT	o__ +0,29 [20]	*** +0,65 [20]	*** +0,63 [19]	*** -0,65 [19]	ooo -0,14 [19]
TSST	*** +0,78 [17]	*** +0,79 [17]	*** +0,76 [16]	ooo +0,13 [18]	*__ +0,38 [18]
TA	*** +0,80 [20]	*** +0,92 [4]	*** +0,91 [4]	o__ +0,00 [4]	*__ +0,38 [18]

En ce qui concerne le respect des critères (fort ou faible, présence ou absence d'effet), on utilise la symbolique proposée dans le Tableau 5.3 et qui permet de saisir d'un coup d'œil si le critère « attendu » a été validé ou non. La présence d'un effet est marquée par une ou plusieurs étoiles (*) et

l'absence d'effet par un ou plusieurs ronds (o). Le respect du critère faible ($\vartheta_X < \vartheta_{seuil}$) est marqué par un 1 (seul) de ces symboles ; et le respect du critère fort ($^{eff}p_X < p_{seuil}$ ou $^{abs}p_X < p_{seuil}$) est marqué par 2 symboles supplémentaires. Pour mémoire, le critère fort est plus conservatif que le critère faible ; et les critères faibles d'absence et de présence d'effet sont incompatibles entre eux. Dans chacun des deux tableaux, le non-respect d'un critère « attendu » est marqué par une écriture en **rouge**. Enfin, ces tableaux affichent en **gris** les résultats pour l'estimateur Contexte (**E**), qui ne peut pas servir de critère pour la sélection des tâches valides (justifications au paragraphe II.B.1).

III.A.2. Sélection des tâches valides

Les résultats pour la mesure EVAS (Tableau 5.4) montrent un effet Tâche significatif (*i.e.*, une augmentation du niveau de stress entre le repos et la tâche stressante), pour chaque tâche stressante excepté SECPT. De même, l'effet Pression (*i.e.*, une augmentation du niveau de stress entre la version Contrôle et Stress) est significatif pour toutes les tâches sauf D2. Les tâches D2 et SCEPT montrent par ailleurs une forte présence d'effet Activité : la version Contrôle de D2 a significativement augmenté le niveau de stress ; et la version Contrôle de SECPT a significativement diminué le niveau de stress des participants par rapport au repos. Ces trois estimateurs concordent pour invalider les tâches D2 et SECPT : d'après les EVAS, elles n'ont pas montré l'effet attendu sur l'état émotionnel des participants. De leur côté, les tâches TSST et TA valident tous les critères attendus : une forte présence d'effet Tâche, Contrainte et Pression ; et une forte absence d'effet Activité sur les niveaux de stress des participants.

Les résultats pour les questionnaires STAI (Tableau 5.5) confirment en substance ceux des mesures EVAS : une diminution du niveau de stress durant SECPT_c et une absence d'effet pour SECPT_s. La tâche D2 fait légèrement mieux que dans le Tableau 5.4 en montrant une absence d'effet durant D2_c. Mais selon l'estimateur Pression (P), la variation du niveau de stress entre D2_c et D_s n'est pas suffisante pour valider cette tâche dans l'opposition Stress / Non-stress qui est garantie par nos critères.

Malgré une taille d'effet quasi-nulle, on remarque que la tâche arithmétique (TA) échoue à satisfaire le critère fort sur l'absence d'effet Activité (**A**). Cela est dû à un nombre trop faible d'exemples disponibles pour cet estimateur et cette tâche ($N_A(TA) = 4$, dû au fait que des sujets n'ont pas rempli le test STAI à la fin de la session Contrôle). On peut donc rejeter le verdict du risque d'erreur ($^{abs}p_A(TA) < 0.1$) et prendre celui de la taille d'effet ($\vartheta_A(TA) > 0.3$) pour valider la condition Contrôle (TA_c) comme condition de non-stress. Ainsi, les mesures EVAS et STAI s'accordent pour valider TSST et TA et rejeter D2 et SECPT. Ainsi, dans la suite des analyses, on a fusionné les données des deux tâches « valides » (TSST et TA) pour sélectionner les caractéristiques les plus sensibles au stress.

III.A.3. Effet du Contexte sur la Charge Affective

Pour finir, les deux évaluations psychologiques s'accordent sur la présence d'un effet du Contexte ($|\vartheta_E| > 0.3$) pour les tâches D2 et TSST (et donc TA, qui a le même repos que TSST). Une explication à ce phénomène peut se trouver dans la condition Contrôle de SECPT, qui est la seule à avoir montré un effet *destressant* sur nos participants (STAI : $\vartheta_A(SECPT) = -0,65$). Comme toutes les tâches Contrôle étaient effectuées à la suite au cours d'une même session (voir Figure 4.1 p. 74), le sentiment de détente causé par SECPT_c a pu se prolonger durant le Repos qui suivait immédiatement (*i.e.*, durant D2_{Rc} ou TSST_{Rc} selon la personne car l'ordre des tâches était aléatoire). Ainsi, la dernière colonne du Tableau 5.4 et du Tableau 5.5 mesure un stress plus élevé dans la condition Rs que dans la condition Rc pour les tâches D2, TSST et TA. Ce résultat montre l'importance de considérer, comme dans le

scénario « réaliste » de la section II.A, que les niveaux de *pression* ne sont pas nécessairement équivalents dans toutes les conditions de *non-stress*.

III.B. Sélection des Caractéristiques

Les critères de sélection validés par chacune des 32 caractéristiques sont donnés en Annexe 7. Comme indiqué dans le paragraphe précédent, ces critères portent sur l'ensemble des données disponibles dans les tâches TSST et TA. Afin d'alléger notre discours, nous ne présenteront ci-dessous que les résultats obtenus sur les caractéristiques sélectionnées comme sensibles au stress.

III.B.1. Caractéristiques sensibles au stress

Les caractéristiques sélectionnées ici sont au nombre de 6 : trois cardiaques (la moyenne, l'écart-type et la RMS des différences successives du rythme cardiaque), deux respiratoires (l'autocorrélation et l'écart-type du signal RIP) ; et une électrodermale (la RMS de la dérivée du signal SC lent). Leurs résultats sont présentés dans le Tableau 5.6 ci-dessous, qui les classe par ordre décroissant de l'effet Pression. Cette table se lit de la même manière que le Tableau 5.4 et le Tableau 5.5 sur les évaluations subjectives. Les résultats de chaque caractéristique sont disposés en ligne, les estimateurs sont en colonne ; et chaque cellule affiche sur la même ligne : (i) le critère satisfait, (ii) la taille d'effet mesurée et (iii) le nombre d'exemples disponibles pour telle caractéristique et tel estimateur. La première ligne (« *Attendu* ») correspond aux critères qui déterminent une caractéristique *spécifique* au stress, c'est-à-dire (i) présentant un effet Tâche (T), Contrainte (K) et Pression (P) ; et (ii) associée à une absence d'effet Activité (A) et Contexte (E). Pour mémoire, on a sélectionné toutes les caractéristiques *sensibles* au stress, c'est-à-dire celles qui présentaient un effet Tâche (T), Contrainte (K) et Pression (P).

Tableau 5.6 : Caractéristiques sélectionnées – Synthèse des tâches valides

Dans chaque cellule : Critère validé | Taille d'effet | [Nombre d'exemples].

'***': forte présence d'effet – '*__': faible présence d'effet – 'o__': faible absence d'effet – 'ooo': forte absence d'effet.

Sélection ¹	Tâche (T)	Contrainte (K)	Pression (P)	Activité (A)	Contexte (E)
Attendu	*** ϑ_T [N _T]	*** ϑ_K [N _K]	*** ϑ_P [N _P]	ooo ϑ_A [N _A]	ooo ϑ_E [N _E]
<i>hrStd</i>	*** +0,86 [40]	*** +0,72 [40]	*** +0,78 [40]	*__ +0,32 [40]	*** -0,46 [40]
<i>hrMean</i>	*** +0,85 [40]	*** +0,71 [40]	*** +0,72 [40]	*** +0,66 [40]	ooo +0,13 [40]
<i>respAutocorr</i>	*** +0,81 [22]	*** +0,76 [22]	*** +0,65 [22]	*** +0,42 [22]	ooo +0,20 [22]
<i>hrRmssd</i>	*** +0,74 [40]	*** +0,51 [40]	*** +0,51 [40]	*__ +0,30 [40]	ooo -0,18 [40]
<i>respStd</i>	*** +0,81 [22]	*__ +0,36 [22]	*** +0,48 [22]	*** +0,80 [22]	ooo -0,20 [22]
<i>sclDerRMS</i>	*** +0,77 [40]	*__ +0,34 [40]	*** +0,46 [40]	*** +0,46 [40]	*__ -0,35 [40]

¹*hrMean* : rythme cardiaque moyen – *hrStd* : écart-type du rythme cardiaque – *hrRmssd* : RMS des différences successives de rythme cardiaque – *respAutocorr* : autocorrélation du signal respiratoire ; *respStd* : écart-type du rythme respiratoire
sclDerRMS : RMS de la dérivée du signal électrodermal à variations lentes.

Sur les six caractéristiques affichées ici, deux ont été sélectionnées sur la base du critère faible ($|\vartheta| < 0,3$) car elles montrent un effet Contrainte modéré (*respStd* et *sclDerRMS*, dernières lignes de la 2nde colonne). Les quatre autres montrent une forte présence d'effet Tâche, Contrainte et Pression.

On constate qu'un effet Activité (A) est présent pour chacune des caractéristiques sélectionnées. Le fort effet Tâche (T) visible dans la 3^e colonne traduit des variations suffisantes pour maintenir l'effet Contrainte (K) et l'effet Pression (P) au-dessus des seuils de sélection. Toutes les caractéristiques

sélectionnées sont donc *sensibles* au stress en tant que charge affective ; mais aucune ne lui est *spécifique*. Enfin, on remarque que l'effet Contexte (E) est absent pour quatre des six caractéristiques sélectionnées.

III.B.2. Congruence inter-tâches

Pour ces six caractéristiques, les résultats tâche par tâche est sont donnés dans le Tableau 5.7. Pour chaque caractéristique (en ligne) et chaque tâche (en colonne), le tableau affiche en ligne les critères satisfaits pour l'effet Tâche (T), l'effet Contrainte (K) et l'effet Pression (P). Afin de faciliter la comparaison des tâches et des caractéristiques entre elles, la taille d'effet ϑ et la taille d'échantillon N ne sont pas affichées. A titre de référence, la première ligne affiche les résultats pour les évaluations psychologiques (EVAS et STAI-Y-A).

Tableau 5.7 : Caractéristiques sélectionnées – Congruence inter-tâches.

Dans chaque cellule : Critère sur Tâche (T) | Critère sur Contrainte (K) | Critère sur Pression (P).

'***': forte présence d'effet – '*_': faible présence d'effet – 'o_': faible absence d'effet – 'ooo': forte absence d'effet.

Sélection ¹	TSST			TA			D2			SECPT		
	T	K	P	T	K	P	T	K	P	T	K	P
EVAS / STAI	***	***	***	***	***	***	***	***	o_	o_	***	***
<i>hrMean</i>	***	***	***	***	***	*_	***	ooo	*_	***	***	***
<i>hrStd</i>	***	***	***	***	***	***	o_	ooo	*_	***	o_	***
<i>hrRmssd</i>	***	***	***	***	***	***	o_	ooo	ooo	***	o_	o_
<i>respAutocorr</i>	***	***	***	***	***	***	***	ooo	ooo	ooo	*_	ooo
<i>respStd</i>	***	***	***	***	ooo	ooo	*_	*_	*_	o_	ooo	*_
<i>sclDerRMS</i>	***	***	***	***	ooo	*_	***	ooo	ooo	***	*_	***

¹*hrMean* : rythme cardiaque moyen – *hrStd* : écart-type du rythme cardiaque – *hrRmssd* : RMS des différences successives de rythme cardiaque – *respAutocorr* : autocorrélation du signal respiratoire ; *respStd* : écart-type du rythme respiratoire ; *sclDerRMS* : RMS de la dérivée du signal électrodermal à variations lentes.

Pour les deux dernières caractéristiques (écart-type du signal respiratoire et RMS de la dérivée du signal électrodermal), le critère fort ($p < p_{seuil}$) qui était satisfait pour l'ensemble {TSST + TA} ne tient plus pour la tâche TA séparément. L'effet Contrainte est absent pour les deux, et l'effet Pression est absent pour la caractéristique respiratoire (voir caractères en rouge, deuxième colonne). La contribution du stress à ces deux caractéristiques n'est donc pas *unique* : on ne peut pas la transposer d'une tâche à l'autre. Ces caractéristiques doivent être écartées sur la base du critère de congruence inter-tâches.

Par ailleurs, on constate qu'un effet Tâche (T) est presque toujours observé dans les quatre tâches expérimentales. Cependant, D2 et SECPT ont rarement montré la présence d'un effet Contrainte (K) ou Pression (P). De fait, aucune des caractéristiques du Tableau 5.7 n'aurait été sélectionnée si cette méthode avait été appliquée séparément à la tâche D2. On aurait sélectionné seulement deux d'entre elles (*hrMean* et *sclDerRMS*) pour la tâche SECPT. Ces résultats appuient l'hypothèse que les tâches D2 et SCEPT n'étaient pas valides pour sélectionner des caractéristiques sensibles au stress.

III.C. Apports du Protocole Expérimental

Parmi les caractéristiques fréquemment citées dans l'état de l'art (voir Chapitre Trois - II.C ou Tableau 5.8 plus bas), une seule a été conservée dans le Tableau 5.6 et Tableau 5.7 : le rythme

cardiaque moyen (*hrMean*). D'où vient ce rejet massif des caractéristiques « standard » par notre processus de sélection ?

III.C.1. Caractéristiques rejetées par nos critères

Le Tableau 5.8 ci-dessous compare les critères satisfaits pour les estimateurs Tâche (**T**), Contrainte (**K**) et Pression (**P**), par certaines caractéristiques déjà mises en évidence dans la littérature et qui n'ont pas été sélectionnées précédemment. Celles-ci sont disposées en lignes. La première colonne (« TSST + TA ») donne les résultats pour la synthèse des tâches TSST et TA, sur laquelle s'appuyait le processus de sélection des caractéristiques. Les colonnes 2 et 3 donnent les résultats pour les tâches TSST et TA considérées séparément. Chaque cellule affiche en ligne les critères satisfaits pour les trois estimateurs, comme dans le Tableau 5.7. Le détail des résultats présentés dans ce tableau, ainsi que les résultats pour l'effet Activité (A), sont donnés en Annexe 7.

Les résultats discordants avec nos critères de sélection des caractéristiques sont présentés en rouge dans ce tableau. L'absence d'effet Pression et Contrainte observée à presque toutes les lignes marque un fort effet de l'activité sur ces neuf caractéristiques ; et souligne la nécessité d'une condition Contrôle dans un protocole expérimental pour l'étude du stress. De même, les résultats discordants entre les deux tâches TSST et TA sont présentés en orange. Ces discordances sont observées pour six caractéristiques sur neuf et soulignent l'utilité d'un protocole multitâche pour l'étude du stress. Ces deux points sont abordés dans les paragraphes suivants.

Tableau 5.8 : Caractéristiques non sélectionnées par nos critères

Dans chaque cellule : Critère sur Tâche (T) | Critère sur Contrainte (K) | Critère sur Pression (P).
 '***': forte présence d'effet – '*_': faible présence d'effet – 'o_': faible absence d'effet – 'ooo': forte absence d'effet.
 En rouge : critères de non-sélection. En orange : effets discordants selon les tâches.

Caractéristiques Non-Sélectionnées ¹	TSST + TA			TSST			TA		
	T	K	P	T	K	P	T	K	P
Attendu	***	***	***	***	***	***	***	***	***
<i>ibiLF</i>	***	ooo	ooo	*_	o_	ooo	***	ooo	ooo
<i>ibiLFnorm</i>	***	*_	o_	*_	*_	ooo	***	*_	*_
<i>ibiHF</i>	ooo	ooo	*_	o_	ooo	ooo	***	ooo	***
<i>ibiHFnorm</i>	***	ooo	ooo	***	o_	ooo	***	o_	ooo
<i>ibiLF/HF</i>	***	o_	ooo	***	ooo	o_	***	***	o_
<i>ibiRmssd</i>	ooo	ooo	ooo	*_	o_	ooo	ooo	ooo	*_
<i>respRate</i>	***	ooo	ooo	***	*_	*_	***	ooo	ooo
<i>sclMean</i>	***	ooo	o_	***	ooo	ooo	***	ooo	o_
<i>scrAmpP50</i>	***	ooo	o_	***	ooo	ooo	***	*_	*_
<i>scrFreq</i>	***	ooo	ooo	***	ooo	ooo	***	ooo	ooo
<i>scrP85</i>	***	ooo	o_	***	ooo	*_	***	ooo	o_

¹*ibiLF* & *ibiLFnorm* : Puissance spectrale LF (resp. brute et normalisée) de l'intervalle inter-battements – *ibiHF* & *ibiHFnorm* : Puissance spectrale HF (resp. brute et normalisée) de l'intervalle inter-battements – *ibiLF/HF* : rapport des puissances spectrales LF et HF – *ibiRmssd* : racine du carré moyen de l'intervalle inter-battements – *respRate* : fréquence respiratoire – *sclMean* : moyenne de la composante électrodermale tonique – *scrAmpP50* : amplitude médiane des réponses électrodermales (SCRs) – *scrFreq* : fréquence globale des SCRs – *scrP85* : 85^e percentile du rythme instantané des SCRs.

III.C.2. Influence de l'activité sur la réactivité des caractéristiques

Sur l'ensemble de tâches {TSST + TA}, presque toutes les caractéristiques (à l'exception de la variabilité cardiaque haute fréquence *ibiHF*, *ibiRmssd*) valident le critère fort sur la présence d'effet Tâche (T). Le sens de leurs variations concorde avec celui généralement observé dans la littérature (voir le signe de ϑ dans l'Annexe 7). Cependant, aucune de ces caractéristiques n'a montré un fort effet Contrainte (K) ou Pression (P) (voir caractères en rouge) ; seulement deux d'entre elles montrent un effet modéré (*ibiLFnorm* pour K, *ibiHF* pour P). En particulier, l'absence d'effet Pression (P) associé à un fort effet Tâche (T) révèle que ce dernier n'a pas prédominé sur l'effet Activité (A) pour presque toutes les caractéristiques.

En d'autres termes : sur les neuf caractéristiques présentées dans le Tableau 5.8, sept ont montré une variation significative entre les conditions Stress et Repos. Malgré cela, aucune ne s'est montrée capable de distinguer un épisode de *stress* (entretien d'embauche + calcul mental avec pression sociale) d'une *activité* triviale et non-stressante (lire à voix haute + compter à l'envers). Ainsi, la condition Contrôle de notre protocole expérimental a permis de montrer que ces caractéristiques étaient davantage des corrélats de l'*activité* que des corrélats du *stress* en tant que charge affective. Ce constat est notamment flagrant pour les caractéristiques électrodermales « standard », comme la conductance moyenne (*sclMean*) et le nombre total de réponses (*scrFreq*). Que ce soit pour l'échantillon de sélection (TSST+TA) ou pour les deux tâches séparées, ces deux caractéristiques montrent un fort effet Tâche (T) mais une absence totale d'effet Contrainte (K) ou Pression (P).

III.C.3. Influence de la tâche sur la réactivité des caractéristiques

Pour les autres caractéristiques du Tableau 5.8, cependant, on remarque que cette absence d'effet Contrainte (K) ou Pression (P) n'est pas toujours observée simultanément pour les tâches TSST et TA (colonnes de droite). Certaines d'entre elles présentent même une forte disparité des résultats entre les 2 tâches (voir caractères orange). Par exemple, l'amplitude médiane des réponses électrodermales (*scrAmpP50*) est une des caractéristiques qui avait été mise en avant par les processus de sélection de Setz et collaborateurs (2010), dans leur tentative de discriminer le stress de la charge cognitive au moyen d'une tâche arithmétique. On retrouve à la fois un effet Contrainte (K) et un effet Pression (P) de notre propre tâche arithmétique (TA) pour cette même caractéristique. Dans une moindre mesure, l'autre caractéristique mise en évidence par leurs travaux (*scrP85*) présente un effet Pression (P) modéré pour TSST et TA malgré l'absence d'effet Contrainte (K). Mais si ces deux caractéristiques *locales* du signal électrodermal font mieux que les caractéristiques « standard », elles ne satisfont plus nos critères lorsque l'on s'appuie sur un protocole à deux tâches.

A nouveau, ces résultats montrent la pertinence d'un protocole multitâche pour mettre en évidence des corrélats physiologiques du stress. Par exemple, les discordances d'effets entre les tâches parmi les caractéristiques de variabilité cardiaques (ex : *ibiHF*, *ibiLF/HF*, *ibiRmssd*) suggèrent un changement de l'équilibre sympathovagal entre les tâches TSST et TA. Cela dit, il faut observer que les tâches TSST et TA ont été effectuées immédiatement à la suite sur un intervalle d'environ 10 minutes. Il est donc possible que les effets mesurés dans le Tableau 5.8 soient liés à une évolution de l'activité sympathique au cours d'un même épisode de stress. Plus que la tâche elle-même, le temps a pu jouer contre la sélection de ces caractéristiques connues de la littérature. L'absence de contrôle sur la dynamique des corrélats du stress constitue une première limite de la méthode de sélection proposée dans ce chapitre.

III.D. Limites de cette Etude

III.D.1. On ignore volontairement les liens entre les caractéristiques

Par ailleurs, une sélection de type *filtre* ne permet pas de mesurer le pouvoir *combiné* de certaines caractéristiques pour prédire l'état de stress. Certaines caractéristiques non sélectionnées ici pourraient se révéler plus efficaces dans un modèle non-linéaire, comme un séparateur à vastes marges. Pour des raisons évoquées au Chapitre Trois -, nous avons pris le parti d'utiliser des modèles en boîte-blanche pour modéliser le stress. Comme on le verra dans la Chapitre Six -, la plupart de ces modèles sont linéaires au niveau de la frontière de décision. Dans cette configuration, chaque caractéristique doit pouvoir séparer les deux classes à sa propre échelle.

Certaines parmi les caractéristiques sélectionnées peuvent mesurer les mêmes phénomènes. Dans cette étude, on s'est abstenu de poser des critères sur la corrélation entre les caractéristiques dans notre base de données. En théorie, la recherche d'une base orthogonale de caractéristiques pourrait permettre de couvrir des manifestations différentes du stress émotionnel. En réalité, la majorité des caractéristiques utilisées dans cette étude est partiellement corrélée avec d'autres. Pour s'en convaincre, le Tableau 5.9 ci-dessous affiche ce lien entre les quatre sélectionnées ici.

On observe une forte corrélation entre la moyenne du rythme cardiaque et son écart-type (*hrMean* & *hrStd* : 0,43), ainsi qu'entre cet écart-type et la RMS des écarts successifs (*hrStd* & *hrRmssd* : 0,66). Bien que *hrStd* et *hrRmssd* soient deux mesures de la dispersion du signal cardiaque, la part du calcul dans les corrélations obtenues est difficile à déterminer. Les caractéristiques cardiaques étant des corrélats connus de l'activité nerveuse autonome, il n'est pas surprenant qu'elles soient également corrélées entre elles.

Tableau 5.9 : Matrice de corrélation pour les quatre caractéristiques finalement sélectionnées

	<i>hrMean</i>	<i>hrStd</i>	<i>hrRmssd</i>	<i>respAutocorr</i>
<i>hrMean</i>	1	0,43	0,14	-0,03
<i>hrStd</i>	0,43	1	0,66	0,27
<i>hrRmssd</i>	0,14	0,66	1	0,11
<i>respAutocorr</i>	-0,03	0,27	0,11	1

Corrélation de Spearman calculée sur les données Stress et Contrôle des tâches TSST et TA, sur les 11 sujets disponibles pour la caractéristique *respautocorr*.

III.D.2. D'autres facteurs ont pu déterminer la sélection d'une caractéristique

Cela ne nous dispense pas de questionner l'origine des variations mesurées en section III.B. Par exemple, les caractéristiques respiratoires dépendent naturellement de la parole. Or, la parole est présente dans les deux tâches TSST et TA sélectionnées en section III.A. Ceci pourrait expliquer les performances de l'autocorrélation du rythme respiratoire (*respAutocorr*), qui pourrait détecter spécifiquement le stress *lié à la parole*. Cette hypothèse est confortée par les résultats de *respAutocorr* pour les tâches D2 et SECPT, tâches dans lesquelles la parole était absente : celles-ci ne présentent pas d'effet Contrainte (P) ou Pression (K) dans le Tableau 5.7.

De même, le rejet de toutes les caractéristiques électrodermales peut poser question. Malgré la proportion importante que celles-ci représentent dans la base de données initiale (19/32 : voir Tableau

4.2 page 74), aucune d'entre elles n'a terminé parmi les lauréates du Tableau 5.6. On peut trouver trois causes possibles à ce rejet, qui sont indépendantes de notre méthode de sélection. (i) Soit ces *caractéristiques* ne sont pas suffisamment fiables pour refléter un niveau de stress. (ii) Soit le *calcul* de ces caractéristiques n'était pas suffisamment fiable pour refléter un niveau de stress. (iii) Soit les *conditions* d'enregistrement n'ont pas permis de lier ces caractéristiques à l'état de stress.

Dans une étude basée sur les mêmes enregistrements, Ollander et collaborateurs (2016) ont montré que les caractéristiques électrodermales standard avaient un pouvoir prédictif plus important avec le capteur Empatica E4 qu'avec le capteur de laboratoire. Dans cette étude, le procédé d'extraction des caractéristiques était différent de celui présenté dans ce manuscrit. Dans une autre démarche de sélection (proche de celle présentée dans cette partie), un calcul encore différent (mais sur la même base de données) avait aussi conduit à rejeter ces caractéristiques (Vila et al., 2018). Ces résultats ne corroborent pas l'hypothèse d'un mauvais calcul des caractéristiques (ii) pour expliquer la non-sélection des caractéristiques électrodermales. Lors des sessions d'enregistrement, cependant, il a été constaté que les participants avaient la peau froide au niveau de l'extrémités des doigts. Cette influence de l'environnement (iii) pourrait expliquer l'absence de lien observé entre le stress et les caractéristiques électrodermales.

III.D.3. Le protocole expérimental n'a pas produit tous les effets attendus

Cette influence du contexte expérimental sur les variations physiologiques mesurées ne se résume peut-être pas à la seule température ambiante. Dans la section I.B, on supposait qu'une variation de la *charge affective* entre les conditions de Non-Stress (Rc, Rs, C), pouvait engendrer une variation parasite dans nos caractéristiques physiologiques. Or, les résultats de la section III.A (Tableau 5.4, Tableau 5.5) montrent qu'une variation du niveau de stress est observable entre les deux conditions Repos des tâches TSST et TA. Cette différence peut être à l'origine d'un « effet du contexte » sur l'estimateur Pression (P), comme on l'a montré au paragraphe II.A.2.

Cette problématique est traitée en Annexe 8, dans laquelle nos résultats expérimentaux sont analysés de sorte à montrer les deux propriétés suivantes. (i) Les contributions de l'environnement (ϵ), de l'activité (α) et de la pression (π) sont indépendantes lorsqu'on les mesure à l'échelle interindividuelle. (ii) L'estimateur Pression (P) pourrait être biaisé en estimant la « part du stress » dans l'évolution d'une caractéristique au travers de nos conditions expérimentales.

Enfin, notre sélection des caractéristiques n'a pu s'appuyer que sur deux tâches parmi les quatre initialement proposées. Les résultats du Tableau 5.7 montrent *a posteriori* qu'aucune des 32 caractéristiques de l'ensemble de départ n'aurait été sélectionnée, si nos critères avaient été appliqués séparément aux quatre tâches disponibles dans la base de données. Les tâches TSST et TA ont été distinguées sur la base des évaluations subjectives, car elles semblent répondre à une définition *consensuelle* du « stress », c'est-à-dire une activité durant laquelle la majorité des participants a ressenti une variation de son propre niveau de « stress ».

Ainsi, les caractéristiques sélectionnées dans ce chapitre sont des corrélats du « stress » au sens *académique* du terme, car on l'a induit au moyen de tâches expérimentales standard puis mesuré avec le questionnaire standard STAI-Y-A. Mais elles sont également des corrélats du « stress » au sens *vernaculaire* du terme, car chaque participant l'a mesuré verbalement sur sa propre échelle. Le rejet des tâches D2 et SECPT ne signifie pas nécessairement que celles-ci ne sont pas « stressantes » : l'étude

de leurs corrélats dans une conception plus large du stress peut constituer un axe de recherche futur sur cette base de données expérimentale.

IV. Synthèse : Principes et Apports de la Méthode de Sélection Proposée

Notre méthode de sélection s'est appuyée sur le raisonnement suivant. Une caractéristique est *réactive* si elle varie en situation de stress. Elle est *sensible* au stress si cette variation est liée à une charge affective. Elle est *spécifique* au stress si cette variation n'est pas liée aux facteurs confondants. Lors du processus de sélection sur les tâches TSST et TA, la plupart des caractéristiques s'est montrée *réactive* en situation de stress (*i.e.*, présence d'un effet pour l'estimateur **T**). Seulement six d'entre elles ont passé le critère de *sensibilité* (*i.e.*, présence d'un effet pour les estimateurs **P** et **K**). Finalement, aucune des caractéristiques *sensibles* au stress n'a passé le critère de *spécificité* (*i.e.*, absence d'effet pour les estimateurs **A** et **E**).

Dans la littérature, la majorité des méthodes de type *filtre* s'est contentée de sélectionner des caractéristiques *réactives* en situation de stress. Dans la présente analyse, le fait que nous n'ayons pas trouvé de caractéristique *spécifique* au stress souligne la nécessité de contrôler l'influence des facteurs confondants dès la campagne expérimentale – en particulier l'influence de *l'activité*. Comme on l'a montré au Chapitre Deux - IV, ignorer ce type de facteurs confondants revient à *redéfinir* implicitement la notion de « stress » en l'associant à celle d'activité. Notre recherche de caractéristiques sensibles au « stress » en tant que *charge affective* nous a permis d'éviter cet écueil, au prix d'une sélection finale restreinte (4 caractéristiques sur 32). Hormis le rythme cardiaque moyen, aucune caractéristique sélectionnées dans ce chapitre n'est fréquemment citée de l'état de l'art.

A notre connaissance, l'influence du *contexte* sur les caractéristiques physiologiques n'avait jamais été mesurée par une démarche de sélection. Dans la présente étude, notre estimateur Contexte (**E**) nous a permis de révéler un effet des tâches Contrôle sur la charge affective de nos participants. Pour prendre en compte ce type d'effet, nous avons pris soin de valider chaque mesure du stress émotionnel par une autre équivalente. Pour l'évaluation psychologique, on a utilisé conjointement les mesures EVAS et STAI. Pour la mesure physiologique, on a utilisé conjointement les estimateurs Pression (**P**) et Contrainte (**K**). Cette démarche de *validation des critères* est souvent absente dans les tentatives existantes pour modéliser le stress. Elle est pourtant essentielle, afin de contrôler l'incertitude qui accompagne toute mesure empirique du stress mental (Broek et al., 2009; Healey & Picard, 2005).

Notre analyse par *estimateurs* est moins contraignante que la recherche d'une matrice orthogonale de contrastes. Couplée à un modèle simple incluant notre concept-cible (le stress) et ses facteurs confondants (l'activité, l'environnement) dans l'évolution d'une caractéristique, elle nous a permis de saisir d'un coup d'œil les variations pertinentes au travers des conditions expérimentales. Ce type d'analyse peut être étendu à la recherche de n'importe quelle réponse physiologique noyée dans ses facteurs confondants – à condition de disposer d'un protocole expérimental adapté. Par ailleurs, le calcul de nos estimateurs ne demande pas d'hypothèse sur la distribution (gaussienne ou non) de nos données. On a ainsi pu faire appel à des tests statistiques non paramétriques, dans le but de mesurer puis comparer l'effet du stress (et celui des facteurs confondants) sur un ensemble de caractéristiques hétérogène.

Enfin, trois propriétés étaient nécessaires pour que nos tests statistiques identifient la présence d'une *charge affective* dans la réponse des caractéristiques au stress. Cette réponse devait être *unique* (i.e, s'applique à toutes les tâches), *universelle* (i.e, s'applique à tout le monde), et *univoque* (i.e, s'applique toujours dans le même sens). Dans le paragraphe III.B.2, on a pu remettre en question le caractère *unique* de la réponse au stress en distinguant nos tâches expérimentales. Cette démarche n'était accessible qu'avec l'appui d'un protocole expérimental multi-tâches. De la même manière, on pourrait questionner le caractère *universel* de cette réponse en distinguant des groupes de participants. Cela pourrait faire l'objet d'une recherche future, sur la base des questionnaires de personnalité qui ont été recueillis par ailleurs. Quant au caractère *univoque* de la réponse au stress, nous pensons que cette propriété est essentielle pour la conception d'un modèle de détection dans une approche *boîte-blanche*. Cette idée a motivé la conception de l'indice de stress, que nous présentons en détail dans le chapitre suivant.

A RETENIR

- Sur les 32 caractéristiques calculées dans le chapitre précédent, 4 ont su distinguer le stress d'une activité cognitive modérée sur les deux tâches TTST et TA. Ces caractéristiques sont la moyenne, l'écart-type et la RMS des différences successives du rythme cardiaque, ainsi que l'autocorrélation du signal respiratoire.
- Cette sélection des caractéristiques nous a conduits à rejeter la grande majorité des caractéristiques fréquemment citées dans la littérature. Ce résultat souligne l'intérêt de prendre en compte l'effet de l'activité lors de l'identification de marqueurs physiologiques du stress.

Chapitre Six - Modèle Adaptatif pour l'Identification du Stress Emotionnel

On dispose désormais d'un ensemble de caractéristiques sensibles à la charge affective, pouvant servir de base au développement d'un modèle interprétable du stress émotionnel. Sans connaissances *a priori* sur le lien entre nos mesures et l'état de stress, notre élaboration d'un modèle de stress doit adopter une démarche empirique basée sur l'apprentissage.

On cherche ici à concevoir un détecteur de stress qui reste fonctionnel avec des mesures de la vie quotidienne. Dans la littérature, la majorité des modèles sont construits sur un impératif de performance : on cherche à maximiser la précision obtenue sur les classes Stress et Non-Stress d'une base de données. Cependant, un modèle performant sur des données de laboratoire n'est pas nécessairement performant en conditions réelles. En particulier, le Chapitre Sept - montrera que les caractéristiques physiologiques peuvent montrer un fort taux d'erreur lorsque celles-ci sont estimées à partir de dispositifs de mesure portables.

Pour répondre à ce problème, nous sommes amenés à proposer un modèle différent des constructions standard, basé sur le potentiel discriminant de chaque caractéristique sélectionnée dans la partie précédente. Ce modèle est conçu de manière à rester flexible à un changement des caractéristiques disponibles, une propriété adaptative dont l'intérêt a été introduit à la fin du Chapitre Trois - . Un défi méthodologique se trouve dans la mesure des performances de ce modèle à partir de conditions expérimentales standard, lorsque son objectif principal est de s'adapter à un environnement de mesure dégradé par rapport aux enregistrements de laboratoire.

En sciences informatiques, on parle de *test fonctionnel* lorsqu'on cherche à savoir si un algorithme remplit la fonction désirée. Dans notre cas, cette fonction peut correspondre à la détection du stress sur la base de mesures physiologiques. Parallèlement, un *test de robustesse* cherche à savoir si l'algorithme remplit toujours cette fonction lorsqu'on lui fournit des entrées inattendues. Dans notre cas, ces entrées peuvent correspondre à des mesures physiologiques bruitées.

Par analogie, notre méthode d'évaluation se déroule en deux phases. Dans un premier temps, on mesure les performances de notre modèle de stress sur nos données de laboratoire, en comparaison à des modèles standard, afin de vérifier que ceux-ci remplissent la fonction attendue. Conformément aux objectifs de ce manuscrit, cette mesure des performances cherche à prendre en compte les facteurs confondants identifiés au Chapitre Deux - : l'effet de l'activité, l'effet de la tâche et l'effet du sujet. Dans un deuxième temps, on cherche à observer dans quelle mesure les performances de notre modèle sont conservées lorsque la qualité des mesures se dégrade. Dans cette démarche, nous serons amenés à introduire un « bruit » dans la valeur de nos caractéristiques physiologiques.

La section I du présent chapitre présente les bases méthodologiques qui permettent la production d'un modèle interprétable du stress ; et identifie les limites théoriques qui se posent lorsque ce modèle est exposé à des mesures bruitées. La section II développe les principes de notre modèle de stress, sa construction et sa méthode d'apprentissage. La section III détaille notre méthodologie de comparaison de modèles dans le but d'évaluer les capacités de ce nouvel algorithme. Enfin, la section IV présente les résultats obtenus et leur discussion.

I. Principes et Limites de l'Approche Standard pour la Production d'un Modèle Boîte-Blanche

I.A. Normalisation des Caractéristiques

Dans le Chapitre Trois -, on soulignait que le genre ou l'âge de la personne étaient susceptibles d'influencer le signal électrodermal (Boucein, 2012), ou la variabilité cardiaque (Acharya et al., 2006). De manière générale, une caractéristique physiologique est souvent liée aux caractéristiques individuelles. A titre d'exemple, la Figure 6.1 ci-dessous représente les quatre caractéristiques sélectionnées dans le Chapitre Cinq - pour deux participants différents : un sujet A (première ligne de graphiques) et un sujet B (deuxième ligne). Les probabilités de chaque caractéristique sont données pour trois classes : les 6 périodes de Repos (histogramme R sans remplissage, en pointillés), la version Contrôle des tâches TSST et TA (histogramme C en bleu) et la version Stress de ces mêmes tâches (histogramme S en rouge). Pour les quatre caractéristiques, les classes Stress et Contrôle sont bien séparables, comme on l'a remarqué dans le Chapitre Cinq -.

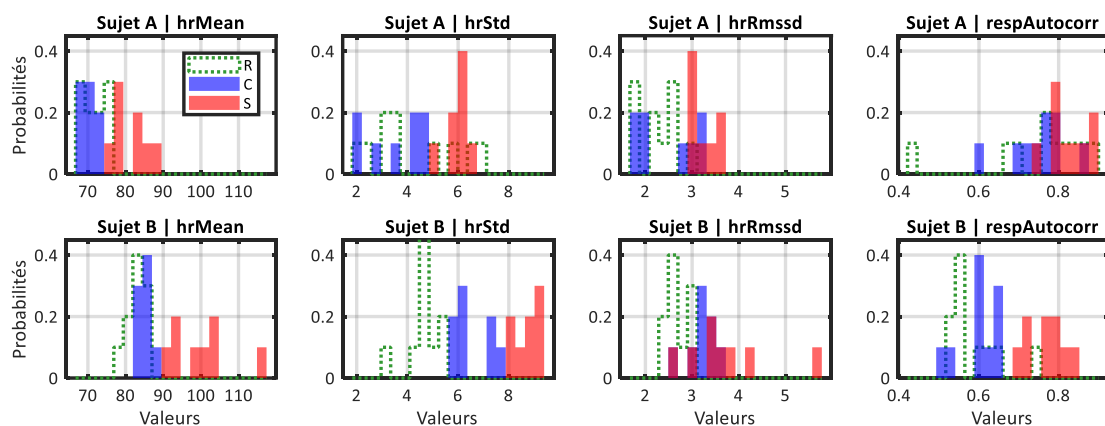


Figure 6.1 : Distribution des valeurs de quatre caractéristiques pour deux sujets différents en situation de Repos, Contrôle et Stress.

R : classe Repos (les 6 périodes) ; *C* : classe Contrôle (TSST et TA) ; *S* : classe Stress (TSST et TA).

hrMean : rythme cardiaque moyen ; *hrStd* : écart-type du rythme cardiaque ; *hrRmssd* : RMS des différences successives du rythme cardiaque ; *respAutocorr* : autocorrélation du signal respiratoire. Caractéristiques calculées sur des fenêtres de 60s.

Mais ici, les supports des histogrammes sont visiblement disjoints entre le sujet A et le sujet B. Pour la moyenne du rythme cardiaque (*hrMean*) et son écart-type (*hrStd*), le support de l'histogramme Stress du sujet A correspond au support de l'histogramme Contrôle du sujet B. En d'autres termes, sur le plan physiologique, l'état Stress du sujet A correspond à l'état Contrôle du sujet B du point de vue de nos quatre caractéristiques.

Comme on l'a suggéré au Chapitre Trois -, la littérature propose deux grandes approches pour répondre à ce problème de variabilité des caractéristiques en fonction des sujets : un modèle « mono-utilisateur » et un modèle « multi-utilisateur ». Ces approches sont comparées dans l'Annexe 3. L'option choisie dans le cadre de ce manuscrit est celle du modèle multi-utilisateur calibré : on utilise une *fonction de normalisation* \mathbf{v} qui projette les mesures brutes dans un référentiel commun à toutes les caractéristiques. Si f est une caractéristique quelconque et f_{τ} une mesure à l'instant τ pour un utilisateur u quelconque, on notera f_{τ}^v la valeur normalisée de f . Conformément aux standards de la littérature, on utilise un score Z paramétré sur les mesures au Repos :

$$f_{\tau}^v = \nu^{[u]}(f_{\tau}) = \frac{f_{\tau} - \mu_R^{[u]}}{\sigma_R^{[u]}} \quad (6-1)$$

Où μ_R et σ_R correspondent à la moyenne et à l'écart-type de f au repos pour l'utilisateur u . Le calcul des paramètres (μ_R , σ_R) implique de réaliser ces mesures au repos sur l'utilisateur u . Dans la suite de ce chapitre, on appellera *calibration* ce procédé mis en oeuvre pour neutraliser l'effet du sujet. Dans ce nouvel espace des entrées, les valeurs ne dépendent plus ni des caractéristiques, ni de l'utilisateur : on a projeté toutes les mesures dans un référentiel commun centré sur Zéro. Une fois chaque caractéristique calibrée par la formule (6-1), on dispose de vecteurs de mesures homogènes (globalement comprises entre -5 et 5, où se confondent plusieurs sujets. Ces données sont désormais formatées pour les algorithmes d'apprentissage.

I.B. Apprentissage en Boîte-Blanche

I.B.1. L'exemple de la régression logistique

Notre problème d'apprentissage possède deux classes : le Stress (**S**) et le Non-Stress (**NS**). On dispose pour cela d'un ensemble \mathcal{F} composé de K caractéristiques physiologiques f^k ($k \in \{1..K\}$). Sur une fenêtre de temps τ quelconque, on peut ainsi extraire un échantillon de K mesures : $\{f_{\tau}^k\}$. En apprentissage statistique, *détecter* le stress revient à calculer la probabilité de Stress à l'instant τ : $P_{\tau}(S)$, à partir des valeurs $\{f_{\tau}^k\}$ mesurées sur la personne. Dans cette perspective, les algorithmes d'apprentissage en boîte blanche s'appuient sur un modèle de la probabilité de stress en fonction des caractéristiques. Parmi les algorithmes usuels, la *régression logistique* utilise le modèle suivant.

$$P_{\tau}(S) = \frac{1}{1 + \exp\left(-\beta_0 - \sum_{k=1}^K \beta_k * f_{\tau}^k\right)} \quad (6-2)$$

On considère qu'une fenêtre de temps τ appartient à la classe Stress si $P_{\tau}(S) \geq \frac{1}{2}$; et à la classe Non-Stress dans le cas contraire. Le travail d'apprentissage consiste à ajuster les $K+1$ paramètres β_k , de sorte à maximiser le nombre d'échantillons bien classés dans leur catégorie. Pour comprendre à quoi correspond le modèle de l'équation (6-2), on peut décomposer ainsi la probabilité de stress :

$$P_{\tau}(S) = \mathcal{L}(\lambda_{\tau}), \quad \begin{cases} \mathcal{L}(x) = \frac{1}{1 + \exp(-x)} \\ \lambda_{\tau} = \beta_0 + \sum_{k=1}^K \beta_k * f_{\tau}^k \end{cases} \quad (6-3)$$

Ici, la *fonction logistique* \mathcal{L} compose une *fonction discriminante* λ qui fusionne les caractéristiques f^k au moyen d'une combinaison linéaire. L'équation : $P_{\tau}(S) = \frac{1}{2}$ équivaut à l'équation : $\lambda_{\tau} = 0$, qui dessine un hyperplan dans l'espace \mathcal{F} des caractéristiques. Ainsi, apprendre un détecteur de stress au moyen d'une régression logistique revient à insérer un hyperplan entre les classes Stress et Non-Stress dans l'espace \mathcal{F} des caractéristiques. Plus de détails sur l'algorithme sont proposés en Annexe 9.

D'autres algorithmes standard (ex. : analyse discriminante, algorithme bayésien naïf) estiment séparément les probabilités de Stress et de Non-Stress, à partir des mesures f_{τ}^k . A la différence de la régression logistique, ces méthodes font des hypothèses sur la distribution des échantillons f_{τ}^k pour

chaque classe (S ou NS). Une présentation plus détaillée des méthodes statistiques et une comparaison de leurs modèles est proposée en Annexe 9.

I.B.2. Limites et Perspectives pour la Conception d'un Modèle Robuste en Vie Réelle

Pour un modèle de stress, la performance mesurée sur les données d'apprentissage n'est pas un gage d'efficacité sur le terrain. Comme on l'a vu au Chapitre Trois - , les enregistrements recueillis dans la vie quotidienne peuvent dévier sensiblement des mesures accessibles en conditions contrôlées. Les mouvements corporels ordinaires, notamment, ainsi que l'utilisation de capteurs portables, altèrent significativement les mesures physiologiques.

Comme la régression logistique, les algorithmes d'apprentissage usuels cherchent à déterminer une frontière « optimale » entre les classes Stress et Non-Stress, dans l'espace des caractéristiques \mathcal{F} à K dimensions. En général, un modèle réalisé sur une base de données restreinte utilise « peu » de caractéristiques ($K = 4$ dans le présent chapitre). Ainsi, chacune d'elles porte une contribution potentiellement élevée sur le processus de décision. Lorsqu'une caractéristique « importante » observe un changement dans la distribution de ses mesures (ex. : bruit intense), celle-ci peut impacter d'autant plus fortement la performance du modèle en situant les échantillons du mauvais côté de la frontière de décision. Cette propriété est directement observable dans l'équation (6-3), lorsque le signe de la fonction discriminante λ détermine la classe en sortie du détecteur. L'influence de chaque caractéristique f^k ($\partial\lambda / \partial f^k$) est alors mesurée par le coefficient β_k associé à f^k .

On peut contrôler l'effet précédent en estimant *en ligne* la qualité de certaines mesures physiologiques : ce sujet sera abordé au Chapitre Sept - . Des indicateurs existent sur le signal cardiaque, comme celui fourni par la ceinture thoracique Zephyr BioHarness 3. Dans ce même dispositif, cela dit, ce type d'indicateur n'est pas disponible pour d'autres mesures comme la respiration. Par ailleurs, le lien entre la qualité de la mesure physiologique et la fiabilité des caractéristiques n'est pas clairement établi. En attendant de futures recherches dans ce domaine essentiel de la mesure en condition ambulatoire, on peut donc envisager deux cas de figure :

(a) En l'absence d'information pertinente sur la qualité des mesures, on peut augmenter la robustesse du modèle en limitant la contribution de chaque caractéristique sur la décision finale. Cette fonction peut être réalisée en appliquant une fonction saturante sur chaque caractéristique.

(b) En présence d'information pertinente sur la qualité des mesures, on peut augmenter la robustesse du modèle en retirant des calculs chaque caractéristique momentanément bruitée. Cette perspective implique de concevoir un modèle *flexible*, dans lequel on peut ajouter ou retirer des caractéristiques sans en modifier le fonctionnement.

Ces deux hypothèses sont à la base du modèle proposé dans la section II.A. Leur pertinence sera vérifiée dans la section III.D.

II. L'indice de Stress : un Modèle Adaptatif

L'indice de stress est un modèle conçu dans l'objectif de rester performant lorsque les conditions de mesure se dégradent. Les sections qui suivent en exposent les principes, la structure et la méthode d'apprentissage.

II.A. Principe Général

II.A.1. Des indices simples et un indice multiple

Pour chaque caractéristique f identifiée comme corrélat du stress, on considère une estimation « locale » de la probabilité de stress qui ne dépend que de f . Pour s'insérer dans une démarche d'apprentissage, la fonction \mathcal{S} qui réalise cette estimation locale peut prendre une forme logistique, sur le modèle de l'équation (6-3).

$$\mathcal{S} : \mathbb{R} \rightarrow]0; 1[, \quad f_\tau \rightarrow \mathcal{S}_\tau = \mathcal{L}(\lambda^{[f]}(f_\tau)) \quad (6-4)$$

Ici, λ est une fonction discriminante qui dépend de la seule caractéristique f ; et τ est un instant de mesure quelconque. Idéalement, \mathcal{S}_τ tend vers 1 lorsque f_τ correspond à une situation certaine de stress et \mathcal{S}_τ tend vers 0 lorsque f_τ correspond à une situation certaine de non-stress. Pour un ensemble de K caractéristiques f^k ($k \in \{1..K\}$), chacune mesurée à l'instant τ , on peut calculer un ensemble d'indices simples : \mathcal{S}_τ^k . Pour aboutir à une sortie unique, une fonction $\mathcal{M} :]0; 1[^K \rightarrow]0; 1[$ fusionne alors les indices simples sous la forme d'un *indice multiple*. La structure d'un tel modèle est illustrée dans la Figure 6.2.

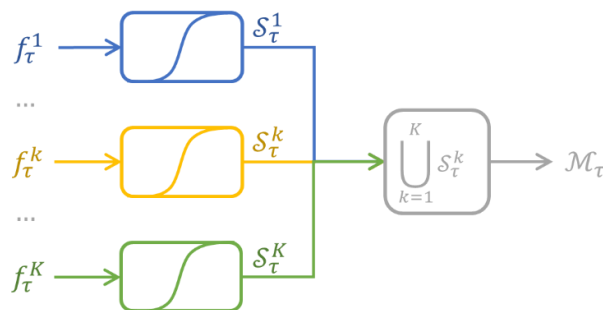


Figure 6.2 : Indices de stress : chaîne de traitements.

f_τ^k : mesure de la caractéristique f^k à l'instant τ ; \mathcal{S}_τ^k : Indice simple de la caractéristique f^k à l'instant τ ; \mathcal{M}_τ : indice multiple à l'instant τ .

Notre indice de stress est ainsi conçu pour présenter les deux propriétés suivantes :

a) Par sa forme *saturante*, l'indice simple \mathcal{S} limite l'influence de chaque caractéristique f sur la valeur de \mathcal{M} . Ainsi, une caractéristique présentant des valeurs aberrantes aura moins d'influence sur la décision du modèle que dans un détecteur standard de type régression logistique. L'indice de stress prévient ainsi la présence potentielle de bruit dans chaque caractéristique.

b) Par sa structure *modulaire*, l'indice multiple \mathcal{M} permet le retrait ou l'ajout d'une caractéristique à n'importe quel instant de mesure. Pour cela, il suffit de concevoir une fonction \mathcal{M} dont les paramètres ne dépendent pas des caractéristiques utilisées. Une manière simple de paramétrer cette fonction est d'attribuer un poids équivalent à chaque caractéristique f^k (voir section II.B). De ce fait, on limite aussi mécaniquement la contribution de chaque caractéristique sur la sortie finale.

II.A.2. Un niveau de confiance sur l'état de stress

D'un point de vue physiologique, l'indice simple \mathcal{S} évalue l'*allostase* liée au stress, reflétée par une caractéristique f pour un utilisateur u . Il répond au problème de détection du stress du « point de vue » de la caractéristique f . D'un point de vue formel, la fonction \mathcal{S} projette toutes les mesures f_τ dans un référentiel commun : un *indice de stress* compris entre 0 et 1, qui ne dépend ni de f ni de u .

Pour donner du sens à ce nouveau référentiel, on considère deux niveaux critiques : un *niveau de référence* \mathcal{S}_{Ref} et un *niveau de certitude* \mathcal{S}_{Cer} . Ceux-ci séparent l'ensemble des valeurs $\mathcal{S}_\tau \in]0 ; 1[$ en trois domaines distincts :

- i. Lorsque $0 < \mathcal{S}_\tau < \mathcal{S}_{\text{Ref}}$, le stress est *indélectable*. On considère que le stress est *indélectable* lorsque la mesure f_τ peut correspondre à une *valeur spontanée* de la caractéristique f .
- ii. Lorsque $\mathcal{S}_{\text{Ref}} \leq \mathcal{S}_\tau \leq \mathcal{S}_{\text{Cer}}$, le stress est *incertain*. On considère que le stress est *incertain* lorsque la mesure f_τ peut correspondre à un *facteur confondant* de notre base de données.
- iii. Lorsque $\mathcal{S}_{\text{Cer}} < \mathcal{S}_\tau < 1$, le stress est *certain*. On considère que le stress est *certain* lorsque la mesure f_τ ne peut correspondre ni à une valeur spontanée, ni à un facteur confondant.

L'indice \mathcal{S} reflète davantage un *niveau de confiance* sur l'état de stress qu'un niveau de probabilité. Comme on le verra plus bas, on cherche à s'extraire des procédures standard d'estimation probabiliste afin d'améliorer la robustesse du modèle fini. De son côté, l'indice multiple \mathcal{M} compile les « décisions » de chaque caractéristique sur l'état de stress. En substance, il propose une mesure unique de la *charge allostatique* liée au stress.

Toutes les caractéristiques sélectionnées dans le Chapitre Cinq - réagissent dans le même sens en situations Stress et Contrôle, mais plus intensément en situation de Stress. Cette propriété suggère de concevoir un indice \mathcal{S} strictement monotone, qui croît dans le sens de cette réaction au stress. Ainsi :

- Le niveau de référence \mathcal{S}_{Ref} correspond à une unique *mesure de référence* (f_{Ref}) ;
- Le niveau de certitude \mathcal{S}_{Cer} à une unique *mesure de certitude* (f_{Cer}).

Dans le cas où la caractéristique f réagit positivement au stress, on a : $\mu_R < f_{\text{Ref}} < f_{\text{Cer}}$. Dans le cas où la caractéristique f réagit négativement au stress, on a : $f_{\text{Cer}} < f_{\text{Ref}} < \mu_R$. Avec ces notations, la projection des trois niveaux de stress précédents (*indélectable*, *incertain*, *certain*) dans l'espace des valeurs de \mathcal{S} est illustrée sur la Figure 6.3.

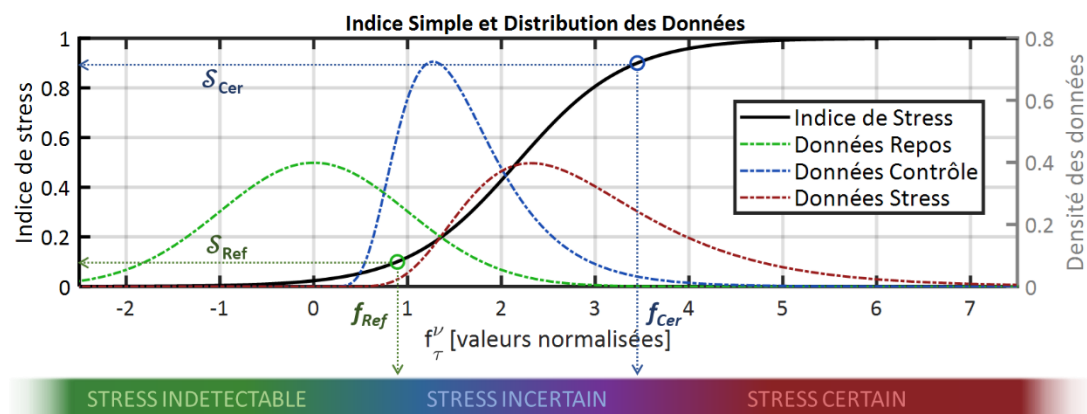


Figure 6.3 : Dimensions de l'indice de Stress.

Les lignes hachées représentent la distribution des données Repos, Contrôle et Stress (axe des ordonnées à droite).
La ligne pleine représente un indice \mathcal{S} généré par ces données (axe des ordonnées à gauche).

II.B. Construction de l'Indice de Stress et Synthèse des Paramètres

II.B.1. L'indice simple

Pour adapter notre indice à l'utilisateur, les mesures f_τ peuvent être normalisées par la fonction v : $f_\tau \rightarrow f_\tau^v$ que l'on a définie dans l'équation (6-1). L'indice \mathcal{S} est alors calculé au moyen de la fonction logistique \mathcal{L} et d'une fonction discriminante λ , sur le principe de l'équation (6-4).

$$\mathcal{S}_\tau^{[f,u]} = \mathcal{L}\left(\lambda^{[f]}(f_\tau^v)\right) = \mathcal{L} \circ \lambda^{[f]} \circ v^{[u]}(f_\tau) \quad (6-5)$$

La fonction \mathcal{L} est sans paramètres. La fonction λ dépend seulement de la caractéristique f , tandis que la fonction v dépend seulement de l'utilisateur u . On peut ainsi dimensionner l'indice \mathcal{S} sur deux jeux de données successifs :

1. Au travers de la fonction λ , l'indice \mathcal{S} est *appris* sur une base de données expérimentale, sans lien avec l'utilisateur final ;
2. Au travers de la fonction v , l'indice \mathcal{S} est *calibré* à partir de mesures au Repos sur l'utilisateur final.

Les paramètres de la fonction v ont été définis en section I.A ; il reste à définir ceux de la fonction λ . Pour cela, il faut considérer la contrainte suivante : l'indice simple \mathcal{S} est destiné à composer un indice multiple \mathcal{M} , qui comporte K caractéristiques. Si la fonction discriminante λ contient N paramètres, alors l'indice \mathcal{M} comportera au moins $N*K$ paramètres. Afin de conserver une approche *boîte-blanche* (et pour limiter le risque de surapprentissage), on suppose que λ prendra une forme affine (sur le modèle de l'équation (6-3)) : ainsi, elle comprendra seulement 2 paramètres (β_0, β_1). Comme λ et v sont deux fonctions affines, la composition de l'une par l'autre engendre une autre forme affine qui dépend à la fois de la caractéristique f et de l'utilisateur u .

$$\begin{aligned} \lambda^{[f]} \circ v^{[u]}(f_\tau) &= \left(\beta_0^{[f]} - \beta_1^{[f]} * \mu_R^{[u]} / \sigma_R^{[u]} \right) + \left(\beta_1^{[f]} / \sigma_R^{[u]} \right) * f_\tau \\ &= B_0^{[f,u]} + B_1^{[f,u]} * f_\tau \\ \mathcal{S}_\tau^{[f,u]} &= \mathcal{L}\left(B_0^{[f,u]} + B_1^{[f,u]} * f_\tau\right) \end{aligned} \quad (6-6)$$

En définitive, l'indice \mathcal{S} ne dépend que de deux paramètres : B_0 et B_1 . Ceux-ci sont partiellement *appris* sur des données expérimentales et partiellement *calibrés* sur l'utilisateur avant l'application du modèle. Au moment de la calibration du modèle, les coefficients (β_0, β_1) ont été *appris* sur des données expérimentales et font office d'hyperparamètres pour le calcul des coefficients B_0 et B_1 . Ces derniers sont alors entièrement déterminés par les paramètres (μ_R, σ_R), issus de mesures sur l'utilisateur au repos. De cette manière, la fonction affine : $[f_\tau \rightarrow B_0 + B_1 * f_\tau]$ réalise la détection du stress : du « point de vue » de la caractéristique f , « à l'échelle » de l'utilisateur u .

Comme on l'a suggéré plus haut, l'indice \mathcal{S} opère ainsi une double normalisation sur la caractéristique f . La première (v) projette les valeurs brutes de f dans un référentiel commun à tous les *individus* : c'est l'espace des valeurs « calibrées » de f . La seconde (λ) projette ces valeurs calibrées dans un référentiel commun à toutes les *caractéristiques* : c'est l'espace des « niveaux de confiance »

sur l'état de stress. Un niveau « local » de confiance sur l'état de stress peut être ainsi déterminé pour chaque caractéristique qui correspond à un corrélat physiologique du stress.

II.B.2. L'indice multiple

Il existe plusieurs manières pertinentes de combiner les K indices simples en un seul indice multiple. Si on considère l'indice \mathcal{S} comme une nouvelle *caractéristique*, on peut appliquer un algorithme d'apprentissage classique pour déterminer l'indice \mathcal{M} . Si on considère l'indice \mathcal{S} comme un *classifieur* à part entière, on peut appliquer un algorithme ensembliste pour déterminer l'indice \mathcal{M} (type *boosting*, forêts aléatoires...). Si on considère \mathcal{S} comme la *fonction de répartition* d'une variable floue, on peut appliquer les règles de la logique floue pour déterminer l'indice \mathcal{M} .

Certaines de ces idées seront reprises en fin de partie pour dessiner des perspectives à ce travail. Afin de simplifier la démarche proposée dans ce chapitre, on calculera l'indice \mathcal{M} en tant que *moyenne* des indices \mathcal{S} . Soit pour un utilisateur u quelconque et un ensemble de caractéristiques f^k :

$$\begin{aligned}\mathcal{M}_\tau &= \frac{1}{K} \sum_{k=1}^K \mathcal{S}_\tau^{[k,u]} \\ &= \frac{1}{K} \sum_{k=1}^K \mathcal{L} \left(B_0^{[k,u]} + B_1^{[k,u]} * f_\tau^k \right)\end{aligned}\tag{6-7}$$

Cette stratégie limite par défaut l'influence de chaque indice \mathcal{S} sur la sortie du modèle. C'est aussi la manière la plus simple pour concevoir un indice *flexible* : en paramétrant \mathcal{M} sans avoir recours à l'apprentissage, on peut ajouter ou retirer une caractéristique sans conséquence sur le fonctionnement du modèle. Il suffit en effet de transformer les poids $1/K$ de chaque indice \mathcal{S} en poids $1/K'$, où K' représente le nouveau nombre de caractéristiques incluses dans le modèle. Dans cette approche, on fait une sorte de pari sur la capacité de l'indice \mathcal{M} à apprendre (puis généraliser) sur les mesures physiologiques au travers de chaque indice \mathcal{S} .

II.C. Apprentissage

II.C.1. Principe

Lorsque les données d'une caractéristique ont été calibrées ($f_\tau \rightarrow f_\tau^v$), l'indice \mathcal{S} dépend seulement des deux paramètres $\{\beta_0, \beta_1\}$ qui déterminent sa fonction discriminante λ (voir II.B.1). Apprendre l'indice \mathcal{S} revient à fixer la valeur de ces deux paramètres.

Afin de donner un sens à cet apprentissage, il faut revenir à la représentation graphique de l'indice \mathcal{S} proposée dans la Figure 6.3. L'indice \mathcal{S} est caractérisé par deux points qui définissent sa « zone d'incertitude » : $(f_{\text{ref}}^v, \mathcal{S}_{\text{ref}})$ et $(f_{\text{cer}}^v, \mathcal{S}_{\text{cer}})$. D'un point de vue formel, *apprendre* les 2 paramètres $\{\beta_0, \beta_1\}$ revient à *interpoler* l'indice \mathcal{S} aux 2 points $(f_{\text{ref}}^v, \mathcal{S}_{\text{ref}})$ et $(f_{\text{cer}}^v, \mathcal{S}_{\text{cer}})$.

En fixant arbitrairement les ordonnées $\{\mathcal{S}_{\text{ref}}, \mathcal{S}_{\text{cer}}\}$, l'allure de l'indice \mathcal{S} dépendra seulement des deux abscisses $\{f_{\text{ref}}^v, f_{\text{cer}}^v\}$, qui sont des *seuils* sur la caractéristique f . Pour que l'apprentissage de \mathcal{S} corresponde à sa définition proposée en II, on a fixé ces seuils de la manière suivante.

- 1) Un **détecteur Stress / Repos** détermine le *seuil de référence* f_{ref} en posant : $P(\mathcal{S}|[f = f_{\text{ref}}]) = 10\%$.
- 2) Un **détecteur Stress / Contrôle** détermine le *seuil de certitude* f_{cer} en posant : $P(\mathcal{S}|[f = f_{\text{cer}}]) = 90\%$.

En pratique, ces deux détecteurs correspondent à des régressions logistiques sur les données de chaque classe. La régression Stress / Repos détermine l'état de « stress indétectable ». La régression Stress / Contrôle détermine l'état de « stress certain ».

II.C.2. Interprétation

Il existe un lien simple entre ces calculs et la définition de \mathcal{S} proposée au paragraphe II. Par convention, on a considéré qu'une probabilité est « faible » lorsqu'elle est inférieure à 10% ; et qu'elle est « forte » lorsqu'elle est supérieure à 90%. Lorsqu'une probabilité de Stress estimée par régression logistique prend des valeurs « fortes » ou « faibles », elle est toujours « forte » ou « faible » *par rapport* à la probabilité du Non-Stress (voir Annexe 9 - partie c)). Ainsi :

- 1) Pour $f \leq f_{Ref}$, la probabilité d'un épisode de stress (classe Stress) est « faible » devant la probabilité que f_t exprime une variation spontanée (classe Repos). L'état de stress est *indétectable*.
- 2) Pour $f \geq f_{Cer}$, la probabilité d'un épisode de stress (classe Stress) est « forte » devant la probabilité que f_t exprime un facteur confondant (classe Contrôle). L'état de stress est *certain*.

Une fois ces seuils connus, on peut assimiler localement l'indice de stress aux probabilités estimées par nos deux détecteurs : $\mathcal{S}_{Ref} = P(S|f=f_{Ref}) = 10\%$; et $\mathcal{S}_{Cer} = P(S|f=f_{Cer}) = 90\%$, comme sur la Figure 6.3. Dans l'espace des données calibrées, on construit alors l'indice \mathcal{S} par interpolation sur les points : $(f_{Ref}^y, \mathcal{S}_{Ref})$ et $(f_{Cer}^y, \mathcal{S}_{Cer})$. Le lien entre ces points et les paramètres $\{\beta_0, \beta_1\}$ est développé dans l'Annexe 10, où l'on montre également que la classe Repos a une utilité concrète pour la robustesse l'indice \mathcal{S} .

A partir de la formule (6-7), l'apprentissage d'un indice simple pour chaque caractéristique donne accès à l'indice multiple. Avec ce choix d'apprentissage, cependant, il n'existe pas de seuil « par défaut » pour lequel cet indice \mathcal{M} permettra de distinguer le Stress du Non-Stress, comme dans une régression logistique (ex : $\mathcal{M}_{seuil} = 50\%$). Pour que l'indice \mathcal{M} fonctionne à son tour comme un *détecteur* de stress, ce seuil reste à déterminer empiriquement sur notre base de données.

III. Comparaison de Modèles pour la Détection du Stress

Dans la suite de ce chapitre, on mesure la *performance* de l'indice de stress (*i.e.*, sa capacité à distinguer les épisodes de stress des épisodes de non stress) et sa *robustesse* (*i.e.*, sa capacité à détecter le stress malgré la présence de bruit) en comparaison avec des algorithmes usuels.

III.A. Question de Recherche et Méthodologie Générale

On cherche à vérifier si notre indice \mathcal{M} peut détecter le stress avec des performances similaires à celles des algorithmes fréquemment rencontrés dans la littérature. Pour cela, on peut mettre l'indice de stress en concurrence avec des modèles plus ou moins complexes et plus ou moins interprétables sur l'état de stress. Dans cette démarche, on ne peut s'attendre à ce que les performances de l'indice de stress surpassent celles d'un modèle plus ajusté aux données d'apprentissage. En effet, la forme contrainte de l'indice \mathcal{M} (une moyenne de sigmoïdes) ne lui permet pas de dessiner une frontière « optimale » entre les classes Stress et Non-Stress dans l'espace des caractéristiques.

L'indice de stress est construit sur l'hypothèse suivante : en transférant le travail d'apprentissage sur plusieurs caractéristiques identifiées comme des corrélats du stress, notre modèle se montrera

plus *robuste* qu'un modèle standard lorsque la qualité des mesures vient à diminuer. Pour vérifier concrètement l'intérêt de notre modèle, il faut donc répondre à la question suivante :

L'indice de stress conserve-t-il ses performances dans un environnement de mesures dégradé ?

Une manière de répondre à cette problématique est d'introduire artificiellement du bruit dans l'une des caractéristiques utilisées par l'indice de stress. Sur ces mesures « volontairement dégradées », la *robustesse* de l'indice \mathcal{M} peut être comparée à celle d'un modèle appris avec un algorithme standard. Dans cette perspective, on peut imaginer deux cas de figure.

- a) Sans information pertinente sur la qualité des signaux physiologiques, toutes les caractéristiques ont été conservées pour le calcul de l'indice \mathcal{M} .
- b) Une information pertinente sur la qualité des signaux physiologiques a permis de retirer la caractéristique bruitée du calcul de l'indice \mathcal{M} .

Dans le premier cas, on mesure la résilience du modèle au bruit introduit par une caractéristique sujette à des artefacts. Dans le second cas, on mesure la résilience du modèle au retrait d'une source d'information sur l'état de stress.

Ainsi, les résultats présentés dans ce chapitre peuvent se décomposer en deux « expériences » successives. La première expérience consiste à mesurer les performances de modèles standards en comparaison à l'indice stress, à partir de données « idéales » qui correspondent à nos enregistrements effectués en conditions contrôlées. C'est l'objet de la section III.C. La seconde expérience consiste à mesurer les performances de l'indice de stress en comparaison avec un modèle standard, à partir de données « dégradées » qui simulent une situation envisageable en vie réelle : celle où une caractéristique physiologique ne reflète plus son paramètre biologique cible (ex. : rythme cardiaque moyen) sur une période prolongée. C'est l'objet de la section III.D. Mais pour s'inscrire dans les objectifs de ce manuscrit, la mesure des performances de chaque modèle doit prendre en compte les trois facteurs confondants identifiés au Chapitre Deux - IV : *l'effet de l'activité, l'effet de la tâche et l'effet du sujet*. C'est l'objet de la section III.B.

III.B. Mesure des Performances d'un Détecteur de Stress

III.B.1. Données support

a) Caractéristiques utilisées

Au Chapitre Cinq - , 25 caractéristiques se sont montrées *réactives* en situation de stress (*i.e.* fort effet Tâche), 6 ont été *sensibles* au stress en tant que charge affective (*i.e.* fort effet Contrainte / Pression) ; et 4 ont été finalement retenues pour montrer une réponse *unique* entre les tâches TSST et TA. On s'intéresse à l'ensemble des échantillons calculés pour ces quatre caractéristiques, sur des fenêtres de temps de 60s. Pour être cohérente avec le Chapitre Cinq - , notre démarche de classification s'est intéressée uniquement aux tâches expérimentales TSST et TA.

Lors des analyses à venir, chaque algorithme d'apprentissage est appliqué sans filtre sur les 4 caractéristiques retenues dans ce chapitre (*hrMean*, *hrStd*, *hrRmssd*, *respAutocorr*). Ce choix de méthode est justifié par le fait que ces caractéristiques sont peu nombreuses. Si nécessaire, nos algorithmes auront donc peu de difficulté à mettre en oeuvre leur propre sélection, en diminuant spontanément la contribution des (1 ou 2) caractéristiques inutiles au fonctionnement des modèles.

b) Echantillons disponibles

Nos classes Stress et Contrôle sont formées des échantillons disponibles dans les versions correspondantes des tâches TSST et TA. Ici, l'ensemble des participants « disponibles » est constitué des personnes sur lesquelles on a pu enregistrer à la fois le signal cardiaque et le signal respiratoire – soit 11 personnes au total. Le nombre d'échantillons typiquement utilisable est de 4 par participant, par tâche (TSST, TA) et par condition (Stress, Contrôle). L'ensemble des échantillons disponibles en condition Stress constitue la classe Stress – et de même pour la classe Contrôle. Enfin, il est nécessaire d'utiliser une classe Repos pour apprendre l'indice \mathcal{M} . Contrairement aux mesures de type Stress ou Contrôle, les mesures au Repos ne dépendent pas de la tâche qui suit. En revanche, elles peuvent dépendre du jour de mesure. Pour éviter cet effet du contexte durant l'apprentissage de \mathcal{M} , la classe Repos a été formée à partir des mesures qui précèdent la tâche TSST en version Stress (TSST_{RS}).

c) Pré-traitement des données

A l'intérieur de chaque classe, on a retiré les valeurs aberrantes sur un critère de déviation absolue. Chaque échantillon f_τ dont la déviation absolue à la moyenne dépassait trois fois la déviation absolue moyenne de la caractéristique (formellement : $|f_\tau - E(f)| > 3 * E(|f - E(f)|)$, où E est l'espérance), a été identifié séparément. Pour chacun des échantillons repérés, la fenêtre de temps correspondante a été retirée de la matrice des caractéristiques. A l'issue de ces traitements, la taille d'échantillon conservée pour chaque classe et chaque tâche est affichée dans le Tableau 6.1.

Tableau 6.1 : Taille de la base d'apprentissage pour chaque tâche et chaque classe

	TSST	TA
Stress	42	44
Contrôle	44	44
Repos	55	

Pour finir, on a *calibré* (normalisé) les mesures de chaque participant au moyen des mesures de repos disponibles pour chacun d'eux. Pour choisir les données de calibration, on peut s'inspirer des résultats du Chapitre Cinq - III.B et de l'Annexe 8.

- La plupart des caractéristiques ont montré des valeurs différentes entre les différents repos. Ainsi, il est pertinent d'utiliser des mesures prises sur plusieurs jours différents pour la calibration.
- Durant la session Contrôle, l'enchaînement des tâches a pu avoir un impact sur les paramètres biologiques des participants lors des repos correspondant aux tâches D2, TSST et TA.

On a ainsi sélectionné les valeurs du premier Repos de chaque session Stress (D2_{RS}, SECPT_{RS} et TSST_{RS}) et celles du repos précédant la version Contrôle du SECPT (SECPT_{RC}). A partir des paramètres (μ_R , σ_R) obtenus sur ces mesures pour chaque participant, les données ont été calibrées selon le procédé décrit en section I.A.

III.B.2. Validation Croisée

a) Motivations

Dans le Chapitre Deux - et le Chapitre Trois - , l'identification de facteurs confondants du stress nous a conduit à définir trois critères pour mesurer l'efficacité d'un algorithme d'apprentissage.

- a. Le modèle distingue une activité stressante d'une activité contrôle.
- b. Le modèle est efficace sur un participant inconnu.

c. Le modèle est efficace sur une tâche inconnue.

Dans notre base de données, le critère (a) est assuré par notre condition Contrôle. Le critère (b) peut être assuré par une validation croisée de type *LOSO*, pour : *Leave-One-Subject-Out*. Le critère (c) peut être assuré par une validation croisée de type *LOTO*, pour : *Leave-One-Task-Out* (Sakri et al., 2018). Un modèle de stress montrant de bonnes performances avec ces deux derniers procédés s'attribue un caractère *unique* et *universel*, à l'image de la « part du stress » recherchée au Chapitre Cinq - . Dans la présente étude, on cherche à ce que ces caractères soient vérifiés simultanément.

b) Principe

Notre protocole de validation s'inspire simultanément des méthodes *LOSO* et *LOTO*. Celui-ci consiste à apprendre le modèle sur l'une des 2 tâches disponibles (TSST ou TA) ; ainsi que sur tous les participants disponibles, sauf un. On valide alors le modèle sur la tâche et le sujet non retenus pour l'apprentissage. En référence aux méthodes standard, on pourrait appeler ce procédé **LOSOTO** (pour *Leave-One-Subject-and-One-Task-Out*). Il permet de contrôler simultanément *l'effet du sujet* et *l'effet de la tâche*. Une illustration du processus est proposée dans la Figure 6.4. A notre connaissance, ce procédé de validation n'a jamais été proposé dans l'évaluation du stress humain.

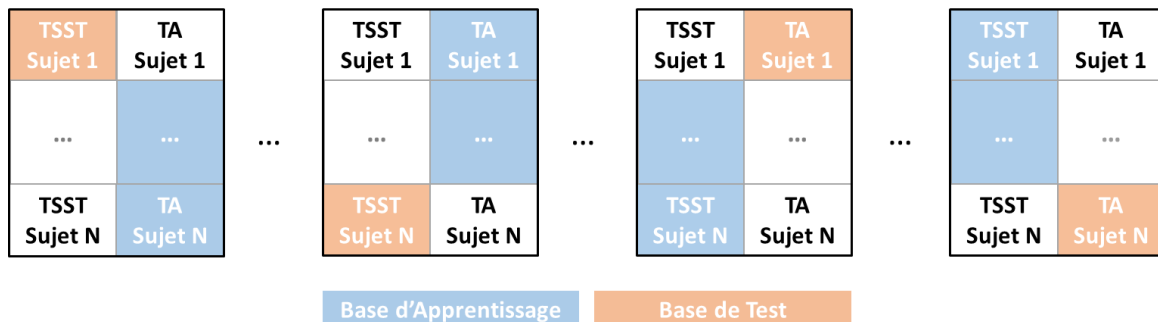


Figure 6.4: Division des données pour la validation croisée LOSOTO

Cette méthode est conçue pour *mesurer* la performance d'un algorithme ; et non pas pour *l'apprendre* de manière efficace. En effet, le sujet et la tâche « laissés-en-dehors » ne peuvent être utilisés pendant la phase d'apprentissage. Afin de limiter la confusion dans les paragraphes qui suivent, le terme « base d'apprentissage » se réfèrera spécifiquement aux cases en **bleu** à chaque étape de la Figure 6.4 ; et le terme « base de test » désignera la case **orange**. Au sein de la base d'apprentissage, un algorithme pourra isoler sa propre « base de validation » pour protéger le modèle final du surapprentissage.

c) Application sur la base de données

Ce découpage des données concerne seulement les classes Stress et Contrôle ; car la classe Repos n'est pas soumise à *l'effet de la tâche*. Pour obtenir un nombre équilibré d'exemples dans chaque classe d'apprentissage, on peut utiliser tous les échantillons de la classe Repos disponibles pour un même sujet. Si l'on nomme i une tâche quelconque et k un participant quelconque de notre base de données, le nombre d'exemples disponibles à l'étape numéro (i,k) de la validation croisée est affiché dans le Tableau 6.2.

Tableau 6.2 : Taille des ensembles d'apprentissage et de test à chaque étape du protocole LOSOTO

Etape	Contenu de l'ensemble	Nombre type d'exemples
Apprentissage	Les données Stress et Contrôle de tous les participants sauf k ; et de toutes les tâches sauf i + Les données Repos de tous les participants sauf k	40 Stress 40 Contrôle 50 Repos
Test	Les données Stress et Contrôle du participant k et de la tâche i	4 Stress 4 Contrôle

Dans une démarche classique, on calculerait directement les performances du modèle sur l'échantillon de test disponible à chaque étape du processus. Ce serait peu pertinent dans notre cas, où la base de données nous donne accès à 8 échantillons en moyenne pour calculer ces performances. Ainsi, la méthode choisie dans ce chapitre consiste à prédire la *sortie* du modèle pour les 8 échantillons de test ; puis à rassembler les prédictions issues de chaque étape.

On peut schématiser ce processus à l'aide de l'Algorithme 1 proposé dans l'Annexe 11. Dans la section IV.A, la conservation des valeurs prédites par l'indice \mathcal{M} sur chaque base de test nous permettra de déterminer un seuil de détection pour ce modèle de stress.

A partir de l'ensemble des prédictions retenues, on peut calculer un chiffre unique de performance sur l'ensemble des données de test : ce chiffre correspond au pourcentage d'échantillons bien classés dans leur catégorie. L'inconvénient de cette méthode est qu'on ne dispose pas d'une mesure de *dispersion* sur la performance mesurée par validation croisée. Ainsi, il sera difficile de juger une *différence* de performances entre deux modèles : cette différence a-t-elle un sens d'un point de vue statistique ? Pour répondre à cette question, une méthode simple consiste à calculer un intervalle de confiance à 95% ($IC^{95\%}$) sur la performance du détecteur à l'aide de la formule suivante⁶ :

$$IC^{95\%} = Perf \pm 1,96 * \sqrt{\frac{Perf * (1 - Perf)}{N}} \quad (6-8)$$

Ici, la largeur de l'intervalle ne dépend que de la performance mesurée (**Perf**, *i.e.* le taux de bien classés) et de la taille d'échantillon **N** sur lequel la performance a pu être mesurée (ici N=174, exemples des classes Stress et Contrôle). A défaut de calculer une dispersion réelle, cet intervalle nous permet d'estimer si la performance d'un modèle peut être « statistiquement » différente d'une autre, compte tenu du nombre d'échantillons qui ont servi à les mesurer.

III.C. Expérience 1 : Comparaison d'Algorithmes pour la Détection du Stress en Milieu Contrôlé

Sur la base du processus LOSOTO, on peut ainsi comparer les performances en détection de l'indice \mathcal{M} à celles de modèles standard. Pour cette démarche, on a sélectionné deux techniques fréquemment utilisées pour détecter le stress : la régression logistique et les réseaux de neurones.

⁶ Ce calcul revient à considérer que le phénomène « bien classé » est une variable de Bernoulli (*i.e.* succès VS échec) ; et que le taux de « bien classés » sur notre base de données est une estimation de la performance réelle du détecteur de stress. L'intervalle de confiance ainsi calculé est l'intervalle de Wald (Agresti & Coull, 1998).

III.C.1. Régressions logistiques

Une régression logistique à 2 classes sur nos $K=4$ caractéristiques est l'une des méthodes les plus simples pour prédire les classes Stress et Contrôle sur notre base de données (voir Annexe 9). On peut alors comparer la sortie de ce modèle \mathcal{R}^2 à celle de l'indice \mathcal{M} sur ces mêmes classes. Dans cette démarche, cependant, l'indice de stress possède un avantage potentiel sur le modèle \mathcal{R}^2 : il inclut de l'information sur la classe Repos, ce qui procure de la stabilité à la sortie du modèle (voir Annexe 10) et pourrait ainsi augmenter ses capacités de généralisation par rapport à \mathcal{R}^2 .

Pour compenser un éventuel déséquilibre entre les deux algorithmes, on peut comparer les performances de l'indice \mathcal{M} à celles d'une régression logistique à trois classes qui inclut les données de la classe Repos. Les fondements d'une régression à N classes sont développés en Annexe 9 (partie c)). Dans ce nouveau modèle \mathcal{R}^3 , on calcule la probabilité de Stress en s'appuyant sur deux régressions à 2 classes : Stress / Contrôle et Stress / Repos. Cela multiplie par autant le nombre de paramètres utilisés par l'algorithme. Si K est le nombre de caractéristiques utilisées pour détecter le stress, alors :

- L'indice moyen \mathcal{M} utilise $2*K = 8$ paramètres ;
- La régression \mathcal{R}^2 (Stress, Contrôle) utilise $K+1 = 5$ paramètres ;
- La régression \mathcal{R}^3 (Stress, Contrôle, Repos) utilise $2*(K+1) = 10$ paramètres.

En sortie des modèles \mathcal{R}^2 et \mathcal{R}^3 , on peut détecter la classe Stress lorsque celle-ci est *plus probable* que le Non-Stress (classes Repos et Contrôle), soit : $P(S) > \frac{1}{2}$. Le seuil de l'indice \mathcal{M} , en revanche, n'a pas de valeur par défaut : celle-ci sera déterminée à l'issue de la validation croisée.

III.C.2. Réseaux de neurones

a) Principe

Dans les trois modèles précédents, la frontière de décision entre les classes Stress et Contrôle est fortement contrainte par la forme logistique utilisée pendant l'apprentissage. En s'affranchissant d'un modèle explicite de la probabilité de stress (comme le fait la régression logistique dans l'équation (6-2)), l'utilisation d'un réseau de neurones artificiel permet de dépasser cette contrainte pour mieux s'ajuster aux données d'apprentissage.

Le réseau que l'on propose dans ce chapitre possède une couche cachée de taille variable (x unités) et une couche de sortie de 2 unités (soit 1 par classe : Stress et Contrôle). Les neurones de la couche cachée utilisent une fonction de transfert de type tangente hyperbolique. Ceux de la couche de sortie utilisent la fonction Softmax pour normaliser la probabilité de chaque classe. D'une couche à l'autre, la rétropropagation se fait par la méthode du gradient conjugué. L'illustration d'un réseau de neurones à 3 unités dans la couche cachée est proposée dans la Figure 6.5.

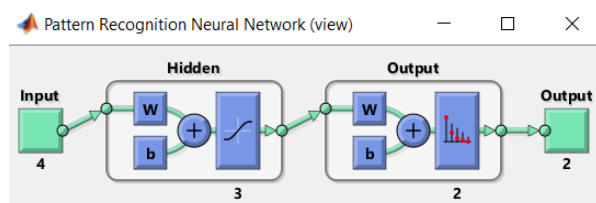


Figure 6.5 : Réseau de neurones artificiel à 1 couche cachée de 3 unités (représentation dans Matlab R2017b)

La couche de sortie recombine de manière optimale les sorties des régressions logistiques opérées par les neurones de la couche cachée. L'algorithme d'apprentissage peut ainsi tracer une frontière de décision arbitrairement complexe dans l'espace des caractéristiques \mathcal{F} . Cette complexité dépend de la quantité x de neurones présents dans la couche cachée, qui augmente en proportion le nombre de

paramètres du modèle. Afin de mesurer l'effet de cette complexité sur la performance du réseau de neurones, celui-ci a été entraîné et validé en faisant évoluer cette quantité x . Si l'on appelle \mathcal{N}^x un réseau possédant x neurones dans sa couche cachée, on a mesuré les performances des modèles \mathcal{N}^1 , \mathcal{N}^3 , \mathcal{N}^5 , \mathcal{N}^7 et \mathcal{N}^9 – où 9 correspond au nombre de sujets disponibles pour l'entraînement de chaque réseau (voir plus bas). L'apprentissage de ces modèles mobilise respectivement 9, 23, 37, 51, et 65 paramètres.

b) Apprentissage sur la base de données

L'utilisation d'une base de validation pendant l'apprentissage est un trait essentiel à l'algorithme des réseaux de neurones. Cette démarche vise à protéger les modèles complexes (ici \mathcal{N}^3 , \mathcal{N}^5 , \mathcal{N}^7 et \mathcal{N}^9) du surapprentissage. Pour ne pas laisser au hasard la sélection des données de validation (cela rendrait la performance du modèle aléatoire), on a utilisé une approche *LOSO* au sein de chaque base d'apprentissage du processus *LOSOTO*. Sur chacun des $N=10$ participants de la base d'apprentissage, on a validé un modèle entraîné sur les 9 autres participants. A l'issue de ces $N=10$ itérations, on a conservé le modèle le plus performant sur la base d'apprentissage constituée des 10 participants. Cette méthode est illustrée dans l'Algorithme 2 disponible en Annexe 11.

Malgré ces précautions, la performance mesurée pour un réseau de neurones possèdera toujours une part de hasard ; car la valeur finale des paramètres dépend de leur initialisation, qui est aléatoire. Pour que les résultats d'un modèle \mathcal{N}^x ne dépendent pas d'un « coup de chance » sur l'initialisation des paramètres, le protocole de validation croisée a été réitéré un certain nombre fois pour chaque réseau. Si M est le nombre de paramètres utilisés pour apprendre un modèle \mathcal{N}^x , alors on a mesuré $2M$ fois sa performance en réitérant le protocole *LOSOTO*. La performance retenue est une moyenne sur l'ensemble des performances ainsi calculées : cette démarche est illustrée dans l'Algorithme 3 proposé en Annexe 11.

III.C.3. Boîtes blanches et boîtes noires

Par construction, notre indice \mathcal{M} utilise une approche *boîte-blanche* pour détecter le stress : il rassemble des « niveaux de certitude » sur l'état de stress exprimé par chaque caractéristique. Idem pour le modèle \mathcal{R}^2 : sa limite de décision $\mathcal{R}^2 = \frac{1}{2}$ correspond à un hyperplan dans l'espace \mathcal{F} des caractéristiques. En ce qui concerne le modèle \mathcal{R}^3 , c'est un peu plus complexe : l'équation $\mathcal{R}^3 = \frac{1}{2}$ correspond à un domaine non-linéaire l'espace \mathcal{F} , qui tend asymptotiquement vers deux hyperplans. Si l'origine des paramètres est connue, la contribution de chaque caractéristique sur la sortie de \mathcal{R}^3 est difficile à mesurer. On peut parler de modèle en « boîte-grise ».

De leur côté, les réseaux de neurones \mathcal{N}^x ne s'appuient pas sur un modèle de la probabilité de stress. Pour les réseaux à 3, 5, 7 unités dans la couche cachée, les paramètres de chaque neurone participent à détecter le stress sans que l'on puisse interpréter leur valeur. Ainsi les réseaux \mathcal{N}^3 , \mathcal{N}^5 , \mathcal{N}^7 et \mathcal{N}^9 sont des modèles de type *boîte-noire*. En revanche, on peut interpréter la valeur des paramètres du réseau \mathcal{N}^1 . A l'issue de l'apprentissage, la frontière de décision dans l'espace \mathcal{F} correspond à un hyperplan dont les $K+1$ coefficients sont connus : leur calcul est donné en Annexe 9 (partie d)). On peut ainsi considérer \mathcal{N}^1 comme un modèle en *boîte-blanche*, au même titre que les modèles \mathcal{M} et \mathcal{R}^2 .

Au travers de cette expérience, la comparaison des modèles \mathcal{M} , \mathcal{R}^x et \mathcal{N}^x nous permettra notamment de déterminer si le recours à un modèle de type *boîte-noire* est nécessaire pour détecter efficacement le stress émotionnel à partir des corrélats identifiés dans le Chapitre Cinq - .

III.D. Expérience 2 : Application de l'Indice de Stress dans un Contexte de Mesures Dégradées

Afin de vérifier la *robustesse* de notre indice de stress, l'expérience précédente peut être reproduite en introduisant du bruit dans les valeurs de chaque caractéristique physiologique.

III.D.1. Fonction de bruit sur une caractéristique physiologique

Imaginons en effet que de mauvaises conditions de mesure (*ex.* : mouvements intempestifs à l'endroit où est positionné le capteur, glissement / décollement d'une électrode sur la peau) altèrent la valeur d'une seule caractéristique f , dont l'évolution ne reflète plus celle de son paramètre biologique cible (*e.g.*, le rythme cardiaque moyen). Cette situation peut être simulée en générant des valeurs aléatoires pour la caractéristique f . Soient μ_f et σ_f la moyenne et l'écart-type de f mesurés sur les classes Stress et Contrôle. On propose d'appliquer à f la « fonction de bruit » suivante :

$$f_t \rightarrow f'_t = \mu_f + \Phi_t, \quad \Phi \sim \mathcal{U}(-3\sigma_f, +3\sigma_f) \quad (6-9)$$

Où Φ suit une distribution uniforme centrée sur Zéro et qui s'étend sur l'intervalle courant des variations de f . On peut assimiler la valeur bruitée f'_t à un tirage aléatoire sur la valeur de f .

On peut alors mesurer l'erreur entre la valeur originale de f et sa valeur bruitée : $\mathcal{E}_t = |f_t - f'_t|$. Pour notre fonction de bruit, la probabilité de générer une erreur contenue dans un certain intervalle est proposée dans le Tableau 6.3 ci-dessous. Ces probabilités d'erreur sont comparables aux probabilités réelles mesurées dans le Chapitre Sept - IV.B, par exemple lorsqu'on observe la qualité générale de la caractéristique *hrRmssd* extraite à partir d'un signal PPG.

Tableau 6.3 : Erreur sur la mesure originale en fonction du bruit introduit pour une caractéristique f quelconque.

Qualité de la mesure	$\mathcal{E} \leq 0,5 * \sigma_f$	$0,5 * \sigma_f < \mathcal{E} \leq \sigma_f$	$\sigma_f < \mathcal{E} \leq 3 * \sigma_f$	$\mathcal{E} > 3 * \sigma_f$
Probabilité d'erreur	16,6%	16,4%	52,5%	14,5%

σ_f : écart-type de f . \mathcal{E} : déviation absolue entre la valeur originale de f et sa valeur bruitée.
Probabilités estimées sur un ensemble de 10^5 échantillons générés aléatoirement entre 0 et 1.

On cherche à observer la propagation de cette erreur jusqu'à la sortie de nos modèles de stress. Dans une logique de test en boîte-blanche, on a appliqué la fonction de bruit de l'équation (6-9) à une seule caractéristique en laissant intacte la valeur des autres (voir justification plus bas).

Naturellement, ce type de bruit n'a pas pour but de reproduire des mesures de terrain. En vie réelle, une altération du signal cardiaque aura des conséquences sur toutes les caractéristiques cardiaques à la fois (ici : *hrMean*, *hrStd* et *hrRmssd*). Il n'est même pas certain que cette altération produira ce bruit centré pour chaque caractéristique. Au travers de cette expérience, on cherche à mener une réflexion sur l'influence de chaque caractéristique dans le processus de décision – et sur la résilience de notre indice de stress, dans le cas où l'une de ses caractéristiques ne reflète plus le paramètre biologique d'intérêt.

Dans notre processus de validation croisée, ces données partiellement bruitées ont été utilisées comme base de *test* pour des modèles appris sur des données non bruitées.

III.D.2. Mesure de performance sur les données bruitées

Comme dans l'expérience précédente, en effet, les modèles de stress comparés ici ne sont pas sensés fonctionner sur une tâche ou un participant connu(e). Afin d'obtenir une mesure valide de leurs performances sur des données bruitées, notre fonction de bruit a été appliquée durant la phase de *test* lors de chaque étape du processus de validation croisée (paragraphe III.B.2). Dans ces conditions, il faut prendre en compte une source d'erreur importante sur la précision mesurée.

La performance des modèles sera largement due au hasard si ceux-ci sont testés sur un petit échantillon de données (avec 1 tâche et 1 sujet, la base de test contiendra seulement 8 mesures

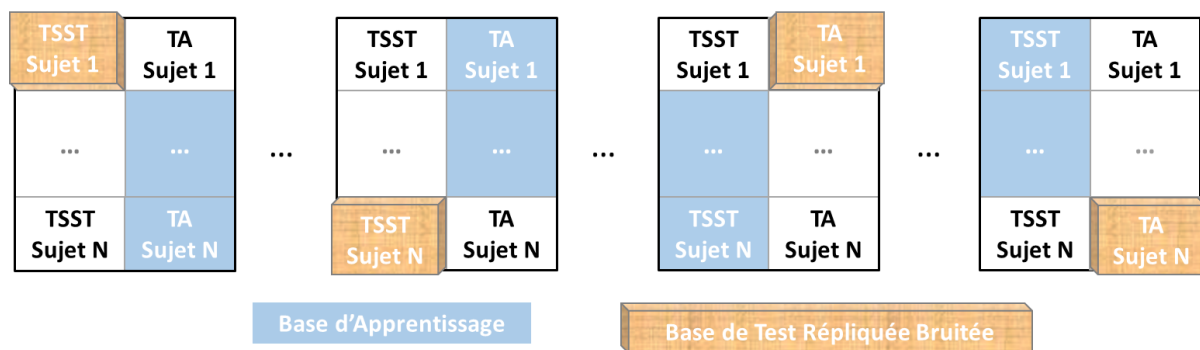


Figure 6.6 : Protocole LOSOTO adapté au bruitage de la base de test

fortement bruitées : voir Tableau 6.2). Pour neutraliser cet effet, on a répliqué 50 fois chaque échantillon de test avant de le soumettre à la fonction de bruit proposée plus haut. Ainsi, la performance de chaque modèle de stress peut être mesurée sur 50 versions bruitées du même échantillon de test. Cette stratégie correspond à une version modifiée de notre protocole de validation croisée, dont une illustration est proposée sur la Figure 6.6.

Dans nos résultats, on conservera la moyenne des performances mesurées sur les 50 réplicats bruités de la base de validation. La démarche générale est synthétisée dans l'Algorithme 4 disponible en Annexe 11. Pour le calcul de l'intervalle de confiance sur la performance mesurée ($IC^{95\%}$: voir équation (6-10)), on conservera la taille initiale de $N=174$ échantillons. En effet, la base de validation provenant de données simulées, sa nouvelle taille ($N' = 50$ réplicats * 174 échantillons) n'a pas de sens d'un point de vue réaliste.

III.D.3. Modification de l'indice de stress et hypothèses à vérifier

Avec ce nouveau protocole de validation croisée, on cherche à vérifier les deux hypothèses formulées dans la section I.B qui ont justifié la construction de notre indice de stress.

- (a) En l'absence d'information sur la qualité des mesures, on peut augmenter la robustesse d'un détecteur de stress en limitant la contribution de chaque caractéristique sur la décision finale.
- (b) En présence d'information pertinente sur la qualité des mesures, on peut augmenter la robustesse du détecteur en retirant des calculs chaque caractéristique bruitée.

Dans cette expérience, le fait de « bruite » une caractéristique en laissant intacte la valeur des autres permet de vérifier ces deux hypothèses à la fois. De cette manière, on peut observer le lien

entre, d'une part, la *contribution* de cette caractéristique sur la sortie des modèles – et d'autre part, la *chute de performance* dans le cas où la caractéristique ne reflète plus le paramètre biologique cible. Cette observation n'est possible que lorsque la contribution de chaque caractéristique est mesurable. Ainsi, on comparera les performances de l'indice de stress à celles du modèle de type *boîte-blanche* qui aura obtenu les meilleurs résultats lors de l'expérience 1.

Dans un second temps, on peut interroger les performances de l'indice de stress lorsqu'on retire la caractéristique bruitée du processus de décision. Contrairement aux autres modèles présentés dans le Tableau 6.4, cette modification de l'espace des caractéristiques peut s'effectuer sans apprentissage : il suffit d'attribuer un coefficient Zéro à la caractéristique bruitée, et un coefficient $1/(K-1)$ aux autres caractéristiques dont la valeur est toujours fiable (voir équation (6-7)). Avec ce nouveau calcul, on fait le pari que la moyenne des indices mono-caractéristiques restants permettra toujours de détecter l'état de stress, à partir du même *seuil* que l'on avait déterminé dans l'expérience 1.

Dans la section IV.B, qui donne les résultats de cette seconde expérience, on appellera « indice entier » (\mathcal{M}) la valeur de \mathcal{M} calculée à partir des K caractéristiques de l'ensemble de départ ; et on appellera « indice partiel » (\mathcal{M}^P) la valeur de \mathcal{M} calculée à partir des K-1 caractéristiques non bruitées.

IV. Résultats et Discussion

Les paragraphes qui suivent donnent les résultats des deux expériences proposées aux sections III.C et III.D. Après avoir observé la sortie de l'indice \mathcal{M} et déterminé un seuil unique pour la détection du stress, on compare ensuite la performance de nos huit modèles (\mathcal{M} , \mathcal{R}^2 , \mathcal{R}^3 , \mathcal{N}^1 , \mathcal{N}^3 , \mathcal{N}^5 , \mathcal{N}^7 et \mathcal{N}^9) à issue du protocole de validation croisée (section IV.A). Après avoir déterminé la contribution de chaque caractéristique sur la sortie des modèles boîte-blanche, on observe enfin la performance de l'indice de stress dans un contexte de mesures dégradées (section IV.B).

IV.A. Expérience 1 : Comparaison des Modèles de Stress

IV.A.1. Détermination d'un seuil sur l'indice de stress

Une fois tous les modèles passés par notre processus de validation croisée, on peut chercher à visualiser les valeurs prédites par l'indice \mathcal{M} à partir des données de test. Les histogrammes (a, b, c) de la Figure 6.6 comparent ainsi la sortie de l'indice \mathcal{M} à celles des régressions logistiques \mathcal{R}^2 (2 classes) et \mathcal{R}^3 (3 classes). Ils affichent la densité de probabilité sur la valeur prédite par ces modèles, pour les classes Stress et Contrôle, dans les deux tâches TSST et TA.

En observant ces trois graphiques, on constate que les modèles \mathcal{R}^2 et \mathcal{R}^3 polarisent fortement leur sortie autour des valeurs 0 et 1. Ce comportement est absent pour l'indice \mathcal{M} , dont les sorties de la classe Contrôle s'étendent surtout sur l'intervalle [0,12 ; 0,40], et celles de la classe Stress sur l'intervalle [0,28 ; 0,64]. A partir des histogrammes du graphique (a), on peut proposer une première valeur seuil pour distinguer les classes Stress et Contrôle à partir de l'indice de stress : $\mathcal{M} = 0,4$.

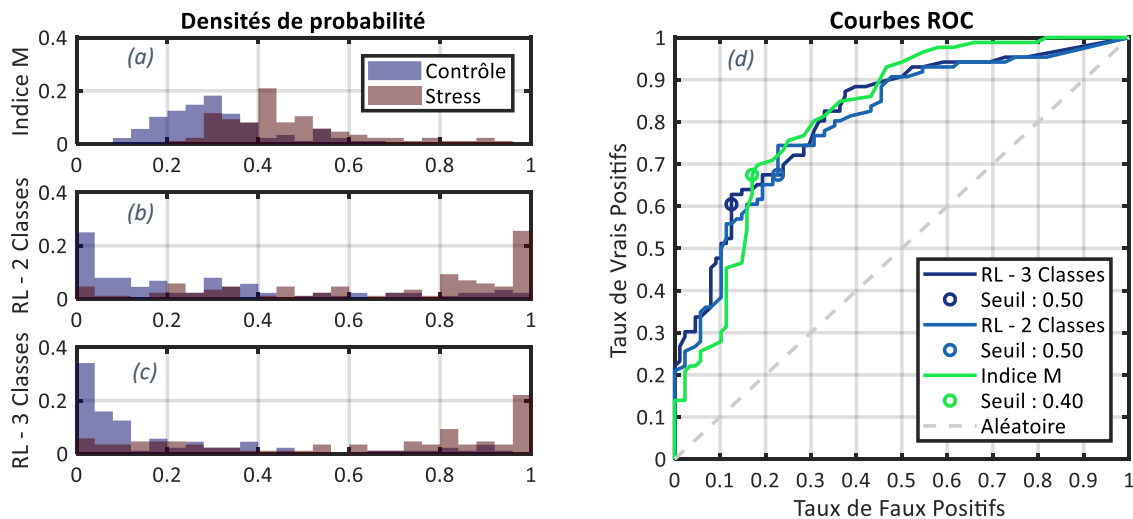


Figure 6.7 : Distribution des classes Stress et Contrôle en sortie de l'indice \mathcal{M} et d'une régression logistique à deux ou trois classes, sur l'ensemble des données de test

A gauche : Densité de probabilité des classes Stress et Contrôle pour \mathcal{M} (a), \mathcal{R}^2 (b) et \mathcal{R}^3 (c).

A droite (d) : courbe ROC tracée en déplaçant un seuil sur les valeurs de ces trois modèles (\mathcal{M} : vert clair ; \mathcal{R}^2 : bleu ; \mathcal{R}^3 : bleu foncé).

En déplaçant cette valeur seuil sur l'abscisse de l'histogramme (a), on peut détecter la classe Stress avec un taux de *vrais positifs* et un taux de *faux positifs*⁷ parcourant chacun l'intervalle [0 ; 1]. Le tracé du second taux en fonction du premier donne une courbe ROC (de l'anglais : *Receiver Operating Characteristic*) qui est affichée dans le graphique (d) de la Figure 6.6, en comparaison avec les courbes ROC des modèles \mathcal{R}^2 et \mathcal{R}^3 .

Le seuil de détection « par défaut » des régressions logistiques est de 0,5. Sur le graphique (d), les points correspondant à ce seuil sont marqués d'un cercle dans la couleur des deux indices. A ce niveau, l'indice \mathcal{R}^2 obtient un taux de faux positifs de 22,7% pour un taux de vrais positifs de 67,4% ; pour l'indice \mathcal{R}^3 , ces taux sont respectivement de 12,5% et de 60,4%. Au seuil de **0,4** proposé plus haut, l'indice \mathcal{M} obtient des taux similaires avec 17,5% de faux positifs pour 67,4% de vrais positifs. On peut donc retenir cette valeur pour l'utilisation de l'indice \mathcal{M} comme détecteur de stress.

IV.A.2. Performance générale des algorithmes

On peut maintenant comparer la performance des huit détecteurs de stress élaborés sur notre base de données. Pour mémoire, le critère de performance retenu est le pourcentage d'échantillons bien classés en validation croisée. Pour chacun des modèles, ce pourcentage est affiché sur la dernière ligne du Tableau 6.4. Les marges d'erreur correspondent à l'intervalle de confiance à 95% calculé à partir du nombre d'exemples disponibles (équation (6-10)).

En ce qui concerne l'indice \mathcal{M} , on observe des précisions supérieures aux modèles \mathcal{R}^2 et \mathcal{R}^3 (resp. +2,8% et +1,2%). Notre méthode d'apprentissage est donc (au moins) aussi efficace qu'une régression logistique standard ; et cette efficacité n'est pas seulement liée à l'inclusion de la classe Repos dans le processus d'apprentissage. En revanche, les réseaux de neurones affichent des précisions moyennes 6% à 11% supérieures à celle de l'indice \mathcal{M} .

⁷ Ces deux notions sont développées en Annexe 9 – partie e)

Tableau 6.4 : Détail des modèles et des précisions mesurées par validation croisée

Algorithme Modèle	Indice de Stress \mathcal{M}	Régressions Logistiques		Réseaux de Neurones				
		\mathcal{R}^2	\mathcal{R}^3	\mathcal{N}^1	\mathcal{N}^3	\mathcal{N}^5	\mathcal{N}^7	\mathcal{N}^9
Nb Paramètres	8	5	10	9	23	37	51	65
Validation LOSO	Non	Non	Non	Oui	Oui	Oui	Oui	Oui
Precision [%]	75,2 ± 6,4	72,4 ± 6,6	74,0 ± 6,5	80,9 ± 5,8	82,9 ± 5,6	84,6 ± 5,4	85,6 ± 5,2	85,7 ± 5,2

$\mathcal{R}^2, \mathcal{R}^3$: Régressions logistiques à 2 et 3 classes.

$\mathcal{N}^1, \mathcal{N}^3, \mathcal{N}^5, \mathcal{N}^7$ et \mathcal{N}^9 : Réseaux de neurones à 1, 3, 5, 7 et 9 unités dans la couche cachée.

Sur la dernière ligne : précision mesurée par validation croisée LOSOTO ± l'intervalle de confiance à 95%.

Cette différence n'est pas surprenante : l'apprentissage des réseaux $\mathcal{N}^3, \mathcal{N}^5, \mathcal{N}^7$ et \mathcal{N}^9 est moins contraint que celui de l'indice \mathcal{M} car il fait appel à plus de paramètres (ligne 2 du Tableau 6.4). En outre, l'augmentation du nombre de neurones montre un effet tendanciel positif sur la performance des modèles \mathcal{N}^x . Le taux d'erreur moyen passe progressivement de 19,1% à 14,3% ; et semble atteindre un optimum dès 7 unités dans la couche cachée. Cette augmentation plafonnée des performances de l'algorithme des réseaux de neurones est la marque d'une protection efficace contre le surapprentissage. En effet, tous les modèles \mathcal{N}^x ont bénéficié d'une validation croisée de type LOSO pendant l'entraînement (ligne 3 du Tableau 6.4).

Sur la base de données utilisée dans ce chapitre, le taux d'erreur minimal obtenu avec un modèle arbitrairement complexe semble ainsi s'établir autour de 14%. De ce point de vue, les réseaux de neurones \mathcal{N}^7 et \mathcal{N}^9 ont dessiné chacun une frontière « optimale » entre les classes Stress et Non-Stress lors des différentes phases d'apprentissage. Au vu du nombre d'exemples disponibles, toutefois, les performances de ces modèles boîte-noire ne sont pas significativement différentes de la performance mesurée pour le modèle \mathcal{N}^1 , qui inclut ces dernières dans son intervalle de confiance. L'augmentation des performances en fonction du nombre de neurones doit être relativisée avec le gain réel obtenu en terme de bien classés. Avec 2,6 fois plus de paramètres que le modèle \mathcal{N}^1 , par exemple, le modèle \mathcal{N}^3 a ajouté en moyenne 3,5 exemples à sa liste de bonnes réponses (sur les 174 disponibles dans la base de données).

IV.A.3. Contribution d'une caractéristique à la détection du stress

En insérant un hyperplan entre les classes Stress et Contrôle dans l'espace des caractéristiques, le modèle \mathcal{N}^1 a ainsi obtenu des performances proches des autres modèles \mathcal{N}^x – qui peuvent théoriquement déterminer une frontière bien plus complexe. Ce résultat montre qu'un séparateur linéaire peut approcher efficacement la frontière « idéale » dans l'espace des corrélats du stress.

Lors de chaque phase de la validation croisée, on a conservé les coefficients de cette frontière, tels qu'ils ont été appris pour les modèles \mathcal{N}^1 et \mathcal{R}^2 . Lorsque l'intercept (l'ordonnée à l'origine) est ramené à -1, ces coefficients présentent des valeurs normalisées qui sont affichées dans le Tableau 6.5.

Les modèles \mathcal{R}^2 et \mathcal{N}^1 ont attribué en moyenne des coefficients forts et équivalents entre les caractéristiques *respAutocorr*, *hrRmssrd* et *hrMean*, tandis que *hrStd* (i.e. : l'écart-type du rythme cardiaque) hérite d'un coefficient faible.

Tableau 6.5: Coefficients de la frontière de décision calculés pour les modèles \mathcal{R}^2 et \mathcal{N}^1 durant la validation croisée

	<i>Intercept</i>	<i>hrStd</i>	<i>respAutocorr</i>	<i>hrRmssd</i>	<i>hrMean</i>
\mathcal{R}^2	-1,0	-0,11 ± 0,30	0.49 ± 0,22	0.46 ± 0.52	0.45 ± 0.14
\mathcal{N}^1	-1,0	-0,08 ± 0,08	0.53 ± 0.10	0.45 ± 0.10	0.45 ± 0.08

Moyenne sur toutes les phases d'apprentissage ± intervalle de confiance à 95% calculé à partir de l'écart-type.

hrMean : rythme cardiaque moyen ; *hrStd* : écart-type du rythme cardiaque ; *hrRmssd* : RMS des différences successives du rythme cardiaque ; *respAutocorr* : autocorrélation du signal respiratoire.

Malgré la différence de performance entre les deux modèles, ceux-ci s'accordent ainsi *en moyenne* sur une frontière de décision commune dans l'espace des quatre caractéristiques. On peut proposer une forme simplifiée de cette frontière dans l'équation suivante.

$$0,5 * (respAutocorr_{\tau} + hrRmssrd_{\tau} + hrMean_{\tau}) - 0,1 * hrStd - 1 = 0 \quad (6-10)$$

Sur l'ensemble de la base de données, ce modèle linéaire distingue les classes Stress et Contrôle avec 82,2% de bien classés à la valeur seuil Zéro, soit une performance proche de celle mesurée pour le modèle \mathcal{N}^1 en validation croisée. Dans le Tableau 6.5, la concentration des coefficients de \mathcal{N}^1 autour de leurs moyennes respectives montre que ce modèle propose fréquemment le séparateur linéaire proposé dans l'équation (6-10). Dans ce modèle, une variation positive de *hrStd* aura cinq fois moins d'impact sur la sortie de ces modèles qu'une variation positive des autres caractéristiques. Ce chiffre préfigure une faible contribution de *hrStd* sur la sortie du modèle \mathcal{N}^1 .

Si l'on fait confiance à l'intervalle de confiance à 95% proposé ici (±1,96 fois l'écart-type calculé sur les coefficients recueillis à chaque apprentissage), le coefficient de cette caractéristique n'est pas significativement différent de Zéro pour les deux modèles \mathcal{R}^2 et \mathcal{N}^1 . Cette neutralisation de *hrStd* par les algorithmes d'apprentissage suggère que cette caractéristique apporte peu d'information supplémentaire sur l'état de stress, par rapport aux trois autres. Les modèles ont ainsi opéré leur propre sélection des caractéristiques dans l'ensemble initial.

De son côté, l'indice \mathcal{M} borne la contribution de toutes ses caractéristiques dans l'intervalle [0 ; 1/K], où K est le nombre de caractéristiques utilisées (ici K=4). A l'intérieur de cet intervalle, cette contribution dépend des seuils f_{Ref} et f_{Cer} qui ont été appris indépendamment pour chaque caractéristique (voir Annexe 10 – partie c)). Le Tableau 6.6 ci-dessous affiche les valeurs de ces seuils durant le processus de validation croisée.

Tableau 6.6: Seuil de référence et seuil de certitude pour chaque caractéristique durant la validation croisée

	<i>hrStd</i>	<i>respAutocorr</i>	<i>hrRmssd</i>	<i>hrMean</i>
f_{Ref}^v	-1.0 ± 0.4	-1.4 ± 0.3	-1.8 ± 1.1	-0.6 ± 1.8
f_{Cer}^v	5.8 ± 1.9	3.2 ± 1.0	5.7 ± 5.0	3.7 ± 1.6

Seuils calculés sur les valeurs calibrées des caractéristiques.

Moyenne sur l'ensemble des phases d'apprentissage ± intervalle de confiance à 95% calculé à partir de l'écart-type.

hrMean : rythme cardiaque moyen ; *hrStd* : écart-type du rythme cardiaque ; *hrRmssd* : RMS des différences successives du rythme cardiaque ; *respAutocorr* : autocorrélation du signal respiratoire.

Malgré une forte variance autour des valeurs f_{Ref} et f_{Cer} pour la caractéristique *hrRmssd*, les intervalles d'incertitude sont peu différents *en moyenne* entre les caractéristiques. Ainsi, on peut supposer que les indices simples \mathcal{S} appris pour chaque caractéristique sont relativement semblables ; et engendrent des contributions relativement équivalentes entre les quatre caractéristiques sur la

sortie de l'indice de stress. Contrairement aux modèles standard du Tableau 6.5, l'indice \mathcal{M} ne minimise pas l'influence de la caractéristique $hrStd$.

IV.B. Expérience 2 : Application dans un Contexte de Mesures Dégradées

IV.B.1. Performances mesurées sur des données bruitées

Au-delà du problème d'apprentissage, on a proposé d'observer la robustesse des modèles proposés lorsque les conditions de mesure se dégradent. Pour représenter cette situation, on a choisi de générer des valeurs aléatoires pour chaque caractéristique tour à tour, tandis que la valeur des autres demeurerait intacte. A chaque fois, on peut comparer la précision mesurée par validation croisée à celle observée sans ajout de bruit (mesures « idéales » : voir Tableau 6.4). Toutes ces données sont représentées dans la Figure 6.8, qui affiche en abscisse le nom de la caractéristique bruitée dans chaque situation. Les résultats sont ici présentés pour trois modèles de type *boîte-blanche*.

- En noir, le réseau de neurones \mathcal{N}^1 qui insère un hyperplan séparateur (voir équation (6-10)) dans l'espace des 4 caractéristiques de départ.
- En vert clair, l'indice de stress « entier » \mathcal{M} calculé à partir des 4 caractéristiques de l'ensemble de départ.
- En vert foncé, l'indice de stress « partiel » \mathcal{M}^P calculé sans nouvel apprentissage, à partir des seules caractéristiques non bruitées. Ce modèle correspond à l'indice qui serait calculé en présence d'informations pertinentes sur la qualité des données.

La hauteur des histogrammes représente la performance moyenne mesurée sur les 50 réplicats bruités de la base de test. Comme dans le Tableau 6.4, les barres d'erreur représentent l'intervalle de confiance à 95% calculé dans l'équation (6-8) et qui tient compte du nombre d'exemples initialement disponibles dans la base de données. Selon cet indicateur, un écart entre deux performances est « significatif » si les deux intervalles de confiance associés sont disjoints.

De manière générale, les modèles montrent des performances moindres lorsque du bruit est introduit dans les mesures, par rapport aux mesures « idéales » (1^{ère} position en abscisse). Lorsque toutes les caractéristiques présentent des valeurs aléatoires (6^e position en abscisse), le modèle \mathcal{M}^P n'est pas défini ; tandis que les modèles \mathcal{N}^1 et \mathcal{M} affichent la précision du hasard (environ 50%). Ce résultat montre que notre mesure de la performance est correcte ; et que les deux modèles \mathcal{N}^1 et \mathcal{M} ne sont pas biaisés dans leur estimation de l'état de stress. Il ne sera pas pris en compte dans la suite des nos analyses.

IV.B.2. Analyse des résultats au regard des hypothèses qui ont motivé la construction de l'indice de stress

Au travers de la Figure 6.8, on a cherché à vérifier les deux hypothèses suivantes.

- (a) En l'absence d'information sur la qualité des mesures, on peut augmenter la robustesse d'un détecteur de stress en limitant la contribution de chaque caractéristique sur la décision finale. Cette fonction est assurée par l'indice « entier » \mathcal{M} .

- (b) En présence d'information pertinente sur la qualité des mesures, on peut encore augmenter la robustesse du détecteur en retirant des calculs chaque caractéristique bruitée. Cette fonction est assurée par l'indice « partiel » \mathcal{M}^P .

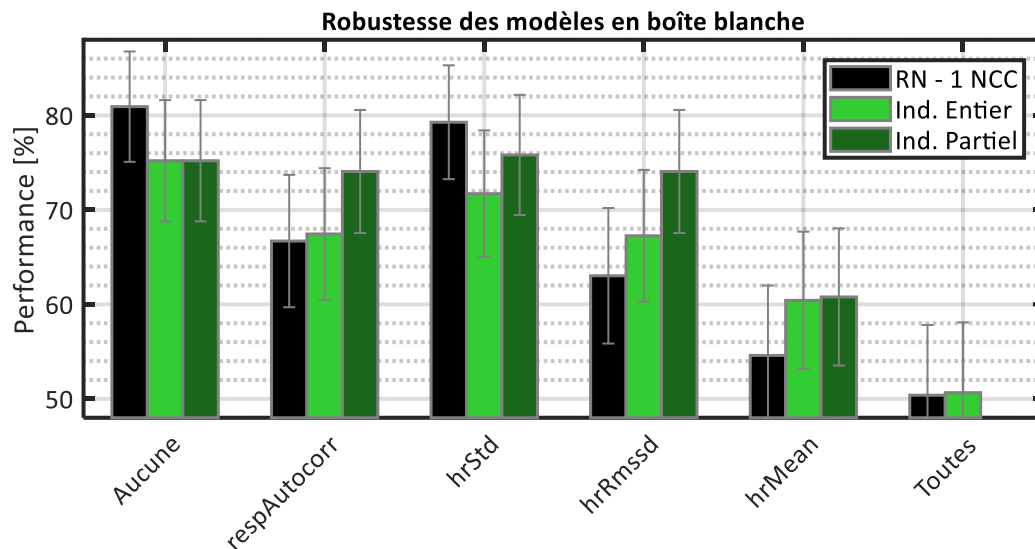


Figure 6.8 : Performance mesurée pour trois modèles de stress lorsqu'une caractéristique présente des valeurs aléatoires

En abscisse : Nom de la caractéristique bruitée.

En ordonnée : Performances mesurées par validation croisée.

Les barres d'erreurs représentent l'intervalle de confiance à 95% tenant compte du nombre d'échantillons disponibles.

En noir : le réseaux de neurones \mathcal{N}^1 . En vert clair : l'indice \mathcal{M} « entier » calculé sur les 4 caractéristiques disponibles. En vert foncé : l'indice \mathcal{M}^P « partiel » calculé seulement sur les caractéristiques non bruitées.

hrMean : rythme cardiaque moyen ; hrStd : écart-type du rythme cardiaque ; hrRmssd : RMS des différences successives du rythme cardiaque ; respAutocorr : autocorrélation du signal respiratoire.

a) Modèle témoin : le réseau de neurones \mathcal{N}^1

Le modèle standard \mathcal{N}^1 montre une forte disparité des performances en fonction de la caractéristique bruitée dans les mesures. Par rapport aux mesures idéales (première colonne), la chute de performance est significative pour toutes les caractéristiques, sauf pour l'écart-type du rythme cardiaque (*hrStd*). La chute de performances s'élève à 14,2% dans le cas de la caractéristique *respAutocorr* ; à 16,9% dans le cas de *hrRmssd* ; et à 26,3% dans le cas du rythme moyen *hrMean*. Dans ce dernier cas, la performance du modèle est proche du hasard (54,6%). Ainsi, ce modèle *performant* sur des données idéales se montre relativement peu *robuste* sur des données bruitées.

Une exception notable est à relever dans le cas de la caractéristique *hrStd*, pour laquelle le modèle \mathcal{N}^1 montre une performance équivalente aux mesures idéales (79,2% contre 80,9%) au regard des barres d'erreur. Ce résultat est cohérent avec ceux de la section précédente sur les coefficients de cette caractéristique (Tableau 6.5) : ils appuient l'hypothèse que l'écart-type du rythme cardiaque (*hrStd*) a *peu* contribué à la sortie modèle \mathcal{N}^1 au cours des différentes phases d'apprentissage. Ainsi, le bruitage de cette caractéristique n'a pas biaisé la décision du modèle \mathcal{N}^1 lors de la détection du stress sur les bases de test. L'inconvénient de ce phénomène est qu'il a mécaniquement augmenté le poids des trois autres caractéristiques dans ce processus de décision, ce qui a probablement diminué la performance du modèle \mathcal{N}^1 lorsque ces caractéristiques étaient bruitées.

b) Impact de la saturation des caractéristiques : performances de l'indice entier \mathcal{M}

Cette disparité des performances observée pour le modèle \mathcal{N}^1 est atténuée dans le cas de l'indice de stress « entier » \mathcal{M} . Bien que ce modèle montre des performances inférieures au modèle \mathcal{N}^1 dans le cas des mesures idéales, il montre des performances supérieures au modèle \mathcal{N}^1 lorsque les caractéristiques $hrRmssd$ et $hrMean$ sont bruitées (resp. 67,5% contre 63,7% pour $hrRmssd$ et 60,8% contre 54,6% pour $hrMean$). La performance est équivalente au modèle \mathcal{N}^1 dans le cas de la caractéristique $respAutocorr$ (67.2% contre 66.5%), mais la chute de performance observée est moins importante pour le modèle \mathcal{M} (7,7%). Ainsi, le modèle \mathcal{M} est plus robuste que le modèle \mathcal{N}^1 lorsque les caractéristiques $respAutocorr$, $hrRmssd$ et $hrMean$ ne reflètent plus le paramètre biologique cible.

Dans le Tableau 6.6, on a pu constater que l'indice \mathcal{M} avait tendance à attribuer des contributions équivalentes aux quatre caractéristiques. Par construction, l'indice \mathcal{M} sature également la valeur de chaque caractéristique au moyen des indices simples \mathcal{S} . Cette stratégie semble payante pour limiter l'impact de données bruitées sur la décision finale du modèle, lorsqu'on compare ses résultats à ceux du modèle \mathcal{N}^1 . Bien que la chute de performance soit ainsi limitée, celle-ci s'établit cependant à moins de 70% dans trois cas de bruit sur quatre.

c) Impact du retrait des caractéristiques bruitées : performances de l'indice partiel \mathcal{M}^p

Le retrait des caractéristiques bruitées change nettement la donne. Pour trois cas de bruit sur quatre, l'indice \mathcal{M}^p affiche une performance équivalente aux mesures idéales. Ces résultats suffisent pour dire que l'indice partiel \mathcal{M}^p est robuste lorsque les caractéristiques $respAutocorr$, $hrStd$ et $hrRmssd$ présentent des valeurs aléatoires. Avec un pourcentage de bien classés proche de 75%, l'indice partiel \mathcal{M}^p se montre relativement performant dans ces trois cas de bruit. Les performances de l'indice \mathcal{M}^p sont toujours supérieures ou équivalentes à celles de l'indice \mathcal{M} . Hormis le cas de la caractéristique $hrStd$, dans lequel les performances sont relativement proches (-3,4%), l'indice \mathcal{M}^p affiche toujours des performances supérieures au modèle \mathcal{N}^1 (+7,3% pour $respAutocorr$, +11,0% pour $hrRmssd$, + 6,2% pour $hrMean$).

Ainsi, la stratégie qui consiste à retirer *sans apprentissage* une caractéristique du calcul de l'indice stress semble avoir porté ses fruits dans la majorité des cas de bruit. Il faut cependant souligner certaines limites dans les résultats affichés sur la Figure 6.8. D'une part, les marges d'erreurs calculées sur la base du nombre d'exemples disponibles ne nous permettent pas d'affirmer que la différence trouvée entre l'indice de stress (\mathcal{M} , \mathcal{M}^p) et le modèle standard (\mathcal{N}^1) est « statistiquement significative ».

D'autre part, aucun des trois modèles ne semble résister au bruitage de la caractéristique $hrMean$. On peut déduire de ce résultat que ces modèles reposent fortement sur l'information apportée par le rythme cardiaque moyen en situation de stress. Contrairement au réseau de neurones, l'indice de stress parvient à conserver des performances significativement différentes du hasard (la valeur 50% n'est pas contenue dans l'intervalle de confiance des modèles \mathcal{M} et \mathcal{M}^p). Mais la performance revendiquée par les indices \mathcal{M} et \mathcal{M}^p (60%) ne permet pas d'envisager une application du modèle dans le contexte simulé par nos données bruitées.

Ces derniers points nous amènent à discuter des limites de la méthode proposée dans ce chapitre.

IV.C. Limites de cette Etude

IV.C.1. Sur les données disponibles

L'utilisation d'une caractéristique respiratoire a nécessité de restreindre fortement le nombre de sujets disponibles dans la base de données (celui-ci est passé de 20 à 11). En outre, on n'utilise que deux tâches parmi les quatre disponibles dans la base de données : TSST et TA, qui ont été sélectionnées dans le Chapitre Cinq - à partir des niveaux subjectifs de stress. Des travaux menés en parallèle sur cette base de données ont montré que des modèles appris sur TSST et TA se montraient moins efficaces sur la tâche SECPT (Sakri et al., 2018) – et les résultats n'étaient pas meilleurs sur la tâche D2. Si le caractère *stressant* de ces deux tâches peut être mis en cause, il est possible également que celles-ci induisent un « type de stress » trop différent de TSST et TA.

En toute rigueur, les résultats présentés dans la section précédente ne sont donc valides que pour ces mêmes personnes (européennes, francophones et majoritairement jeunes) et ces mêmes tâches (impliquant le jugement social et la parole). Cela dit, la capacité de généralisation de nos modèles sur des populations et des activités similaires est garantie par notre protocole de validation croisée. La base de données utilisée dans cette partie reste multitâche et dotée de conditions Contrôle multiples : en ce sens, elle est plus riche que la majorité des bases utilisées dans la littérature.

Cela étant dit, les données disponibles dans cette partie étaient réduites au quart de la taille initiale de la base de données. Durant le processus de validation LOSOTO, chaque base d'apprentissage remplissait un huitième de cette taille initiale (environ 40 exemples par classe et par tâche). Cette base de données réduite a constitué un handicap durant l'apprentissage des modèles, qu'il convient de prendre en compte en regardant les résultats du Tableau 6.4. Enfin, la faible quantité d'exemples disponibles s'est aussi ressentie lors de l'évaluation des performances. Au regard de l'intervalle de confiance défini dans l'équation (6-8), la taille d'échantillon associée à la base de test était parfois insuffisante pour donner un sens statistique aux différences de performances observées entre les modèles de stress.

IV.C.2. Sur le jeu de caractéristiques utilisé

De même que les tâches disponibles, l'ensemble de caractéristiques sélectionné à l'issue du Chapitre Cinq - a été conservé tel quel et livré sans filtre aux algorithmes d'apprentissage utilisés dans cette étude. Ces caractéristiques ont en commun de proposer une réponse *univoque* au stress : elles varient toujours dans le même sens en cas d'activité, mais de manière plus intense en cas de stress émotionnel. Bien qu'elle soit nécessaire pour concevoir l'indice de stress, cette propriété n'est pas nécessaire pour les modèles de type *boîte-noire* (comme \mathcal{N}^3 , \mathcal{N}^5 , \mathcal{N}^7 et \mathcal{N}^9) qui peuvent tracer des frontières arbitrairement complexes dans l'espace des caractéristiques. Ils peuvent ainsi tirer profit de caractéristiques sans lien direct avec le niveau de stress.

Ainsi, le jeu de caractéristiques utilisé dans cette partie ne correspond pas forcément au jeu *optimal* que l'on aurait obtenu, par exemple, au moyen d'une sélection de type *wrapper* pour ces modèles. En particulier, la neutralisation de la caractéristique *hrStd* (i.e, l'écart-type du rythme cardiaque) par les algorithmes standard \mathcal{R}^2 et \mathcal{N}^1 montre que cette caractéristique n'apporte pas plus d'information que les trois autres réunies pour la détection du stress. Ce résultat s'oppose ainsi à ceux

du Chapitre Cinq - III.B, où le Tableau 5.6 affichait l'écart-type du rythme cardiaque en première position des caractéristiques sélectionnées sur la « part du stress » mesurée dans leurs variations.

IV.C.3. Sur l'introduction de bruit dans la valeur des caractéristiques

L'utilisation des quatre caractéristiques sélectionnées au Chapitre Cinq - III.B était cependant nécessaire pour mettre en évidence le potentiel de l'indice de stress dans une application en vie réelle. Afin de comparer l'impact d'une caractéristique bruitée sur la sortie de plusieurs modèles de stress, ceux-ci devaient également être construits sur le même ensemble de caractéristiques. Dans cette démarche, le « test de robustesse » mis en place dans la section III.D pour mesurer les performances de nos modèles sur des mesures dégradées, nous a permis de justifier les hypothèses qui ont motivé la construction de l'indice de stress.

Cela dit, l'introduction de bruit dans des données de laboratoire ne suffit pas à reproduire des enregistrements de la vie quotidienne. Le fait que l'indice de stress se soit montré *robuste* sur nos données simulées ne garantit pas que celui-ci sera *robuste* en vie réelle. Comme on le soulignait dans la section III.D, un défaut du capteur de rythme cardiaque peut entraîner la défaillance de plusieurs caractéristiques à la fois. Ainsi, une extension pertinente du travail proposé dans la section III.D serait de mesurer la réaction des modèles de stress à une combinaison aléatoire de caractéristiques bruitées.

Par ailleurs, on a constaté dans la section IV.B que la performance des modèles \mathcal{M} et \mathcal{N}^1 reposait essentiellement sur l'information portée par le rythme cardiaque moyen (*hrMean*). En vie réelle, cette caractéristique est pourtant largement dépendante de l'activité physique. Difficilement prévisibles à partir des contributions des caractéristiques mesurées au paragraphe IV.A.3, ces chutes de performance montrent ainsi l'intérêt de multiplier les sources d'information sur l'état de stress dans la construction d'un détecteur en boîte-blanche. Elles posent également la question du poids de cette caractéristique dans les modèles de stress proposés dans l'état de l'art.

Toutefois, nous n'avons pas trouvé dans la littérature de méthode équivalente à celle qui nous a permis de mesurer les performances affichées sur la Figure 6.8. Ce manque de référence disponible nous empêche de discuter plus profondément les résultats de la section IV.B.

IV.D. Discussion

IV.D.1. Comparaison à la littérature des performances mesurées sur la base de données initiale

En ce qui concerne la performance des modèles sur des données non bruitées, il est également délicat de comparer nos résultats à ceux de l'état de l'art ; car notre protocole de validation croisée n'a pas non plus d'équivalent dans les études existantes en détection du stress.

Si l'on s'intéresse aux algorithmes d'apprentissage, on peut observer que la performance affichée dans le Tableau 6.4 pour le modèle \mathcal{R}^2 est comparable à celle de 74,5% mesurée par Wijsman et collaborateurs (2013) avec le même algorithme, en utilisant une condition Contrôle dans un protocole expérimental multitâche. Cela dit, ces derniers auteurs avaient utilisé des modèles mono-utilisateurs.

On peut alors s'intéresser aux études de l'Annexe 4 qui ont impliqué une condition contrôle, plusieurs tâches et une validation croisée de type *LOSO*. Deux d'entre elles ont reporté des précisions

inférieures à nos huit modèles : 67,7% et 68% (Mohino-Herranz et al., 2015; Shi et al., 2010). En fournissant un ensemble de 63 caractéristiques à un algorithme de forêts aléatoires, Gjoreski et collaborateurs (2016) ont obtenu des performances analogues à nos modèles \mathcal{N}^x (83%). La performance de l'indice de stress \mathcal{M} (75%) se situe entre ces deux niveaux. Elle correspond à la moyenne des performances observées dans l'état de l'art, parmi les études ayant contrôlé l'effet d'activité, sans nécessairement contrôler l'effet de la tâche ou l'effet du sujet (voir Tableau 3.6, p.55).

Avec un ensemble de caractéristiques restreint, nos modèles montrent ainsi des performances comparables à celles habituellement mesurées dans la littérature avec un protocole de test moins rigoureux. Ces performances valident ainsi la méthode proposée tout au long de ce chapitre pour construire un modèle de détection du stress, alors même que cette méthode ne cherche pas à maximiser la performance mesurée.

IV.D.2. Sur la pertinence des modèles de type boîte-noire

Notre comparaison de modèles nous a d'ailleurs conduit à opposer des détecteurs de type *boîte-blanche* à des détecteurs de type *boîte-noire*, dont on a pu faire évoluer le nombre de paramètres.

Dans le Tableau 6.4, les modèles \mathcal{N}^x ont affiché les meilleurs scores de précision. En sélectionnant à chaque étape, la version de chaque modèle \mathcal{N}^x qui obtenait les meilleures performances sur toute la base d'apprentissage, on pouvait s'attendre à voir la performance de ces modèles augmenter avec leur degré de complexité. Outre une protection efficace contre le surapprentissage, l'application d'une validation *LOSO* pendant l'entraînement des modèles \mathcal{N}^x a probablement augmenté la résilience de ces modèles à l'effet du sujet, partiellement contrôlé par le processus de validation *LOSOTO*. Malgré cette stratégie avantageuse pour les modèles *boîte-noire*, cependant, multiplier par six le nombre de paramètres n'a diminué que marginalement le taux d'erreur des modèles \mathcal{N}^7 et \mathcal{N}^9 par rapport au modèle \mathcal{N}^1 (de 19% à 14%).

Dans un contexte où le coût d'une erreur est faible (on cherche à repérer des épisodes de stress sur une journée d'enregistrement) ; et où la complexité d'un modèle peut diminuer ses performances dans un contexte différent des données d'apprentissage, il n'est pas aberrant de retenir le modèle qui propose la frontière de décision la plus simple. Sur nos données expérimentales, par exemple, le séparateur linéaire proposé dans l'équation (6-10) affiche une performance comparable à de l'état de l'art, parmi les protocoles qui impliquent une condition Contrôle (voir Annexe 4). Ces résultats viennent appuyer d'autres études existantes, comme celle de Setz et collaborateurs (2010), qui avaient obtenu de meilleures performances avec une analyse discriminante qu'avec les séparateurs à vastes marges.

Tous ces éléments appuient l'idée que l'utilisation d'un modèle de type boîte-noire n'est pas nécessaire pour détecter l'état de stress. Nos résultats justifient ainsi *a posteriori* la démarche en boîte-blanche dans laquelle s'inscrit l'indice de stress.

IV.D.3. Pistes d'amélioration pour l'indice de stress

Malgré tout, l'implémentation proposée dans la section II est encore largement perfectible. En conditions de mesures standard, par exemple, l'indice \mathcal{M} a montré des performances bien inférieures au modèle \mathcal{N}^1 . Ce résultat suggère d'inclure des données de validation dans le processus d'apprentissage – un élément de méthode absent des régressions logistiques actuellement utilisés

pour dimensionner l'indice simple. Partant de ce constat, il pourrait être intéressant de dimensionner les indices simples avec des détecteurs de type \mathcal{N}^1 .

Sur la Figure 6.7, l'indice de stress n'atteint que rarement des valeurs supérieures à 0,6 ou inférieures à 0,1. Les valeurs Stress et Contrôle sont relativement concentrées autour du seuil de décision fixé par validation croisée. Ainsi, une petite modification du seuil optimal peut engendrer une chute drastique de la performance. D'après les résultats de la Figure 6.8, cela dit, le retrait de trois caractéristiques différentes n'a pas semblé modifier le seuil optimal de décision. En revanche, ce seuil pourrait être à l'origine de la chute drastique de performance observée lors du retrait du rythme cardiaque moyen. Cette chute pourrait justifier un traitement différencié du rythme cardiaque moyen dans le calcul de l'indice multiple.

Plusieurs théories pourraient permettre de combiner les indices simples de sorte à obtenir un indice multiple plus performant. Si l'on considère, par exemple, chaque indice simple \mathcal{S} comme un détecteur de stress à part entière, on peut utiliser les méthodes de l'apprentissage ensembliste. Une ouverture intéressante se trouve, par exemple, dans l'apprentissage dit *structuré* (Sutton et al., 2006), ou *partitionné* (Chang et al., 2008), qui cherche à recombinaison des probabilités « locales » de stress issues de certaines partitions de caractéristiques. Mais dans cette démarche, il faudrait veiller à ce que l'indice final ne nécessite pas un nouvel apprentissage pour chaque cas où une caractéristique est retirée ou rajoutée au modèle final.

Une autre perspective intéressante se trouve dans les méthodes de la logique floue (Zadeh, 1988). Dans les limites de notre problème à deux classes, la conception d'un modèle flou peut être décrite en deux étapes : (i) Pour chaque caractéristique $f \in \mathcal{F}$, on postule la *fonction d'appartenance* de f à l'état de stress. L'indice \mathcal{S} développé dans cette partie est largement analogue à une fonction d'appartenance. (ii) Etant donné un ensemble de mesures, on utilise des *règles d'inférence* pour déterminer l'état de stress à partir des fonctions d'appartenance. Par défaut, ces règles ne nécessitent pas d'apprentissage. Un modèle de stress basé sur la logique floue a été implémenté avec succès par Salazar-Ramirez et collaborateurs (2018), qui n'ont pas proposé de score de performance.

La logique floue illustre efficacement le changement de paradigme proposé par notre indice de stress dans la logique de l'apprentissage. Plutôt qu'un *point* unique dans un espace \mathcal{F} à K dimensions, on se propose de représenter les données par un ensemble de *flèches* pointant plus ou moins dans la même direction.

IV.D.4. Synthèse des principes essentiels et preuve de concept

Contrairement aux algorithmes usuels, l'indice de stress ne cherche pas à dessiner une frontière optimale entre les classes dans l'espace des caractéristiques. L'apprentissage de l'indice de stress s'effectue au niveau de chaque caractéristique, au travers de l'indice simple \mathcal{S} . Cette fonction transforme les *mesures* de chaque caractéristique en *niveaux de certitude* sur l'état de stress. L'indice de stress suppose ainsi que les caractéristiques utilisées ont été identifiées comme des corrélats du stress. Cette opération a été réalisée au Chapitre Cinq - III.B, au prix d'un filtrage sélectif basé sur un protocole expérimental rigoureux.

L'indice de stress poursuit un double objectif.

- (i) Détecter le stress en présence de facteurs confondants tels que l'activité cognitive et motrice, sur un sujet inconnu, dans un environnement inconnu.

(ii) Conserver sa performance dans un contexte de mesures dégradées.

L'efficacité de l'indice de stress doit être jugée à l'aune de ces deux critères. Dans la section IV.A, l'indice de stress a montré des performances comparables à de l'état de l'art et à la régression logistique, en distinguant le stress d'une activité cognitive modérée, sur un sujet et une tâche qui n'étaient pas inclus dans la base d'apprentissage. Dans la section IV.B, l'indice de stress a montré des performances équivalentes à celles de la section IV.A, pour trois cas de figure dans lesquels l'une de ses caractéristiques présentait des valeurs aléatoires. Pour parvenir à ce dernier résultat, la conception de l'indice de stress a reposé sur deux idées-clés.

- (a) Limiter la contribution de chaque caractéristique sur la décision du modèle.
- (b) Retirer des calculs chaque caractéristique qui n'est momentanément pas pertinente pour la détection du stress.

Comme on l'a vu précédemment, cette stratégie s'est révélée gagnante sur le modèle standard \mathcal{N}^1 dans la plupart des cas où l'une des caractéristiques ne reflétait plus le paramètre biologique cible. Lorsqu'un modèle définit une frontière optimale dans l'espace des caractéristiques, on peut considérer la présence de plusieurs mesures comme une multiplication des sources d'erreur si cette situation de bruit s'avère courante en vie réelle. Pour l'indice de stress, c'est une multiplication des sources d'information : sa nature *flexible* lui permet de s'adapter « en ligne » à un changement brusque dans la qualité des données. Encore faut-il un indicateur fiable pour signaler ce changement.

Cette nécessité préfigure les analyses proposées au Chapitre Sept - , où l'on cherche un critère qui permettra de définir si une fenêtre temporelle est propre au calcul de certaines caractéristiques cardiaques.

A RETENIR

- Sur des données propres, notre « indice de stress » a montré une performance supérieure à la régression logistique et inférieure à un réseau de neurone.
- Sur des données volontairement bruitées, les performances de l'indice de stress « entier » étaient généralement meilleures et plus stables que celles du réseau de neurones. En retirant du calcul les caractéristiques bruitées, la perte de performance de l'indice de stress était nulle pour trois caractéristiques sur quatre.
- Ces résultats montrent la capacité de l'indice de stress à conserver ses performances dans un contexte de mesures dégradées. Pour concrétiser son caractère adaptatif, nous avons besoin d'un indicateur de qualité sur les caractéristiques utilisées.

Chapitre Sept - Qualité des Mesures dans la Vie Quotidienne

Dans le Chapitre Six - , on a présenté un modèle de stress ayant la capacité de s'adapter à un ensemble évolutif de caractéristiques physiologiques. La structure modulaire de l'indice de stress lui permet en effet de rejeter (en temps réel) certaines caractéristiques dont la qualité ne serait pas suffisante pour détecter correctement le stress dans un contexte de mesures dégradées.

Cette adaptation au contexte de mesure n'est possible que lorsque l'indice de stress est couplé à des indicateurs pertinents sur la qualité des signaux mesurés. Ces indicateurs ne sont pas systématiquement fournis par les constructeurs de capteurs physiologiques ; et lorsque c'est le cas, le mode de calcul de l'indicateur est rarement communiqué. Ainsi pour ne pas dépendre du dispositif de mesure, un modèle de stress doit pouvoir s'appuyer sur des méthodes simples permettant d'estimer *en ligne* la qualité des signaux physiologiques utilisés.

Sur les quatre caractéristiques sélectionnées au Chapitre Cinq - , trois sont d'origine cardiaque. En laboratoire, ces caractéristiques sont presque toujours calculées à partir d'un électrocardiogramme (**ECG**), signal encore considéré comme l'*étalon-or* pour la mesure du rythme cardiaque. A l'heure actuelle, cependant, la grande majorité des dispositifs de mesure du rythme cardiaque disponibles en vie réelle utilisent la photopléthysmographie (**PPG**) : une mesure particulièrement sensible au mouvement ; et dont la capacité à représenter fidèlement la variabilité cardiaque fait encore débat. De plus, si des dispositifs ECG portables sont également accessibles au grand public, ces derniers ne sont pas non plus immuns aux artéfacts de mouvement.

Ainsi, la mesure du rythme cardiaque peut être citée en exemple dans ce qui constitue un défi majeur pour les sciences affectives : la transférabilité d'un modèle depuis des enregistrements de laboratoire vers des dispositifs de mesure en vie réelle. Dans ce contexte, nous aborderons deux problèmes successifs.

- (i) La conception d'un indicateur *boîte-blanche* sur quantité d'information disponible dans une fenêtre temporelle du signal de rythme cardiaque.
- (ii) Une mesure de la qualité du signal cardiaque issu de la mesure PPG, qui montrera comment l'indicateur précédent peut être utilisé pour prédire si une caractéristique peut être extraite de ce signal cardiaque dans un contexte ambulatoire.

Dans cette perspective, la section I de ce chapitre dresse un tableau succinct des défis relatifs à la mesure cardiaque en condition ambulatoire. Pour y répondre, on s'appuiera sur une nouvelle base de données recueillies dans la vie quotidienne de trois participants. La section II détaille le protocole expérimental associé à cette base ; ainsi que les traitements effectués sur les signaux cardiaques. La section III présente d'une part, notre indicateur de qualité du signal cardiaque, associé à un critère qui nous permettra de déterminer les fenêtres de temps sur lesquelles le signal ECG pourra servir de référence. Elle présente d'autre part, une méthode de comparaison des signaux de rythme cardiaque issus de la PPG et de l'ECG, dans les portions de signal de signal où l'ECG peut servir de référence. Enfin, les sections IV et V présentent nos résultats et leur discussion.

I. Contexte

I.A. Propriétés des Signaux Cardiaques

I.A.1. Extraction du rythme cardiaque à partir des signaux ECG et PPG

La Figure 7.1 représente dix secondes d'un signal ECG et d'un signal PPG, enregistrés sur un même sujet et synchronisés entre eux. Sur le signal ECG, les battements de cœur peuvent être repérés à l'aide des pics très nets : les complexes *QRS*, qui marquent la contraction du muscle cardiaque (systole ventriculaire). Ces pics sont rapidement suivis par une oscillation de faible amplitude : l'onde *T*. La précision temporelle offerte par le complexe *QRS* en fait un motif de choix pour la détection automatique des battements de cœur, comme dans l'algorithme de Pan & Tompkins (1985).

De même, chaque battement de cœur est repéré par un motif pulsatile sur le signal PPG. Une élévation rapide du signal (phase anacrotique) alterne avec une redescente plus lente (phase catacrotique), souvent marquée par une ou plusieurs « encoches » (Allen, 2007; Schäfer & Vagedes, 2013). La phase ascendante marque une onde de pression qui parcourt les artères suite à la contraction du muscle cardiaque ; une encoche marque une réflexion de cette onde dans les parties périphériques du système sanguin. La taille des encoches dépend ainsi de la rigidité des parois artérielles, ainsi que du lieu de la mesure PPG.

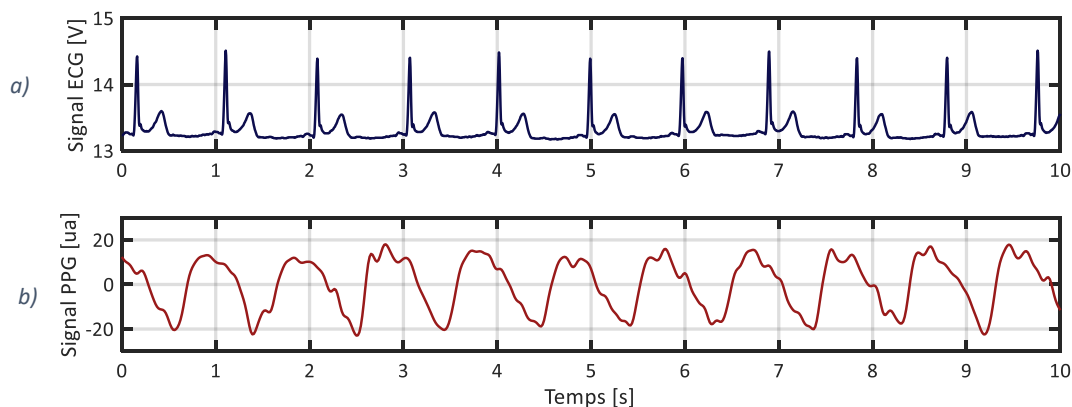


Figure 7.1 : ECG thoracique (a) et PPG du poignet gauche (b) enregistrés simultanément sur un même sujet
L'ECG provient du capteur de Zephyr BioHarness 3. Le PPG provient du capteur d'Empatica E4.

Ainsi, même lorsque les conditions de mesure sont idéales, le signal PPG présente au moins deux inconvénients par rapport à l'ECG.

- Sa forme d'onde est irrégulière, à cause de phénomènes mécaniques propres au système sanguin.
- Sa forme d'onde est oscillatoire : elle marque chaque battement de cœur avec moins de précision.

Contrairement à l'ECG, il n'existe pas de méthode standard pour extraire un signal de rythme cardiaque à partir de la mesure PPG (Schäfer & Vagedes, 2013). A partir des oscillations du signal PPG, des méthodes simples permettent d'estimer un rythme cardiaque moyen sur une fenêtre de temps donnée. On peut calculer, par exemple, la composante fréquentielle principale (analyse de Fourier), ou le délai qui maximise l'autocorrélation du signal. Mais ces méthodes ne donnent pas accès au rythme cardiaque *instantané*, essentiel à l'étude de la variabilité cardiaque. Pour repérer localement

chaque forme d'onde, on peut s'appuyer, par exemple, sur un seuillage dynamique du signal, sur les passages à Zéro de la dérivée du signal ou sur les maxima de la dérivée du signal.

I.A.2. Variabilité cardiaque

En théorie, la phase ascendante de la forme d'onde permet de détecter chaque battement avec davantage de précision. Ainsi, la pente maximale du signal PPG apparaît comme un bon marqueur de la contraction du muscle cardiaque. Mais comme dans n'importe quel fluide, l'onde de pression mesurée par le signal PPG nécessite un *temps de transit* entre la systole ventriculaire et sa propagation jusqu'aux vaisseaux sanguins périphériques. Le principe est illustré sur la Figure 7.2 qui reprend les deux premières secondes des signaux de la Figure 7.1. Entre autres paramètres, ce temps de transit dépend de la pression artérielle. Ainsi, certains auteurs ont tenté d'estimer les variations de pression sanguine à partir du délai mesuré entre l'ECG et le PPG à chaque battement de coeur (Thomas et al., 2016).

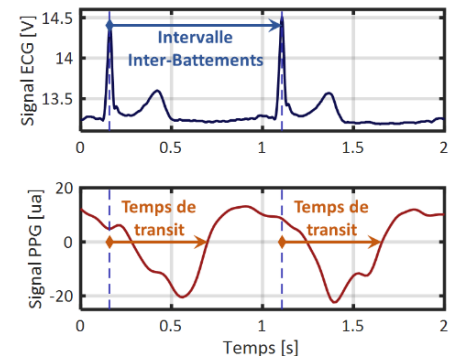


Figure 7.2 : Temps de transit de l'onde PPG

Ainsi : même dans l'hypothèse où une méthode standard permettrait d'extraire un rythme cardiaque du signal PPG, les valeurs mesurées dépendraient encore de la pression artérielle. L'interférence de ce paramètre vasculaire avec le signal cardiaque a mené certains auteurs à remplacer la notion de « variabilité cardiaque » (*heart rate variability*, ou **HRV**) par celle de « variabilité du pouls » (*pulse rate variability*, ou **PRV**), lorsqu'on mesure celle-ci *via* un signal PPG (Gil et al., 2010). A l'état de repos, les variations lentes de la pression sanguine permettent d'identifier la PRV à la HRV avec une précision satisfaisante. Cependant, cette précision diminue en condition d'exercice physique ou mental modéré (Schäfer & Vagedes, 2013). Ainsi, la mesure en ambulatoire est susceptible de générer des caractéristiques physiologiques différentes de celles observées en conditions stationnaires.

I.B. Défis de la Mesure Cardiaque en Condition Ambulatoire

I.B.1. Artéfacts de mouvement

Bien que cette mesure soit plus contraignante que la PPG, l'ECG reste ainsi l'étalon-or pour mesurer la variabilité cardiaque en condition ambulatoire. La technologie Holter enregistre ainsi l'activité cardiaque dans la vie quotidienne depuis plus d'une cinquantaine d'années (Kennedy, 2006). Mais l'installation de tels capteurs nécessite des connaissances hors de portée du grand public et l'utilisation d'électrodes intégrées dans des textiles (textrodes) fait aujourd'hui référence dans l'électrocardiographie portative (Gruetzmann et al., 2007).

Aucune méthode de mesure n'est cependant immune aux artéfacts de mouvement. Pour illustrer cet effet, la Figure 7.3 montre l'apparition d'un artéfact sur des signaux ECG et PPG enregistrés simultanément (secondes 6 à 14 sur l'ECG, 7 à 17 sur le PPG).

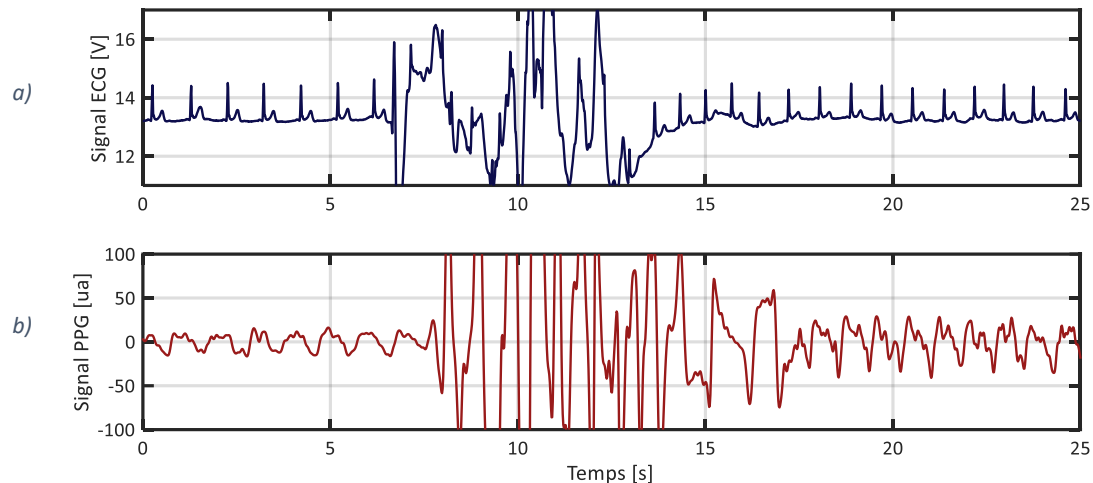


Figure 7.3 : Artéfacts de mouvement sur un signal ECG (a) et PPG (b)

L'ECG provient du capteur de Zephyr BioHarness 3. Le PPG provient du capteur d'Empatica E4.

L'amplitude du bruit généré par le mouvement dépasse largement celle du motif utile à la détection des battements cardiaques, ce qui rend temporairement *indisponible* l'extraction du rythme cardiaque à partir des signaux bruts. Ce problème est particulièrement étudié sur le signal PPG, pour lequel les techniques de retrait des artéfacts se multiplient depuis une quarantaine d'années (Allen, 2007). Les techniques les plus récentes utilisent l'apprentissage automatique, basée sur des caractéristiques temporelles et fréquentielles du signal PPG, ou l'addition de signaux d'accéléromètre (Periyasamy et al., 2017). Cela dit, aucune méthode standard n'a pour l'instant été dégagée de la littérature, chacune de ces méthodes étant validée sur une base de données différente.

Pour la mesure ECG comme pour la mesure PPG, la plupart des fabricants de capteurs livrent aujourd'hui leur produit avec une estimation propriétaire du rythme cardiaque. Mais comme le montre la Figure 7.3, la présence de mouvement en condition ambulatoire peut rendre cette estimation soit incorrecte, soit indisponible durant ces intervalles de temps. Pour ne pas fausser le calcul de caractéristiques en temps réel, un enjeu essentiel se trouve dans la conception d'un indice de qualité sur la mesure ECG ou PPG. Un certain nombre d'auteurs a travaillé dans ce sens.

I.B.2. Indices de Qualité

La plupart des indices de qualité existants s'appuient sur le signal brut renvoyé par la mesure ECG ou PPG. Sur le signal ECG, un certain nombre de caractéristiques fréquentielles ou statistiques peuvent être extraites du signal pour refléter sa qualité sur une fenêtre de temps (Aldecoa Sánchez del Río et al., 2011). Ces caractéristiques peuvent être incluses dans un algorithme d'apprentissage automatique, afin de calculer « l'acceptabilité clinique » de chaque portion de signal. Par exemple, Behar et collaborateurs (2012) ont obtenu 93% de bonnes classifications en comparant la sortie de leur algorithme à une annotation manuelle.

En ce qui concerne le signal PPG, chaque forme d'onde peut être comparée à un gabarit pré-enregistré, basé sur des connaissances *a priori* et calibré sur les formes d'onde adjacentes. Avec cette technique, par exemple, Li & Clifford (2012) ont bien classifié 95% de leurs portions de signal par rapport à l'annotation manuelle. Des variantes de cette technique permettent de concevoir un indice de qualité continu (Papini et al., 2017). Cela dit, toutes ces études centrées sur le signal ont été

conduites sur des enregistrements effectués en conditions contrôlées. Leur validité reste à démontrer en conditions ambulatoires, lorsque la qualité du signal vient à se dégrader de manière significative.

Une autre approche consiste à déterminer l'acceptabilité d'un segment de signal directement sur l'estimée du rythme cardiaque. Il s'agit d'exploiter certaines propriétés connues de la variabilité cardiaque. Par exemple : Orphanidou & Drobnjak (2017) ont fourni à un algorithme SVM des caractéristiques fréquentielles du signal cardiaque, calculées sur des segments de 30 secondes enregistrés *via* un dispositif Holter en condition ambulatoire. La précision obtenue était de 94% par rapport à l'annotation manuelle. L'avantage de cet indice uniquement basé sur le signal cardiaque, est que celui-ci s'applique indifféremment à tout type de mesure, ECG ou PPG.

Cependant, ce type de modèle est généralement développé à l'aide d'une instrumentation standard vraisemblablement mise en place par un personnel qualifié. Un tel modèle doit donc être validé (sinon *réappris*) sur des mesures de la vie réelle, enregistrées *via* des dispositifs grand public. Dans cette perspective, l'utilisation de techniques basées sur la donnée (apprentissage automatique), généralement de type *boîte-noire* (comme les séparateurs à vastes marges), implique que beaucoup d'échantillons fiables soient disponibles pour l'apprentissage. Comme on l'a vu dans les paragraphes précédents (et comme on le confirmera dans les paragraphes qui suivent), cette hypothèse est difficilement vérifiée pour le signal PPG.

I.C. Synthèse et Objectifs

Les travaux présentés dans ce chapitre n'ont pas pour but de surpasser la littérature existante en matière d'indices de qualité sur le signal de rythme cardiaque. L'un des principaux défis de la mesure en ambulatoire est l'absence totale de contrôle sur l'activité de la personne, et par conséquent sur la qualité du signal enregistré. Lorsque le signal de rythme cardiaque est défaillant, le calcul des caractéristiques cardiaques s'en trouve affecté – et les conséquences peuvent être dramatiques sur la sortie d'un modèle de stress. Ainsi, notre objectif est de proposer un critère boîte-blanche qui permettrait de garantir en temps réel la fiabilité des caractéristiques utilisées dans les chapitres précédents pour détecter le stress.

Nos analyses portent plus particulièrement sur la mesure PPG, dont la capacité à mesurer la variabilité cardiaque peut être mise en question. Sur un tel signal, l'application en temps réel d'un critère de qualité semble être indispensable à la mesure d'un rythme cardiaque valide. C'est la solution opérée par le dispositif Empatica E4 (Empatica, 2013), qui sélectionne seulement les formes d'ondes qui correspondent à un motif prédéfini avant de calculer l'intervalle inter-battements. Avec cette stratégie, une faible qualité du signal se traduit par une grande quantité d'échantillons IBI manquants. Le signal IBI_{e4} illustre très bien cette propriété : l'algorithme d'Empatica ne renvoyant un IBI que lorsque le signal est jugé suffisamment fiable, il en résulte un signal IBI parfois très rare sur une période prolongée.

Ollander et collaborateurs (2016) ont montré ce problème sur la base de données utilisée dans les chapitres précédents. Avec notre mesure ECG pour référence, ils ont pu compter le nombre d'échantillons IBI manquants sur des fenêtres de temps de 30 secondes dans chacune des conditions Stress, Contrôle et Repos. Le nombre moyen d'échantillons manquants variait fortement en fonction de la condition expérimentale, de 26% (Repos) à 62% (Stress).

Ainsi lorsqu'on sélectionne les échantillons valides du signal de rythme cardiaque, le problème de *qualité* relatif aux signaux bruts se transforme en problème de *disponibilité* des échantillons IBI sur une fenêtre de temps donnée. Ce constat est à la base de notre méthode boîte-blanche pour identifier les segments de signal exploitables sur un enregistrement donné. Contrairement à la majorité des travaux sur la qualité des signaux cardiaques, nos analyses concernent des mesures effectuées en ambulatoire. Une base de données a été conçue à partir d'une campagne expérimentale en vie réelle. Celle-ci fait l'objet de la section suivante.

II. Matériel : Base de Données en Vie Réelle

II.A. Protocole Expérimental

Les paragraphes qui suivent présentent succinctement le protocole expérimental de la base de données Semaine de Travail, développée dans le cadre de ces travaux de thèse dans le but de tester / adapter les algorithmes de détection du stress proposés au Chapitre Six - . On évoque ici seulement les problématiques qui concernent l'enregistrement et l'analyse de signaux cardiaques en vie réelle.

II.A.1. Equipement et signaux disponibles

Nos mesures cardiaques ont été obtenues au moyen d'un bracelet instrumenté et d'une ceinture thoracique, deux dispositifs proposés sur le marché public des capteurs biologiques. La ceinture thoracique est une Zephyr BioHarness 3 (Zephyr™, 2012), ici abrégée **BH3**. Le bracelet électronique est un Empatica **E4** (Empatica, 2013). Un schéma simplifié du matériel, de ses mesures et de son positionnement est donné dans la Figure 7.4. Entre autres mesures, la ceinture **BH3** fournit un

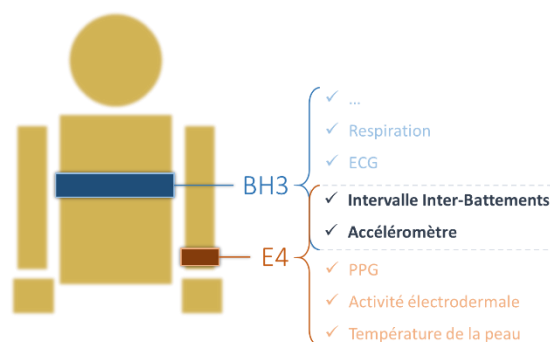


Figure 7.4 : Matériel de mesures physiologiques et signaux disponibles pour chaque appareil

électrocardiogramme (**ECG_{bh3}**) échantillonné à 250Hz, issu d'une paire de textrodes intégrée au tissu principal. Cette mesure est stockée dans un module central accroché à la ceinture, dont la capacité en mémoire et en énergie est supérieure à une journée d'acquisition. Ce module central intègre également des traitements temps-réel du signal ECG_{bh3}, parmi lesquels se trouvent deux estimations différentes du rythme cardiaque. La première est un intervalle inter-battements (**IBI_{bh3}**) échantillonné à chaque nouvelle détection d'un battement de cœur. La seconde est une approximation du rythme moyen sur 15 secondes (**HR_{bh3}**), échantillonnée toutes les secondes, calculé à partir des échantillons adjacents du signal IBI_{bh3}. La validité du signal HR_{bh3} comme mesure continue du rythme cardiaque a été vérifiée par Johnstone et collaborateurs (2012), en conditions d'activité physique faible et forte.

En parallèle, la ceinture BH3 fournit un *niveau de confiance* (C_{bh3}) sur le signal HR_{bh3} , aussi échantillonné toutes les secondes, assimilable à un indice de qualité. Selon la notice de l'appareil, cet indice est calculé à partir du maximum et du rapport signal-sur-bruit du signal ECG_{bh3} . Enfin, BH3 fournit un signal d'accéléromètre 3 axes (A_{bh3}), échantillonné à 100 Hz. Ce signal est indispensable pour synchroniser correctement l'appareil et a servi de base au calcul d'un indice d'activité (section II.B).

De son côté le bracelet de mesures **E4** fournit un photoplétysmogramme (PPG_{e4}) échantillonné à 64Hz, obtenu à partir de diodes vertes et rouges et d'un algorithme de traitement propriétaire d'Empatica. De la même manière, un algorithme propriétaire renvoie un signal d'intervalle interbattement (IBI_{e4}). Contrairement à la ceinture BH3, cependant, ce signal est échantillonné seulement lorsque la qualité du signal PPG_{e4} est jugée suffisante par l'algorithme d'Empatica, qui s'appuie notamment sur une comparaison de motifs (Empatica, 2019). Implicitement, le bracelet E4 utilise ainsi un indice de qualité dont la valeur n'est pas connue de l'utilisateur. Pour finir, E4 fournit un signal d'accéléromètre 3 axes (A_{e4}), échantillonné à 32Hz, qui a servi de base au calcul d'un indice d'activité (section II.B).

La ceinture BH3 était portée sur la partie inférieure du buste, le boîtier d'enregistrement situé au niveau de l'aire précordiale. Le bracelet Empatica E4 était porté sur le poignet de la main non-dominante, le boîtier d'enregistrement situé sur la face postérieure (dos de la main). Ces deux dispositifs étaient positionnés par les participants eux-mêmes au début de chaque journée d'enregistrement ; puis enlevés à chaque fin de journée, selon un protocole bien défini.

II.A.2. Procédure expérimentale et marquage des signaux

La Figure 7.5 résume la journée standard d'un(e) participant(e) à la campagne expérimentale. Il était demandé aux participants d'installer les capteurs le plus tôt possible après le lever ; puis de les mettre en route une fois installés. Après la phase d'initialisation des capteurs (1min), les participants effectuaient trois sauts sur eux-mêmes, séparés par des intervalles d'environ une seconde. Cette manœuvre avait pour but de marquer les signaux d'accéléromètre de chaque capteur, en vue d'une synchronisation *a posteriori* des deux dispositifs. Après une période de repos de cinq minutes, les participants pouvaient oublier la présence des capteurs et suivre librement leur emploi du temps.

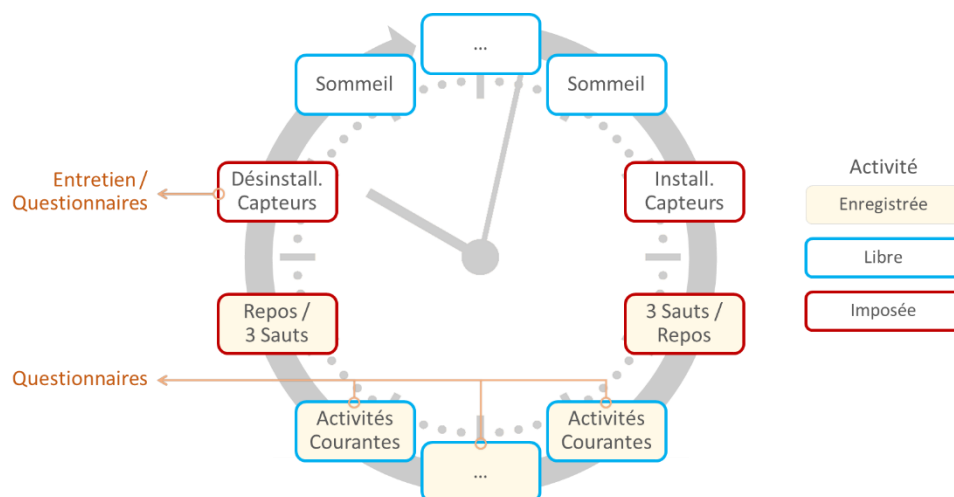


Figure 7.5 : Protocole Semaine de Travail – Journée de mesures du point de vue d'un participant

La plus grosse partie du protocole commençait le soir, le plus tard possible avant le coucher. De la même manière qu'en matinée, les participants effectuaient cinq minutes de repos puis trois sauts sur

eux-mêmes espacés d'une seconde. Cette manœuvre laissait une seconde marque sur les signaux d'accéléromètre de chaque capteur. Ceux-ci étaient alors éteints puis retirés par les participants, installés sur leurs supports de charge et connectés à l'ordinateur. Les participants devaient télécharger eux-mêmes les données dans l'ordinateur, puis les envoyer sur un serveur sécurisé qui permettait à l'expérimentateur de vérifier chaque jour la qualité des données enregistrées.

Dans cette routine quotidienne, les participants n'étaient pas livrés à eux-mêmes. Chaque élément du protocole était guidé par un manuscrit, reçu le premier jour d'expérience. Tout au long de la journée, il était demandé aux participants de reporter chaque événement pertinent sur une application smartphone conçue pour l'expérience. Après retrait des capteurs, le soir, un questionnaire leur demandait de retracer les principales séquences de la journée. Pour terminer, l'expérimentateur les contactait pour un court entretien (par téléphone, mail ou SMS, selon les besoins et disponibilités), après avoir observé les données téléchargées sur le serveur distant.

II.A.3. Recrutement et données disponibles

Les participants étaient recrutés sur annonce. Ceux-ci ne devaient pas présenter de maladie neurologique ou cardiaque récente ou passée, ni suivre de traitement susceptible d'altérer leurs fonctions cérébrales ou neurologiques. Après les formalités de recrutement, chaque participant(e) venait au laboratoire pour un entretien de deux heures au cours duquel le protocole expérimental était expliqué en détail (théorie et pratique). A la fin de cet entretien, le(la) participant(e) signait un manifeste de consentement éclairé et les capteurs étaient remis en mains propres à l'expérimentateur. Le(la) participant(e) gardait les capteurs pour environ une semaine, au terme de laquelle il(elle) revenait au laboratoire pour un entretien final et la restitution des capteurs.

Les données utilisées dans le cadre de cette étude concernent à ce jour trois participants de sexe masculin, âgés de 25, 27 et 33 ans au moment de l'expérience. Deux d'entre eux étaient des sportifs de très haut niveau, partis effectuer un stage d'entraînement à l'étranger. Le troisième était un *business developer* travaillant près du laboratoire. Ainsi, les mesures obtenues au moyen de ce protocole couvrent un très large spectre des activités quotidiennes et d'effort physique : du voyage en camion aux trajets en vélo, du développement musculaire au travail de bureau.

Comme chaque base de mesures en vie réelle, provenant (qui plus est) de capteurs commerciaux qui ne communiquent pas entre eux, ces données ont nécessité un travail rigoureux d'inspection visuelle et de synchronisation. Leurs pré-traitements seront développés dans la section suivante. Lors de l'expérience, des problèmes techniques liés aux capteurs eux-mêmes, ou liés au non respect du protocole expérimental, ont compliqué la synchronisation *post-hoc* de certaines journées d'acquisition. Afin de travailler sur des données valides, les journées dont les signaux n'ont pas été parfaitement synchronisés au moyen du protocole ci-après ont été retirées de notre base de données. Au total, celle-ci cumule 11 journées d'enregistrement – soit 124 heures de mesures pour nos deux capteurs.

II.B. Traitements des Signaux

Un signal de référence et une estimation de l'activité physique sont pertinents pour interroger la fiabilité des signaux IBI collectés. D'autre part, les analyses proposées en section III s'appuient sur des

signaux IBI parfaitement synchronisés et débarrassés de leurs valeurs aberrantes. Ces prétraitements font l'objet des paragraphes suivants.

II.B.1. Signaux additionnels

a) Référence académique sur le rythme cardiaque

Dans cette base de données, les signaux IBI disponibles ont été calculés par des algorithmes non divulgués par les fabricants des capteurs. Le calcul d'une référence académique peut fournir un élément de comparaison pour mesurer l'efficacité de ces « boîtes-noires » en conditions réelles. L'algorithme de Pan & Tompkins (1985) a été appliqué sur tous les enregistrements disponibles du signal ECG_{bh3}, à partir de l'implémentation Matlab proposée par Sedghamiz (2014). Après détection des pics R, un troisième signal (IBI_{p&t}) pouvait se superposer à ceux renvoyés par E4 et BH3.

b) Mesure de l'activité physique

En condition ambulatoire, l'activité physique est un facteur d'erreur qu'on ne peut ignorer pour ces trois estimées du rythme cardiaque. Une observation continue de ce facteur peut être pertinente pour expliquer d'éventuelles défaillances dans la détection des battements cardiaques ; et ceci d'autant plus que la quantité de mouvement n'est pas la même pour nos deux capteurs, BH3 étant située sur le torse et E4 sur le poignet de la main non-dominante.

A partir de ses signaux d'accéléromètre, BH3 renvoie son propre indicateur de l'activité physique, échantillonné à 1Hz. Selon la documentation du capteur, cette estimation se base sur la norme euclidienne des trois composantes du signal, après filtrage passe-bande. Ces informations ne sont pas suffisantes pour reproduire l'estimation sur le bracelet E4, qui ne fournit de son côté que des signaux d'accéléromètre bruts. Afin de mesurer la quantité de mouvement des deux capteurs sur des échelles comparables, on a constitué notre propre indicateur d'activité d'après la formule suivante :

$$Act(\tau) = \frac{1}{L} \sum_{i \in \{x,y,z\}} \int_{t=t_n}^{t_n+L} |A^i(t)| dt \quad (7-1)$$

Où τ représente la fenêtre de temps de longueur L qui commence à l'instant t_n . Chaque terme A^i représente la i -ème composante du signal d'accéléromètre ($i \in \{x,y,z\}$), après filtrage passe-bande dans l'intervalle de fréquences : $[0.1 ; 10]$ Hz. D'après certains travaux sur la quantification du mouvement humain (Bouten et al., 1997; Karantonis et al., 2006), l'estimateur Act proposé par l'équation (7-1) est corrélé à la dépense d'énergie liée au mouvement ; et distingue efficacement les périodes avec et sans activité physique. Sur l'ensemble des fenêtres de 60 secondes disponibles dans la base de données, la corrélation de Spearman mesurée entre l'indicateur Act_{bh3} et la moyenne de l'indicateur natif du capteur BH3 est de 0,91.

II.B.2. Synchronisation des mesures

a) Inter-Capteurs

Les analyses à suivre nécessitent une synchronisation rigoureuse des trois estimations de l'IBI (au battement de cœur près). Afin d'assurer cette contrainte, on a procédé à un marquage des signaux d'accéléromètre bruts fournis par E4 et BH3. On a d'abord calculé la norme euclidienne des trois composantes fournies par nos deux capteurs : $\|A_{bh3}\|$ et $\|A_{e4}\|$, afin d'obtenir un signal d'accélération

unique et temporellement précis pour chacun d'eux. Ces deux signaux sont affichés sur le graphique (a) de la Figure 7.6, sur une journée standard d'acquisition.

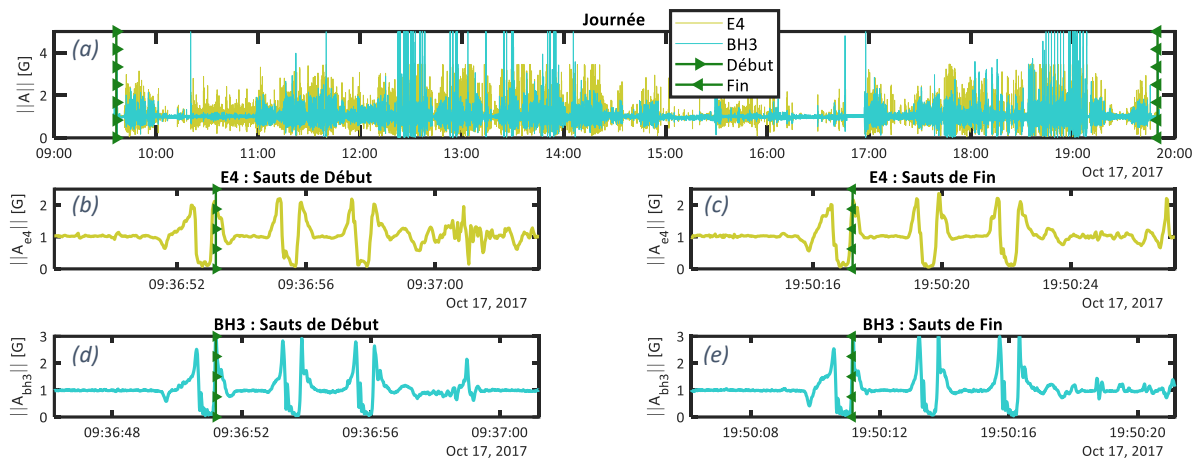


Figure 7.6 : Synchronisation des signaux d'accéléromètre pour E4 et BioHarness

(a) : Journée complète d'enregistrement.

(b) : Marque de début pour Empatica E4. (c) : Marque de fin pour Empatica E4.

(d) : Marque de début pour Zephyr BioHarness 3. (e) : Marque de fin pour Zephyr BioHarness 3.

Pour chacun de ces signaux, les trois sauts prévus par le protocole expérimental au début et à la fin de la journée sont affichés sur les graphiques (b, c, d, e). Chaque saut présente un motif bien particulier sur la norme euclidienne de l'accélération. (i) Un premier pic positif marque l'action de décollage. (ii) Durant la chute libre (*i.e.* : dès que les pieds quittent le sol), l'accélération gravitationnelle n'est plus compensée par un support et le signal montre un creux proche de 0G. (iii) L'atterrissage est marqué par un second pic positif avant que le signal ne retourne à 1G. L'enchaînement de 3 sauts correspond à un motif unique, identique pour chaque signal d'accélération, facilement repérable à proximité d'une période de repos.

A partir de ces motifs, une marque de début et de fin d'enregistrement a été placée sur le pic d'atterrissage correspondant au premier des trois sauts (ligne verticale sur chacun des graphiques). Ces marques ont permis de recalibrer précisément l'horloge d'un capteur sur l'autre, avec une précision de l'ordre du dixième de seconde. Comme les signaux d'accéléromètre, les signaux cardiaques des deux capteurs étaient ainsi (théoriquement) synchronisés. Pour vérifier la qualité de cette synchronisation, les quatre estimées disponibles du rythme cardiaque (IBI_{e4} , IBI_{bh3} , $IBI_{p\&t}$ et HR_{bh3} , cette dernière ayant été inversée pour correspondre aux valeurs des 3 autres) ont été juxtaposées sur un même graphique, puis inspectées visuellement pour chaque enregistrement.

b) Intra-Capteur

Lors de cette inspection, il est apparu que le signal IBI_{bh3} pouvait se désynchroniser soudainement des trois autres signaux, en affichant un retard de plusieurs minutes jusqu'à la fin de l'enregistrement. Ce décalage ne concernant pas l'indicateur de référence IBI_{pt} , ni l'estimée moyenne HR_{bh3} , les signaux cardiaques de la ceinture BH3 nécessitaient une synchronisation *intra-capteur* avant d'être comparés au signal du bracelet E4.

Lors du téléchargement des données, le signal IBI_{bh3} était présenté sous la forme d'un *tacogramme* : un horodatage initial (ex. : 10/10/2016 – 08:30:00.000) suivi d'une séquence de durées en millisecondes (ex. : [950, 990, ...]), correspondant chacune à l'intervalle qui sépare chaque nouveau battement détecté du battement précédent. Ce signal, vraisemblablement codé sur 16 bits, ne

dépassait jamais la valeur : $2^{15} \approx 33000$ ms. Sur certains enregistrements, cependant, il est arrivé que la qualité du signal ECG reste suffisamment faible pour empêcher la détection de nouveaux battements durant plusieurs minutes. Ainsi, la désynchronisation du signal IBI_{bh3} survenait dès que le signal ECG demeurait inutilisable durant plus de 33 secondes.

Suite à ce constat, une méthode semi-automatique a été mise en œuvre pour repérer chaque désynchronisation sur le signal BI_{bh3} . Les portions de signal concernées étaient isolées, puis recalées sur le signal HR_{bh3} en maximisant leur corrélation avec ce dernier. Cette translation du signal était affinée à la main pour tenir compte d'éventuelles imprécisions dans le calcul des délais.

Pour terminer, le signal cardiaque du bracelet E4 a été recalé sur ceux de la ceinture BH3 par maximisation de la corrélation entre les deux signaux (calcul vérifié puis affiné à la main). Cette manoeuvre avait pour but de neutraliser le *temps de transit* entre les deux signaux cardiaques. Après ces manoeuvres, la trame temporelle de nos deux signaux IBI était optimisée pour une identification battement-à-battement (voir section III.B).

II.B.3. Retrait des valeurs aberrantes

La présence de valeurs aberrantes est à prévoir sur tout signal IBI mesuré en conditions réelles. Or, les analyses à venir supposent que nos signaux reflètent des valeurs correctes du rythme cardiaque. Toute valeur impossible doit être considérée comme un battement manqué. Pour nettoyer nos signaux de ces mauvaises mesures, on peut s'appuyer sur des propriétés connues du rythme cardiaque.

Notre méthode de détection des valeurs aberrantes est automatique et susceptible d'être appliquée en temps réel. La stratégie adoptée et les critères choisis s'inspirent de recherches analogues pour une méthode boîte-blanche de nettoyage des signaux cardiaques (Karlsson et al., 2012; Kemper et al., 2007). Les valeurs aberrantes sont retirées du signal à partir de deux critères successifs.

- Critère d'intervalle : toute valeur d'IBI qui correspond à un rythme cardiaque en-dessous de 30 ou au-dessus de 250 battements par minute est retirée du signal.
- Critère de variations : toute valeur d'IBI qui dévie de plus de 30% des valeurs voisines est retirée du signal.

L'intervalle [30 ; 250]bpm a été choisi large pour prendre en compte des activités sportives intenses. De son côté, le critère de variations repose sur trois étapes clés. i) La moyenne du signal IBI a été calculée sur une fenêtre glissante de 5 secondes autour de chaque échantillon. ii) Une déviation limite de 30% a été définie au-dessus et en-dessous de cette valeur moyenne. iii) Chaque échantillon hors-limite a été temporairement défini comme valeur aberrante et retiré du signal IBI. Selon les

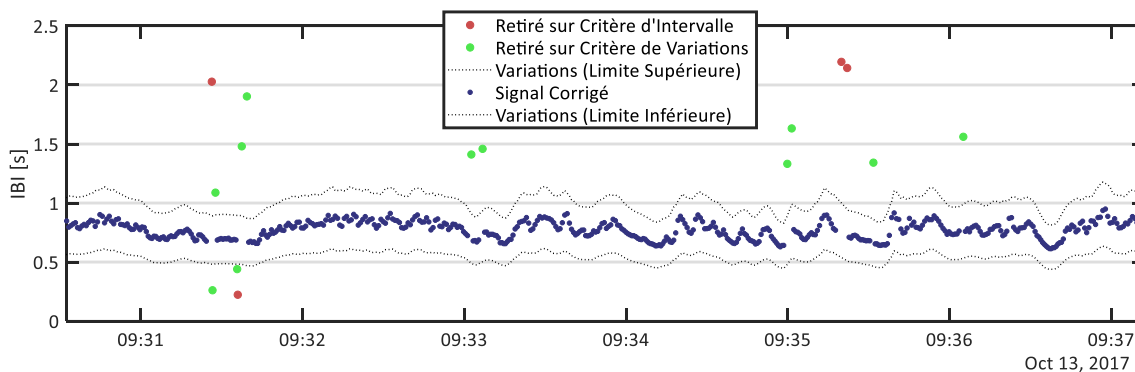


Figure 7.7 : Retrait des valeurs aberrantes sur un signal d'intervalles inter-battements

recommandations de Karlsson et collaborateurs (2012), cet algorithme a été appliqué de manière récursive pour tenir compte des portions les plus bruitées du signal. Ainsi, les étapes i), ii) et iii) ont été réitérées jusqu'à ce qu'aucune valeur aberrante ne soit plus trouvée, ou que l'ensemble du processus dépasse 20 itérations. Une illustration de cette méthode est proposée dans la Figure 7.7.

Le processus de retrait des valeurs aberrantes a été appliqué de la même manière aux trois signaux IBI disponibles dans cette étude : IBI_{bh3} , $IBI_{p\&t}$ et IBI_{e4} . Implicitement, celui-ci applique aux signaux IBI un premier critère de qualité. Il introduit des échantillons *manquants* dans les signaux IBI, aux endroits où ceux-ci ne validaient pas les propriétés d'intervalle et de variation du rythme cardiaque. A l'inverse, il laisse *disponibles* les échantillons IBI qui vérifient ces propriétés.

III. Méthodes

Dans les sections qui suivent, une méthode en boîte-blanche est proposée pour identifier des segments de signal IBI qui ne présentent aucun défaut dans la détection des battements de cœur (III.A). Ces segments IBI_{ref} issus de la mesure ECG serviront de référence pour mesurer la qualité des segments IBI_{e4} issus de la mesure PPG (III.B).

III.A. Estimation de la Part d'Echantillons Manquants dans un Signal IBI

III.A.1. Segment de signal complet : définition et propriétés

Soit un signal IBI quelconque, observé sur une fenêtre de temps de longueur L qui contient N échantillons de ce signal. On considère que cette portion est *complète* lorsqu'il ne « manque » aucun échantillon IBI sur la fenêtre de temps. Cela revient à dire que tous les échantillons sont « disponibles » ou que la détection automatique des battements de cœur n'a pas été corrigée sur cette portion de signal. La valeur de chaque échantillon s'exprime ainsi : $IBI^n = t_n - t_{n-1}$; où : $\{t_1, t_2, \dots, t_N\}$ marquent l'instant de détection de chacun des N battements. Cette situation est représentée sur la Figure 7.8.

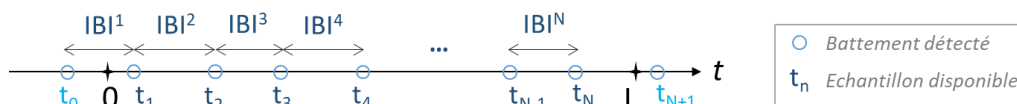


Figure 7.8 : Portion complète de signal IBI

Ici, la fenêtre de temps démarre à l'instant $t=0$. Les instants t_0 et t_{N+1} n'appartiennent pas à la fenêtre $[0 ; L]$; mais le calcul du premier échantillon (IBI^1) implique l'instant t_0 . Lorsque la fenêtre $[0 ; L]$ contient suffisamment de battements, une propriété intéressante du signal IBI se traduit dans l'approximation suivante sur la somme de ses éléments.

$$\sum_{n \in \{1..N\}} IBI^n = \sum_{n \in \{1..N\}} (t_n - t_{n-1}) = t_N - t_0 \approx L \tag{7-2}$$

Littéralement : la somme des échantillons IBI sur la fenêtre $[0 ; L]$ vaut exactement $t_N - t_0$. A un battement de cœur près, cette somme vaut également L . Cette relation, qui n'est valable que pour les segments *complets*, est à la base de notre méthode pour identifier ces segments.

Pour préparer les calculs à suivre, on peut considérer une situation où l'on dispose seulement de l'IBI moyen sur la fenêtre de temps (μ_{IBI} , c'est-à-dire la somme des IBI^n divisée par N), sans plus avoir accès au signal original. A partir de l'équation (7-2), on peut alors *estimer* (au battement de cœur près) le nombre d'échantillons disponibles sur la fenêtre de temps. On définit ainsi le nombre théorique N_t .

$$N_t \stackrel{\text{def}}{=} L / \mu_{IBI}$$

$$N_t \approx \frac{\sum_{n \in \{1..N\}} IBI^n}{\mu_{IBI}} = N \quad (7-3)$$

III.A.2. Critère de qualité d'un signal IBI : l'indice de carence (\mathcal{X})

Supposons maintenant qu'après avoir été nettoyé par un processus automatique, le signal IBI présente des échantillons *manquants* sur la fenêtre de temps $[0 ; L]$. Dans le cas des signaux IBI_{bh3} et $IBI_{p\&t}$, ce processus correspond au retrait des valeurs aberrantes. Dans le cas du signal IBI_{e4} , c'est l'effet conjugué de ce nettoyage et de l'algorithme Empatica (inconnu) qui a sélectionné seulement les formes d'onde valides sur le signal PPG. Pour nos trois signaux IBI, on peut toujours compter le nombre d'échantillons *disponibles* ; mais pas le nombre d'échantillons *manquants*.

Cette situation est illustrée dans la Figure 7.9. On imagine ici qu'à l'instant t_2 , un battement de cœur a été manqué par l'algorithme de détection sur la mesure PPG ou ECG. L'IBI calculé entre les instants t_1 et t_3 a été identifié comme valeur aberrante (soit par l'algorithme d'Empatica, soit par nos propres traitements) ; puis retirée du signal, qui est donc incomplet sur l'intervalle $[0 ; L]$. Au total, notre signal IBI comporte N' échantillons disponibles sur cette fenêtre de temps ($N' < N$).

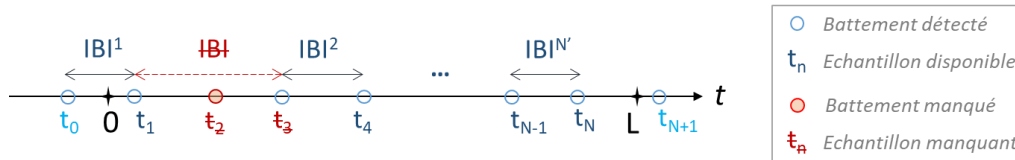


Figure 7.9 : Portion incomplète de signal IBI

Pour connaître la quantité d'information perdue lorsque la qualité de la mesure (PPG ou ECG) n'était pas suffisante, on souhaite estimer la *proportion d'échantillons manquants* sur cette portion de signal IBI. Pour cela, on peut calculer le nombre théorique N_t vu dans l'équation (7-3). En toute rigueur, il faudrait alors considérer que la moyenne $\mu_{IBI'}$ du signal (celle des échantillons *disponibles* sur cette fenêtre de temps) est différente de la moyenne μ_{IBI} qu'on aurait calculée si ce segment de signal était *complet* (comme dans l'équation (7-3)).

Cela étant dit, on peut supposer que : (i) les valeurs aberrantes ayant été retirées du signal IBI, les échantillons disponibles reflètent un rythme cardiaque réel. (ii) La variabilité cardiaque étant limitée par la physiologie, un petit nombre d'échantillons suffit à représenter le rythme moyen si la fenêtre de temps est suffisamment courte. On peut alors identifier la moyenne du signal *disponible* à celle du signal *complet* : $\mu_{IBI'} \approx \mu_{IBI}$. Cette hypothèse de conservation sera vérifiée en section IV.B.

Dans les paragraphes qui suivent, on considère que l'estimation N_t définie dans l'équation (7-3) est correcte quel que soit le nombre d'échantillons disponibles sur la fenêtre de temps. On définit alors l'*indice de carence* (\mathcal{X}) comme déviation relative entre ce nombre de battements théorique et le nombre d'échantillons disponibles (N'). Ce calcul permet d'estimer la part d'échantillons manquants.

$$\mathcal{X} \stackrel{\text{def}}{=} \frac{L/\mu'_{IBI} - N'}{L/\mu'_{IBI}} \quad (7-4)$$

$$\mathcal{X} \approx \frac{N - N'}{N}$$

III.A.3. Critère de recherche des segments de signal complets

Lorsque l'algorithme de détection des battements du cœur est correct, l'indice de carence a valeur d'indice de qualité. Un indice \mathcal{X} élevé révèle une incapacité de l'algorithme à extraire un signal IBI de la mesure initiale (ECG, PPG). Lorsqu'aucun échantillon IBI n'est extrait sur une fenêtre de temps, sa valeur limite est : $\mathcal{X}=1$. Sur une fenêtre où le signal IBI est *complet*, on obtient : $\mathcal{X}\approx 0$. Réciproquement, peut-on définir les valeurs de \mathcal{X} qui montrent que le segment de signal IBI est *complet* ?

On trouve une solution à ce problème en revenant à l'exemple de la Figure 7.8. Lorsque le segment IBI est *complet*, les échantillons disponibles sont tous les $\{IBI^n\}$ mesurés aux instants $\{t_n\}$, où $n \in \{1..N\}$. On peut également exprimer l'indice \mathcal{X} à partir de la somme des échantillons $\{IBI^n\}$: l'équation (7-5) ci-dessous vient directement de l'équation (7-4) du paragraphe précédent.

$$\mathcal{X} = \frac{L - \sum_{n \in \{1..N\}} IBI^n}{L} \quad (7-5)$$

On cherche à encadrer la valeur de \mathcal{X} avec des valeurs connues. Par définition, les instants t_0 et t_{N+1} sont à l'extérieur de la fenêtre de temps de longueur L . Ce qui nous amène à l'inégalité ci-dessous :

$$(t_N - t_1) < L < (t_{N+1} - t_0) \quad (7-6)$$

Et aux deux inégalités suivantes, qui sont équivalentes.

$$\frac{t_0 - t_1}{L} < \frac{L - \sum_{n \in \{1..N\}} IBI^n}{L} < \frac{t_{N+1} - t_N}{L} \quad (7-7)$$

$$\frac{-IBI^1}{L} < \mathcal{X} < \frac{IBI^{N+1}}{L}$$

La dernière inégalité définit toutes les valeurs de l'indice \mathcal{X} qui correspondent à un segment IBI *complet*. Comme la présence d'un échantillon manquant augmente la valeur de \mathcal{X} , on peut supposer qu'un segment IBI est *complet* si l'indice \mathcal{X} est *inférieur* à la quantité : IBI^{N+1}/L . En pratique, cependant, on ne peut garantir que l'échantillon IBI^{N+1} sera bien *disponible* sur la fenêtre de temps suivante. A partir des seuls échantillons disponibles sur la fenêtre courante, la valeur minimale des échantillons $\{IBI^n\}$ permet de produire un critère conservatif pour déterminer si cette fenêtre est bien *complète*.

Ainsi, soit un signal IBI_k issu d'un algorithme d'extraction quelconque (IBI_{bh3} , IBI_{e4} , $IBI_{p\&t}$) ; et découpé en fenêtres de temps τ de taille L . Pour déterminer l'ensemble \mathcal{C} des fenêtres sur lesquelles ce signal est *complet*, on utilisera le critère d'intégrité ci-dessous.

$$k \in \{bh3, p\&t, e4\}, \quad \tau \in \mathcal{C}(\mathcal{X}_k) \Leftrightarrow \mathcal{X}_k(\tau) < \frac{1}{L} * \min_{IBI_k^n \in \tau} (IBI_k^n) \quad (7-8)$$

A nouveau, cette relation constitue un critère de qualité sur les signaux cardiaques.

III.B. Comparaison PPG-ECG

III.B.1. Identification battement à battement

La sélection de segments IBI complets (à partir de l'indicateur de carence \mathcal{X} présenté ci-dessus) nous permet de définir un signal de référence IBI_{ref} issu de l'ECG (ici enregistré via la ceinture BioHarness), que l'on sait être d'excellente qualité. Ce signal de référence va pouvoir être utilisé pour évaluer la qualité du signal IBI_{e4} (issue du bracelet Empatica), sur ces mêmes fenêtres de temps. Dans cet objectif, la synchronisation de ces signaux (IBI_{ref} et IBI_{e4}), à la fraction de seconde près (voir section II.B), nous permet d'identifier chaque battement de cœur en superposant les enregistrements. Une illustration de cette démarche est proposée dans la Figure 7.10.

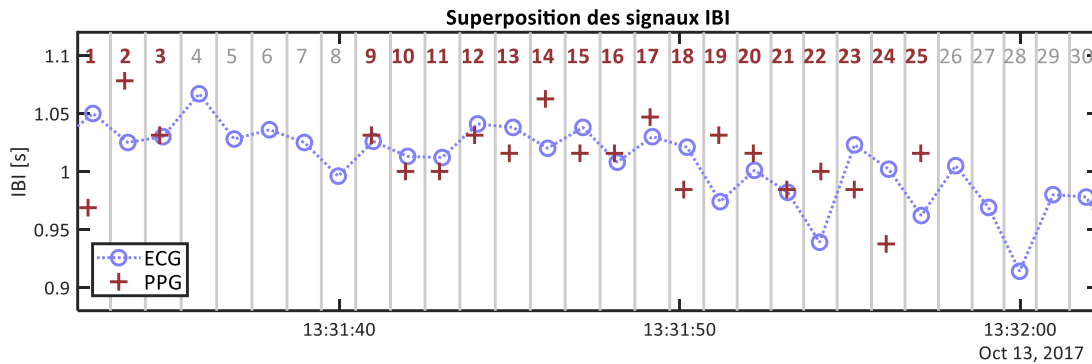


Figure 7.10 : Identification battement à battement des signaux IBI_{bh3} et IBI_{e4} sur une période de trente secondes

Un numéro est attribué à chaque battement en haut du graphique. Les numéros affichés en rouge marquent une correspondance entre les échantillons des deux signaux. Les numéros affichés en gris marquent un échantillon manquant sur le signal IBI_{e4} .

Sur une fenêtre de temps τ quelconque, on peut compter le nombre d'échantillons du signal IBI_{e4} proches de chaque échantillon IBI_{ref} . A chaque itération, ce décompte possède trois cas de figure :

- Soit l'échantillon IBI_{ref} ne correspond à *aucun* échantillon IBI_{e4} . Dans ce cas, on peut incrémenter un compteur d'échantillons manquants (N_m) sur la fenêtre de temps.
- Soit l'échantillon IBI_{ref} correspond à *plusieurs* échantillons IBI_{e4} . Dans ce cas, on peut incrémenter un compteur d'échantillons surnuméraires (N_s) sur la fenêtre de temps.
- Soit l'échantillon IBI_{ref} correspond à *un* (seul) échantillon IBI_{e4} . Dans ce cas, on peut mesurer la différence absolue entre les valeurs de ces deux échantillons.

A la fin du décompte, N_{e4} échantillons IBI_{e4} ont été liés à autant d'échantillons IBI_{ref} (sur les N_{ref} battements de la fenêtre de temps). La qualité du signal IBI_{e4} sur la fenêtre τ peut être évaluée par trois indicateurs : la *part d'échantillons manquants* dans le signal IBI_{e4} (**PEM**), le *nombre d'échantillons surnuméraires* (N_s) et la *déviaton absolue moyenne* (**DAM**) entre les échantillons IBI_{e4} et IBI_{ref} .

$$PEM(\tau) = \frac{N_m}{N_{ref}}, \quad DAM(\tau) = \frac{1}{N_{e4}} \sum_{n \in \{1..N_{e4}\}} |IBI_{e4}^n - IBI_{ref}^n| \quad (7-9)$$

La part d'échantillons manquants est l'équivalent de l'indice de carence \mathcal{X} lorsqu'un signal est disponible pour servir de référence. Si l'algorithme d'Empatica est efficace, alors aucun échantillon surnuméraire ne devrait être trouvé sur la majorité des fenêtres de temps. Enfin, la déviaton moyenne correspond à l'erreur qui peut être attendue sur la valeur de IBI_{e4} dans la fenêtre de temps. Cette erreur est due à l'imprécision de la détection des battements de cœur sur la forme d'onde du signal PPG (voir section I.A).

III.B.2. Caractéristiques cardiaques

Pour finir, notre objectif principal dans cette seconde partie était d'estimer l'erreur qui doit être attendue lorsqu'on calcule, les caractéristiques cardiaques sélectionnées au Chapitre Cinq -, à partir du signal IBI_{e4} . Ces caractéristiques sont : la moyenne du rythme cardiaque (**hrMean**), son écart-type (**hrStd**) et le carré moyen de ses différences successives (**hrRmssd**). On définit l'erreur \mathcal{E}_f sur la caractéristique f comme la déviation absolue entre la valeur f_{e4} calculée à partir du signal IBI_{e4} ; et la valeur f_{ref} calculée à partir du signal IBI_{ref} . Pour que cette erreur soit comparable entre les trois caractéristiques, on l'exprime en unités d'écart-type avec la formule suivante :

$$f \in \{hrMean, hrStd, hrRmssd\}, \quad \mathcal{E}_f(\tau) = \left| \frac{f_{e4} - f_{ref}}{\sigma_f} \right| \quad (7-10)$$

Où τ représente la fenêtre de $L = 60s$ sur laquelle on a calculé f et σ_f est l'écart-type de f estimé sur toute la journée d'acquisition correspondant à cette fenêtre τ . L'erreur \mathcal{E} a ainsi été calculée pour nos trois caractéristiques, sur toutes les fenêtres de temps où le signal IBI_{ref} pouvait servir de référence.

IV. Résultats

IV.A. Analyse des Segments Complets des Signaux IBI

Dans la section III.A, on a proposé un critère de sélection des segments *complets* du signal IBI à partir de l'indice de carence \mathcal{X} . Cet indice et ce critère sont respectivement définis aux équations (7-4) et (7-8). Ce dernier critère peut être appliqué indistinctement sur le signal IBI issu de la ceinture BioHarness 3 (IBI_{bh3}), sur le signal IBI extrait de l'ECG via l'algorithme de Pan & Tompkins (1985) ($IBI_{p\&t}$) et sur le signal IBI issu du bracelet Empatica E4 (IBI_{e4}).

Pour valider ce critère avec une référence existante, on peut utiliser l'indice de qualité natif de BioHarness : C_{bh3} , qui fournit chaque seconde un « niveau de confiance » (entre 0% et 100%) sur l'extraction du signal IBI_{bh3} à partir de l'ECG. Sur la base de cet indicateur, le critère le plus conservatif consiste à sélectionner les fenêtres de temps τ pour lesquelles C_{bh3} n'est jamais inférieur à 100%.

$$\tau \in \mathcal{C}(C_{bh3}) \Leftrightarrow \min_{C_{bh3} \in \tau} (C_{bh3}) = 100\% \quad (7-11)$$

Comme $\mathcal{C}(\mathcal{X}_{bh3})$, ce critère doit garantir l'absence d'échantillons manquants sur le signal IBI_{bh3} .

A partir de ces critères, on peut représenter la proportion de segments IBI *complets* parmi l'ensemble des segments contenus dans la base de données. Cette proportion est affichée sur la première ligne du Tableau 7.1 (plus bas). Par cohérence avec la taille des fenêtres de temps utilisées depuis le Chapitre Quatre -, la sélection est ici proposée pour des segments de longueur $L=60s$.

En parallèle, on peut estimer la quantité de mouvement disponible sur ces ensembles \mathcal{C} de segments qui ne contiennent pas d'échantillon manquant. Cette estimation nous permet d'évaluer l'impact de l'activité physique sur la qualité des mesures – et donc sur la perte d'information dans le signal IBI. Le niveau d'activité physique est estimé par l'indicateur Act proposé dans l'équation (7-1). Pour chaque critère, la valeur maximale de cet indicateur sur les segments *complets* est affichée à la ligne 2 pour la ceinture BH3 et à la ligne 3 pour le bracelet E4. Il s'agit d'une estimation optimiste sur

la quantité de mouvement (au niveau de la poitrine et du poignet) que l'on peut espérer observer via nos deux capteurs sans perte d'information sur le signal IBI.

Tableau 7.1 : Sélection des segments complets pour chaque indice de qualité et niveau maximal d'activité sur ces segments

Indice de qualité	C_{bh3}	\mathcal{X}_{bh3}	$\mathcal{X}_{p\&t}$	\mathcal{X}_{e4}
Taux de segments complets	51,4 %	52,4 %	47,6 %	0,60 %
Activité maximale sur BH3 (poitrine)	137 mG	137 mG	137 mG	7,05 mG
Activité maximale sur E4 (poignet)	415 mG	415 mG	415 mG	19,1 mG

Ligne 1 : part des fenêtres de 60s sélectionnées par le critère d'intégrité, sur toute la base de données, pour chaque indice de qualité (C , \mathcal{X}) donné sur la ligne d'en-tête. Lignes 2 et 3 : valeur maximale de l'indicateur d'activité physique Act, calculé séparément pour les capteurs BH3 (poitrine) et E4 (poignet), sur l'ensemble des fenêtres de temps sélectionnées en Ligne 1. En l'absence de sélection, le niveau maximal d'activité mesuré est de 382mG pour BH3 et 825mG pour E4 ($1mG = 10^{-3}G$).

Les deux indices de qualité sur le signal IBI_{bh3} (resp. : C_{bh3} et \mathcal{X}_{bh3}) ont chacun permis de retenir un peu plus la moitié des segments de 60s contenus dans la base de données (resp. : 51,4% et 52,4%). Les deux critères proposés dans les équations (7-8) et (7-11) ont relativement bien convergé : 90,5% des segments sélectionnés sur la base de l'indice \mathcal{X}_{bh3} ont également été sélectionnés sur la base de l'indice C_{bh3} . Ainsi, le « niveau de confiance » proposé par BioHarness pour le signal ECG valide à sa manière notre « indice de carence » comme indice de qualité sur le signal IBI issu de l'ECG.

De son côté, l'indice de carence $\mathcal{X}_{p\&t}$ a sélectionné un peu moins de la moitié (47,6%) des segments sur le signal $IBI_{p\&t}$ issu de l'algorithme de Pan & Tompkins (1985). Par comparaison avec cette référence académique, l'algorithme de BioHarness est donc un peu plus efficace pour estimer un rythme cardiaque à partir du signal ECG. Comme cet algorithme obtient de meilleures performances, le signal IBI_{bh3} a été conservé pour la section IV.B et le signal $IBI_{p\&t}$ a été retiré de la suite des analyses.

A côté de ces trois résultats similaires pour le capteur BH3, le Tableau 7.1 montre que la part de segments sélectionnés par notre critère sur le signal IBI_{e4} est très faible (0,6%). Un regard sur la quantité de mouvement maximale observée (sur les segments complets) pour chaque capteur montre exactement le même effet. Le niveau d'activité « que l'on peut observer » sur la poitrine lorsque la détection des battements est sans défaut est de 137mG pour les trois indices du capteur BH3 (ligne 2) ; et de 7mG pour le capteur E4, le niveau maximal enregistré sur la base de données étant de 382mG. Le rapport 20 entre ces deux chiffres est conservé pour le niveau d'activité « que l'on peut observer » sur le poignet (ligne 3 : 415mG lorsque les segments sont sélectionnés par BH3 contre 19mG lorsque les segments sont sélectionnés par E4, le niveau maximum enregistré étant de 825mG). Tous les éléments du Tableau 7.1 confirment ainsi l'incapacité du capteur E4 à fournir des segments *complets* de signal IBI en condition ambulatoire.

IV.B. Fiabilité de la Mesure PPG

A partir du signal IBI_{bh3} , 3370 segments complets d'une minute ont ainsi pu être définis comme fenêtres de référence pour mesurer la qualité du signal IBI_{e4} . Les paragraphes qui suivent donnent des statistiques sur les indicateurs de qualités calculés avec cette nouvelle référence.

IV.B.1. Qualité générale du signal IBI

Sur les 3370 fenêtres de référence, on a calculé trois indicateurs permettant de comparer le signal IBI_{e4} au signal IBI_{bh3} : la déviation absolue moyenne (DAM), le nombre d'échantillons surnuméraires (N_s) et la part d'échantillons manquants (PEM, voir section III.B et équation (7-9)). Les résultats obtenus pour chacun des 3 indicateurs sont résumés dans les histogrammes de la Figure 7.11.

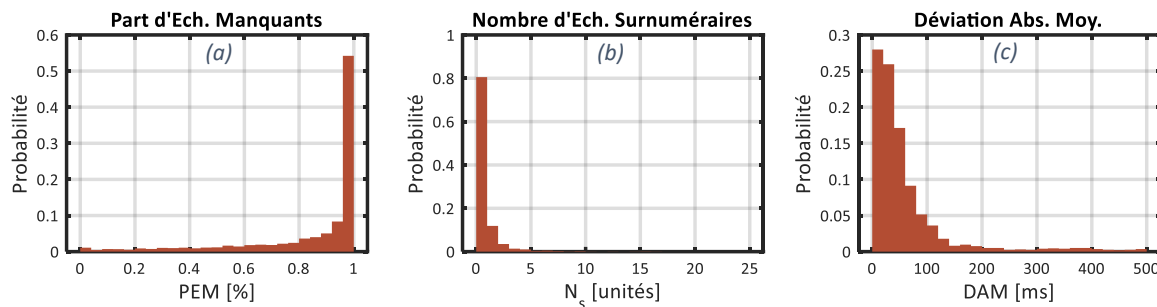


Figure 7.11 : Qualité générale du signal IBI_{e4}

- (a) Part d'échantillons manquants dans le signal IBI_{e4} par rapport au signal IBI_{bh3} .
 (b) Nombre d'échantillons manquants par fenêtre de temps dans le signal IBI_{e4} par rapport au signal IBI_{bh3} .
 (c) Déviation absolue moyenne entre les échantillons IBI_{e4} et les échantillons IBI_{bh3} .
 Statistiques obtenues sur les 3370 fenêtres de 60s pour lesquelles le IBI_{bh3} peut servir de référence.

Sur le graphique (a), on observe que la plupart des segments du signal IBI_{e4} ont une part d'échantillons manquants supérieure à 80%. Ceci montre qu'une part significative de l'information est perdue dans le signal IBI_{e4} par rapport au signal IBI_{bh3} , pour une grande majorité des fenêtres de temps. En moyenne, la PEM est de 85%. Au total, une fenêtre sur dix possède moins de 50% d'échantillons manquants ; 4 fenêtres sur 10 ne contiennent aucun échantillon pour le IBI_{e4} (PEM = 100%).

Sur le graphique (b), on observe que 8 segments sur 10 ne contiennent aucun échantillon surnuméraire (parmi les segments du signal IBI_{e4} qui contiennent des échantillons). Ceci montre que l'algorithme d'Empatica E4 est très spécifique lors de la détection des battements cardiaques. La surdétection de battements est un phénomène marginal pour ce capteur : on peut donc écarter cet indicateur de qualité pour la suite de notre analyse.

Sur le graphique (c), on constate que 90% des segments affichent une déviation absolue moyenne inférieure à 130ms, lorsqu'un échantillon IBI_{e4} correspond à un échantillon IBI_{bh3} . L'erreur observée sur le signal IBI_{e4} oscille typiquement entre 1% et 15% de la valeur standard d'un intervalle inter-battement (1s). La moyenne de cette erreur (DAM) est de 67ms, soit environ 7% de la valeur standard d'un IBI. Cette erreur n'est pas négligeable pour le calcul de la variabilité cardiaque

Pour finir, le coefficient de corrélation (Pearson, Spearman) entre la part d'échantillons manquants (PEM) et l'indice de carence (\mathcal{X}_{e4}) est de 0,999 sur les fenêtres de référence. Les deux indicateurs dévient en moyenne de 0,6%. Ces résultats valident l'indice de carence \mathcal{X} comme estimateur de la part d'échantillons manquants sur le signal IBI issu de la PPG. Par ailleurs, l'erreur précédente (0,6%) est inférieure à la part d'un IBI ($\sim 1s$) sur une fenêtre de temps (60s). Ainsi, nos résultats valident le critère de l'équation (7-8) pour déterminer les fenêtres sur lesquelles *aucun* échantillon IBI n'est manquant.

IV.B.2. Influence de la quantité de mouvement

Ces résultats montrent que l'algorithme d'Empatica E4 fournit une estimation relativement précise, mais rare, de l'intervalle inter-battements en condition ambulatoire. Dans quelle mesure peut-

on lier la qualité de cette estimation à la quantité de mouvement mesurée sur le poignet ? Pour répondre à cette question, l'indicateur Act_{e4} a été calculé en même temps que la part d'échantillons manquants (PEM) et la déviation absolue moyenne (DAM) sur chaque fenêtre de temps.

Ainsi, la Figure 7.12 affiche l'évolution des paramètres PEM et DAM en fonction de l'activité mesurée au niveau du poignet. Pour obtenir ces résultats, on a d'abord divisé en déciles (i.e. en tranches de 10%) les valeurs Act_{e4} mesurées sur les fenêtres de référence. Ces déciles ont défini l'échelle d'activité affichée en abscisse sur chaque graphique de la Figure 7.12. Les valeurs affichées en ordonnée correspondent à des statistiques sur les indicateurs PEM et DAM (médiane \pm intervalle interquartile), calculées indépendamment pour chacun de ces déciles d'activité.

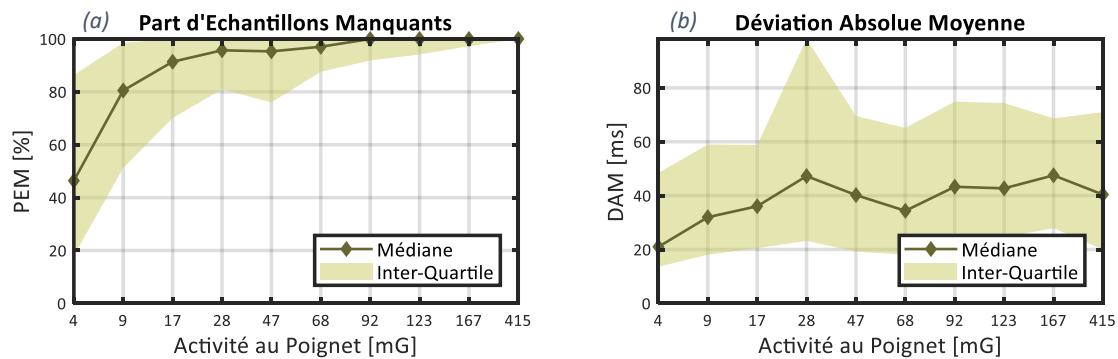


Figure 7.12 : Qualité du signal IBI_{e4} en fonction de la quantité de mouvement

(a) Part d'échantillons manquants dans le signal IBI_{e4} par rapport au signal IBI_{bh3} .

(b) Déviation absolue moyenne entre les échantillons IBI_{e4} et les échantillons IBI_{bh3} .

Chaque valeur d'activité correspond à un décile de l'indicateur Act_{e4} .

Ligne pleine : médiane de l'indicateur. Zone orangée : intervalle inter-quartile de l'indicateur.

Statistiques obtenues sur les 3370 fenêtres de 60s pour lesquelles le signal IBI_{bh3} peut servir de référence.

Sur le graphique (a), on observe une croissance nette de la part d'échantillons manquants (de 46% à 96% pour la médiane) sur les quatre premiers déciles du niveau d'activité. Cette part s'établit à presque 100% au-delà du cinquième décile ($Act_{e4}=47\text{mg}$, niveau d'activité faible comparé au maximum observable de 415mG). La déviation moyenne (graphique (b)) poursuit la même logique : sa valeur médiane augmente sur les quatre premiers déciles d'activité (21 à 47ms) puis stagne autour des 40ms. Cependant, cette augmentation reste noyée dans l'intervalle interquartile. La valeur médiane, limitée à 40ms pour les plus forts niveaux d'activité, montre que ce facteur influence finalement assez peu la précision de la mesure IBI par l'algorithme d'Empatica.

On peut déduire de ces résultats que si l'activité diminue fortement la *quantité* d'information présente dans le signal IBI_{e4} , elle n'influence pas la *qualité* de cette information. L'intervalle inter-battement est estimé avec une fiabilité relativement constante même aux plus forts niveaux d'activité.

IV.B.3. Qualité des caractéristiques cardiaques

Après avoir montré la rareté des échantillons disponibles dans le signal IBI_{e4} , on peut s'interroger sur la qualité des caractéristiques extraites à partir d'un tel signal. Quelle erreur peut être attendue sur une fenêtre de temps quelconque ? Peut-on diminuer cette erreur en sélectionnant seulement les fenêtres où l'on constate une faible activité au poignet ? Quelle utilité peut avoir l'indice de carence pour prédire les fenêtres où le calcul des caractéristiques sera valide ?

Pour répondre à ces questions, on a calculé l'erreur \mathcal{E}_f définie en section III.B (équation (7-10)) pour les caractéristiques $hrMean$, $hrStd$ et $hrRmssd$, sur l'ensemble des fenêtres où le signal IBI_{bh3}

pouvait servir de référence. Puis on a sélectionné ces fenêtres selon trois critères qui devraient permettre une sélection des niveaux d'erreur les plus faibles.

- i) Les fenêtres qui observent une « faible » activité du poignet. Il s'agit du premier décile de l'indicateur Act_{e4} sur les fenêtres de référence.
- ii) Les fenêtres qui observent un « faible » indice de carence. Il s'agit du premier décile de l'indice \mathcal{X}_{e4} sur les fenêtres de référence.
- iii) Les fenêtres complètes du signal IBI_{e4} . C'est l'ensemble \mathcal{C}_{e4} des fenêtres qui valident le critère d'intégrité : $\mathcal{X}_{e4} < \min(IBI_{e4}) / L$.

Pour comparer directement la probabilité d'erreur dans ces quatre ensembles, la Figure 7.13 présente la fonction de répartition des valeurs \mathcal{E}_f pour chaque caractéristique f . Cette fonction

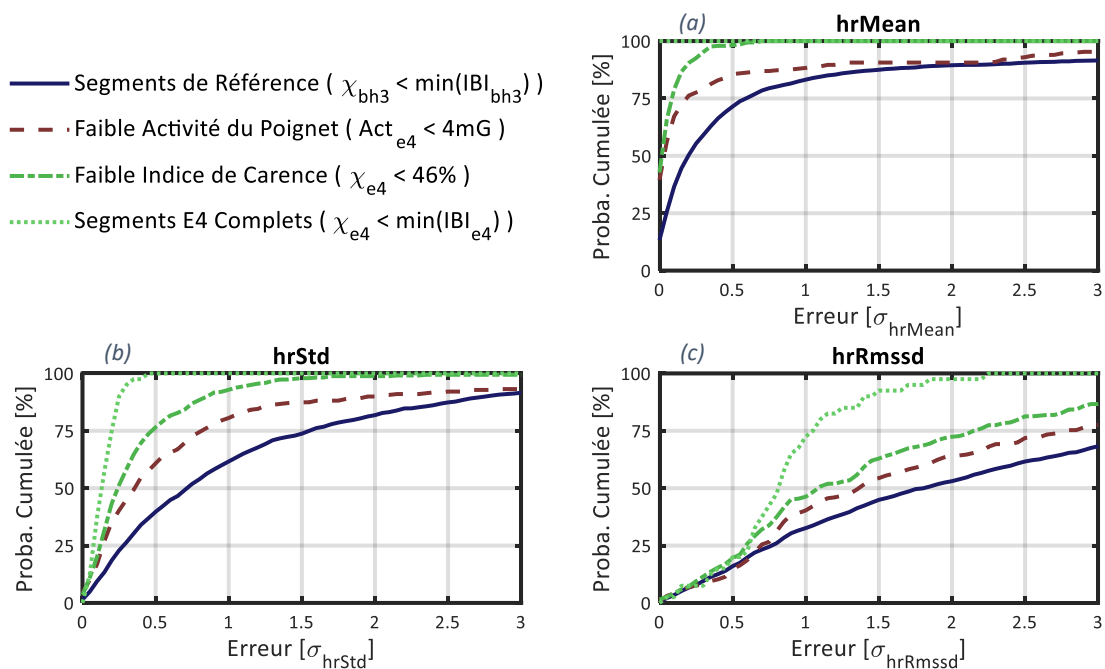


Figure 7.13 : Erreur sur les trois caractéristiques cardiaques calculées à partir du signal IBI_{e4}

hrMean : rythme cardiaque moyen ; *hrStd* : écart-type du rythme cardiaque ; *hrRmssd* : RMS des différences successives du rythme cardiaque. Calculs sur les 3370 fenêtres de 60s pour lesquelles le signal IBI_{bh3} pouvait servir de référence. Chaque courbe représente la probabilité cumulée d'erreur sur une portion de cet ensemble. La ligne pleine rassemble tous les segments. Les lignes discontinues correspondent aux trois critères de qualité développés dans le corps du texte.

correspond à la probabilité P_f que l'erreur sur la caractéristique soit inférieure ou égale à une certaine valeur, c'est-à-dire la proportion des fenêtres de temps qui présente un certain niveau maximal d'erreur. L'erreur médiane est donnée par le point d'ordonnées $P_f = 50\%$. Plus la courbe s'approche du point : $\{ \mathcal{E}_f = 0 ; P_f = 100\% \}$, plus la valeur des erreurs mesurées dans cet ensemble est faible pour un grand nombre de segments IBI_{e4} .

Lorsqu'on calcule la moyenne du rythme cardiaque (*hrMean*) sur tous les segments pour lesquels une référence est disponible (graphique (a), ligne bleu marine), on constate que l'erreur typiquement mesurée est faible : l'erreur médiane s'établit à $0,2\sigma$. Cela dit, on observe un comportement asymptotique de la courbe bleue dans les grandes valeurs de \mathcal{E}_{hrMean} : une poignée de fenêtres (env. 10%) conserve une valeur d'erreur élevée. Lorsqu'on ne sélectionne que les fenêtres avec un niveau d'activité « faible » (ligne pourpre en tirets), on diminue largement l'erreur médiane sur *hrMean* ($< 0,05\sigma$). A nouveau, une petite partie des fenêtres (environ 5%) conserve une valeur d'erreur élevée.

En revanche, si l'on ne sélectionne que les fenêtres ayant un indice de carence « faible » (inférieur à 46% pour le signal IBI_{e4}), l'erreur observée ne dépasse jamais les 0.5σ (ligne verte en tirets-pointillés). Enfin, 100% des fenêtres complètes pour le signal IBI_{e4} montrent une erreur inférieure à $0,05\sigma$ pour le calcul de $hrMean$ (ligne vert pâle en pointillés).

Losqu'on la calcule sur tous les segments (graphique (b), ligne bleu marine), l'erreur typiquement mesurée est plus élevée pour l'écart-type du rythme cardiaque ($hrStd$) : la valeur médiane est de $0,7\sigma$. A nouveau, la sélection de fenêtres sur critère *d'activité* permet de diminuer ce taux d'erreur de moitié (ligne pourpre en tirets). Cependant, seul un *indice de carence* faible garantit une erreur faible sur le calcul de cette caractéristique (ligne verte en tirets-pointillés). Enfin, 100% des fenêtres complètes pour le signal IBI_{e4} obtiennent une erreur inférieure à $0,5\sigma$ (ligne vert pâle en pointillés).

Enfin, le taux d'erreur est bien plus fort pour la RMS des différences successives ($hrRmssd$: valeur médiane à $1,8*\sigma$ sur tous les segments). Contrairement aux deux autres caractéristiques, aucun de nos critères ne permet d'augmenter la part de fenêtres sur lesquelles l'erreur est faible ($<0,5\sigma$). Bien qu'on observe une amélioration significative parmi les segments IBI_{e4} complets, un quart d'entre eux conserve une erreur comprise entre 1σ et 2σ .

On peut résumer ainsi les résultats observés dans la Figure 7.13. Le manque d'échantillons disponibles explique totalement l'erreur observée sur le rythme cardiaque moyen ($hrMean$) pour le signal IBI_{e4} . Cela dit, cette erreur demeure faible pour la majorité des segments analysés, quel que soit la part d'échantillons manquants. Les résultats du graphique (a) semblent ainsi confirmer l'hypothèse formulée au paragraphe II.A.2 pour construire l'indice de carence, à savoir que la moyenne des échantillons IBI sur une fenêtre de temps dépend peu du nombre d'IBI manquants. Par ailleurs, l'indice \mathcal{X} permet de proposer des critères efficaces pour prédire les fenêtres de temps sur lesquelles le calcul des caractéristiques $hrMean$ et $hrStd$ peut être considéré comme valide. Cependant, le manque d'échantillons disponibles n'explique que très partiellement l'erreur observée sur la variabilité haute fréquence du rythme cardiaque ($hrRmssd$) calculée à partir du signal PPG. Sans signal de référence, il semble difficile de proposer un critère permettant de déterminer si le calcul de cette caractéristique correspond effectivement à de la variabilité cardiaque.

V. Discussion

V.A. Sur l'Observation de l'Activité Cardiaque en Vie réelle

V.A.1. Disponibilité de la mesure IBI

L'algorithme de détection des battements cardiaques à partir du signal PPG d'Empatica semble efficace. Dans nos enregistrements en vie réelle, la surdétection des battements cardiaques à partir du signal PPG est restée marginale sur les fenêtres de temps pour lesquelles l'ECG pouvait servir de référence. Les battements ont été repérés de manière suffisamment précise pour que l'erreur mesurée sur l'IBI dépasse rarement le dixième de seconde, quel que soit le niveau d'activité physique (Figure 7.12). Ainsi, la sélection des ondes de pouls conformes par l'algorithme d'Empatica a permis de produire un signal IBI suffisamment fiable pour que l'utilisation de notre indice de carence ait du sens (voir justifications plus bas). Cependant, cette sélection des ondes de pouls se fait au prix d'un signal IBI particulièrement rare en condition ambulatoire.

En effet, sur 90% des fenêtres de temps pour lesquelles l'ECG a pu servir de référence, il manquait au signal IBI_{e4} plus d'un échantillon sur deux. Nos résultats contrastent ainsi nettement avec ceux d'Orphanidou et al. (2015), qui ont compté au maximum 1 battement manquant sur la moitié de leurs segments IBI de 10 secondes. Bien que le protocole impliquait des conditions semi-ambulatoires, ces derniers résultats concernent un signal PPG mesuré en milieu hospitalier à l'aide d'une instrumentation standard. Nos résultats montrent ainsi l'impact considérable de la mesure en ambulatoire *via* des capteurs portables sur la quantité d'information présente dans le signal IBI issu de la PPG. Cette perte d'information dans le signal IBI est largement déterminée par la quantité de mouvement mesurée sur le poignet, comme on a pu le constater dans la Figure 7.12.

En ce qui concerne le signal IBI issu de la mesure ECG, les résultats du Tableau 7.1 sont proches de ceux d'Orphanidou et al. (2015), pour qui la moitié des segments de 10s enregistrés par un dispositif Holter a montré au maximum 1 échantillon IBI manquant. Ces résultats ont notamment permis à l'algorithme de détection de battements de BioHarness de se positionner comme référence ; et donnent un certain crédit à l'observation de l'activité cardiaque en vie réelle au moyen de l'ECG thoracique. Même pour BioHarness, cela dit, la présence d'échantillons IBI manqués sur la moitié des fenêtres de temps montre qu'il est pertinent de coupler l'estimation du rythme cardiaque à un indice temps-réel qui permet de repérer les situations où de l'information est manquante dans le signal. En effet, les résultats de la section IV.B montrent que la quantité d'échantillons IBI manquants impacte le calcul des caractéristiques cardiaques sur une fenêtre de temps.

V.A.2. Qualité des caractéristiques cardiaques issues de la mesure PPG

Nos résultats sur la qualité des caractéristiques cardiaques concernent seulement la mesure PPG, car nous ne disposons pas d'une mesure ECG supplémentaire pour valider les signaux de la ceinture BioHarness. Néanmoins, ils nous permettent de tirer des conclusions intéressantes sur l'observation temps-réel de l'activité cardiaque au moyen de cette technologie (PPG) qui est encore très largement dominante sur le marché des capteurs physiologiques portables.

D'après la Figure 7.12, la quantité de mouvement dans le poignet détermine la *quantité* d'information contenue dans le signal IBI_{e4} (proportion d'échantillons manquants) mais pas la *précision* de cette information (déviations absolues moyennes). Quel que soit le degré d'activité, le signal IBI_{e4} présente une erreur constante dans l'estimation du rythme cardiaque. On peut trouver deux explications à ce phénomène : (i) la forme d'onde PPG, qui ne permet pas de marquer chaque battement de cœur avec autant de précision que sur l'ECG ; et (ii) le temps de transit de l'onde PPG entre le cœur et le poignet (voir section I.A).

Cette erreur d'estimation est presque sans conséquence pour évaluer le rythme cardiaque moyen sur une période prolongée (erreur médiane faible pour le calcul de $hrMean$ sur la Figure 7.13). Ainsi, la mesure PPG introduit une variance *sans biais* dans le signal de rythme cardiaque. Ceci explique le succès actuel de la mesure PPG pour des applications bas-niveau, comme l'estimation de l'activité physique ou les rendus graphiques de type *life log*. En revanche, les conséquences de cette (petite) erreur d'estimation de l'IBI sur le signal PPG sont désastreuses pour la mesure de variabilité cardiaque haute fréquence (calcul de la caractéristique $hrRmssd$ sur la Figure 7.13 : erreur médiane forte, même lorsque tous les échantillons IBI sont disponibles). Ainsi, nos résultats donnent du crédit à l'opposition parfois faite entre la notion de *variabilité cardiaque* et celle de *variabilité du pouls*.

Si l'on réinterprète les résultats de la Figure 7.13 (graphique (c)) concernant les segments où l'activité physique était faible, on observe qu'environ 10% des mesures de la caractéristique *hrRmssd* ont une erreur en-dessous de $\frac{1}{2}$ écart-type ; et 25% ont une erreur au-dessus de 3 écarts-types. Ces probabilités d'erreur sont du même ordre que celles affichées dans le Tableau 6.3 (Chapitre Six - III.D page 119), lorsqu'on introduisait des mesures aléatoires parmi nos caractéristiques mesurées en laboratoire. Dans ces conditions, l'indice de stress « entier » s'était révélé plus efficace que le réseau de neurones \mathcal{N}^1 lorsque cette même caractéristique *hrRmssd* était bruitée. Ainsi, les résultats du présent chapitre ont montré l'intérêt de l'indice de stress en tant que modèle résilient à des données bruitées. En couplant l'indice de stress \mathcal{M} avec l'indice de carence \mathcal{X} , on peut enfin concrétiser son utilisation en tant que détecteur à caractère *modulaire* – c'est-à-dire adapté à un ensemble variable de caractéristiques cardiaques.

V.B. Sur l'Indice de Carence

V.B.1. Limites

Le calcul de l'indice de carence \mathcal{X} suppose que les échantillons IBI disponibles dans la fenêtre de temps sont des mesures *valides* du rythme cardiaque. Cela revient à formuler deux hypothèses.

- a) L'algorithme de détection des battements sur la mesure initiale (ECG, PPG) est efficace.
- b) L'algorithme de détection des valeurs aberrantes sur le signal IBI est efficace.

En effet, l'indice \mathcal{X} ne peut être utilisé sans un traitement préalable qui garantit que le signal IBI a bien toutes les propriétés d'un rythme cardiaque (et nettoie le signal de toute mesure invalide). La procédure implémentée en section II.B est un exemple perfectible de ce type de traitement. Tel qu'on l'a proposé, en effet, cet algorithme ne peut pas repérer d'échantillons invalides lorsque la variance du signal IBI est trop élevée. Cela pourrait se produire (par exemple) sur un signal PPG de mauvaise qualité, d'où l'importance que l'algorithme de détection des battements de coeur soit efficace. En particulier, la production de faux positifs (battements surdétectés) doit être rare ; car notre algorithme de rejet des valeurs aberrantes est incapable de repérer une succession trop longue de ces mesures cardiaques invalides.

V.B.2. Avantages

Cela étant dit, les résultats de la section IV ont permis de montrer les deux propriétés suivantes.

D'une part, l'indice \mathcal{X} est un estimateur valide de la quantité d'échantillons manquants dans le signal IBI sur une fenêtre de temps. Ainsi lorsque l'algorithme de détection des battements cardiaques produit peu de faux positifs (c'est le cas pour le bracelet E4 et pour la détection de battements sur le signal ECG), le couplage de l'indice \mathcal{X} avec un algorithme de retrait des valeurs aberrantes en fait un estimateur valide de la qualité du *signal* de rythme cardiaque sur la fenêtre de temps.

D'autre part, l'indice \mathcal{X} est un prédicteur efficace de la qualité d'une *caractéristique* cardiaque sur une fenêtre de temps. En effet, on peut obtenir une mesure exacte du rythme cardiaque moyen et une mesure précise de son écart-type à partir du signal PPG lorsque suffisamment d'échantillons sont disponibles sur la fenêtre de temps. La quantité d'échantillons disponibles augmente également la précision des mesures de la variabilité cardiaque. Ainsi, l'indice de carence \mathcal{X} peut aider à déterminer

si une caractéristique cardiaque peut être calculée avec suffisamment de précision pour l'application souhaitée. Par exemple, pour la détection du stress en condition ambulatoire.

V.B.3. Indice de carence dans un indice de stress

Couplé avec notre procédé de retrait des valeurs aberrantes, l'indice de carence (\mathcal{X}) peut être estimé en temps réel avec un besoin très faible en ressources de calcul. A partir de la valeur de \mathcal{X} sur une fenêtre de temps, on peut donc appliquer en temps réel un critère pour indiquer si le calcul d'une caractéristique f est compatible avec une bonne détection de l'état de stress.

Il suffit, par exemple, de proposer un seuil de tolérance x tel la caractéristique f ne serait incluse dans l'indice de stress que lorsque $\mathcal{X} \leq x$. Sur la Figure 7.13, les deux courbes qui dépendent de l'indice \mathcal{X} correspondent à une sélection des segments IBI sur la base des critères : $x = 46\%$ (indice « faible ») et $x \approx 0\%$ (segments complets). L'utilisation de ces courbes peut nous permettre d'ajuster x à la caractéristique $f \in \{hrMean, hrStd, hrRmssd\}$, dans le cas où celle-ci serait calculée à partir du capteur Empatica E4.

Au Chapitre Six -, la version « entière » de l'indice de stress s'est montrée robuste à des données très bruitées. Si un critère suffisamment rigoureux (ex. : $x \approx 0\%$) est imposé pour le calcul de la caractéristique $hrRmssd$ à partir du signal IBI_{e4} , on peut espérer une détection sporadique (mais juste) de l'état de stress à partir des trois caractéristiques proposées dans ce chapitre ($hrMean$, $hrStd$, $hrRmssd$). Lorsque ce critère de tolérance ($\mathcal{X} \approx 0\%$) n'est pas satisfait, il suffirait de retirer la caractéristique $hrRmssd$ du calcul de l'indice de stress, tant que l'indice \mathcal{X} respecte le critère de la seconde caractéristique la plus bruitée du signal IBI_{e4} (par exemple : $\mathcal{X} < 10\%$ pour $hrStd$).

Avec cette stratégie pour l'estimation du stress en vie réelle, on peut se poser la question suivante : sur une journée d'enregistrement quelconque, quelle est la probabilité pour que l'indice de carence valide un certain seuil x ? Une réponse détaillée est proposée en Annexe 12, où le critère proposé ici est concrètement appliqué aux segments IBI de notre base de données. Pour la mesure PPG comme pour la mesure ECG, cette probabilité dépend du seuil de tolérance mais aussi de la journée d'enregistrement et de la taille des fenêtres de temps. Pour la ceinture BH3, elles sont relativement fréquentes aux seuils les plus bas (environ 7/10 pour $x=5\%$), ce qui confirme l'ECG thoracique comme une mesure à privilégier pour l'observation de l'activité cardiaque dans la vie quotidienne. Pour la mesure PPG, les fenêtres sélectionnées sont encore rares à des seuils élevés (moins de 1/10 pour $x=50\%$), à l'image des fortes proportions d'échantillons manquants relevés en section IV.B. En attendant une amélioration substantielle de la technologie PPG, ces résultats compromettent un suivi précis de la variabilité cardiaque au moyen d'un bracelet connecté.

V.B.4. Synthèse

La qualité du signal de rythme cardiaque peut être estimée en temps réel au moyen d'algorithmes simples (ici, un retrait des valeurs aberrantes couplé à une estimation de la proportion d'échantillons manquants) et dont la sortie (ici, un indice de carence) est facilement interprétable. Cette conclusion ouvre la voie à une application de l'indice de stress sur des mesures de la vie quotidienne. Cependant, l'absence d'échantillon manquant ne suffit pas à garantir une estimation précise de certaines caractéristiques sensibles au bruit, comme la variabilité cardiaque haute fréquence, pour la mesure PPG (actuellement dominante sur le marché des capteurs cardiaques portables). De plus, le signal de

rythme cardiaque issu de la PPG est susceptible de rester indisponible sur de longues périodes (4 segments de 60s sur 10 ne contiennent aucun échantillon dans notre base de données). Des perspectives de recherche se trouvent dans l'amélioration de cette mesure, ou dans la mise en évidence d'un substitut fiable pour l'estimation du rythme cardiaque dans la vie quotidienne.

A RETENIR

- Notre « indice de carence » fournit une estimation précise de la proportion d'échantillons manquants sur un segment de signal du rythme cardiaque
- Couplé à un algorithme de retrait des valeurs aberrantes, cet indice prédit efficacement la fiabilité d'une caractéristique du rythme cardiaque. Il peut ainsi fournir un critère efficace sur la disponibilité d'une caractéristique cardiaque pour une utilisation dans « l'indice de stress ».

Chapitre Huit - Conclusions & Perspectives

I. Synthèse des Contributions

Nous avons cherché à développer un détecteur de stress dans une approche en *boîte-blanche*, qui prend en compte la qualité des mesures physiologiques et la présence de facteurs confondants. Dans cette démarche, les travaux de cette thèse nous ont permis de proposer les trois éléments suivants.

- i) Un ensemble de caractéristiques sensibles au stress et peu sensibles aux facteurs confondants. Cette sélection des caractéristiques a fait l'objet du Chapitre Cinq.
- ii) Un détecteur de stress robuste à des données bruitées, dont la structure lui permet de s'adapter à un ensemble variable de caractéristiques. La conception de ce modèle et la mesure de ses performances a fait l'objet du Chapitre Six.
- iii) Un algorithme permettant de détecter en temps réel si le calcul d'une caractéristique cardiaque est suffisamment fiable pour intégrer celle-ci à notre modèle de stress. La conception de cet algorithme a fait l'objet du Chapitre Sept, accompagnée d'une réflexion plus générale sur la fiabilité en vie réelle de la mesure PPG.

Dans cette démarche, les analyses se sont appuyées sur deux bases de données différentes. La base issue de mesures en laboratoire (Chapitres Cinq et Six), nous a permis de prendre en compte des facteurs tels que : *l'effet de la tâche* (les différences observées sur une caractéristique entre deux activités stressantes), *l'effet du sujet* (les différences observées entre plusieurs individus) et surtout *l'effet de l'activité* (la réponse réflexe de l'organisme à une activité non-stressante). La base de données recueillie dans le cadre de cette thèse (Chapitre Sept), qui comprend l'enregistrement de plusieurs individus dans la vie quotidienne, nous a permis d'estimer la qualité qui peut être attendue en vie réelle sur des mesures cardiaques issues de capteurs portables.

II. Principaux Résultats

II.A. Recherche de Caractéristiques Sensibles au Stress

En examinant la réponse au stress des trente-deux caractéristiques physiologiques calculées au Chapitre 4, on a montré qu'un tout petit nombre d'entre elles était capable de distinguer le stress émotionnel d'une activité cognitive modérée. Les quatre caractéristiques ainsi sélectionnées sont le rythme cardiaque moyen, l'écart-type du rythme cardiaque, la RMS des différences successives du rythme cardiaque et l'autocorrélation du signal respiratoire. Nos résultats montrent que ces caractéristiques répondent au stress de manière universelle (*i.e.*, varient pour tout le monde), unique (*i.e.*, varient dans différents contextes) et univoque (*i.e.*, varient toujours dans le même sens). Cependant, toutes ces caractéristiques se sont aussi montrées sensibles à l'effet d'activité.

Notre sélection des corrélats du stress nous a conduits à rejeter la plupart des caractéristiques fréquemment citées dans la littérature. A contre courant de nombreuses études sur la détection du stress mental, les résultats du Chapitre 5 montrent ainsi l'importance de sélectionner des caractéristiques en tenant compte de l'effet d'activité. En outre, la recherche d'une réponse physiologique cohérente entre deux tâches différentes nous a conduit à rejeter certaines

caractéristiques qui avaient déjà participé à distinguer le stress de l'activité mentale, mais sur la base d'un protocole monotâche. Ce résultat appuie l'idée qu'une caractéristique physiologique doit être observée dans différents contextes avant d'être identifiée comme corrélât du stress.

II.B. Proposition d'un Modèle Adaptatif : l'Indice de Stress

Les quatre caractéristiques précédentes ont servi de base à l'élaboration d'un détecteur de stress. Dans la littérature, un certain nombre de techniques d'apprentissage permet déjà la conception de modèles dans une approche *boîte-blanche*. Cependant, la structure figée de ces modèles les rend sensibles à la présence de bruit dans le calcul des caractéristiques. Notre objectif était de proposer un modèle *performant* sur nos mesures de laboratoire, mais également *robuste* lorsque la qualité des mesures vient à se dégrader. Pour cela, on s'est appuyé sur deux hypothèses :

- (a) En l'absence d'information sur la qualité des mesures, on peut augmenter la robustesse du détecteur de stress en limitant la contribution de chaque caractéristique sur la décision finale.
- (b) En présence d'information pertinente sur la qualité des mesures, on peut augmenter la robustesse du détecteur en retirant des calculs chaque caractéristique bruitée.

L'indice de Stress conçu dans cette démarche tire parti du pouvoir discriminant de chaque caractéristique sélectionnée dans le Chapitre Cinq. En appliquant une fonction saturante ajustée à chaque caractéristique, on limite de fait la contribution de chacune d'elles sur la sortie du modèle. En utilisant une relation simple (ici une moyenne) pour fusionner les caractéristiques, on permet au modèle de s'adapter sans apprentissage à un ensemble de caractéristiques variable.

Sur nos données de laboratoire, l'Indice de Stress a bien classé 75% des exemples Stress et Contrôle dans leurs catégories respectives. Cette performance est supérieure à celle d'une régression logistique à deux classes (72,4%) mais inférieure à celle d'un réseau de neurones à une unité dans la couche cachée (80,9%). Mais la performance sur des données « propres » n'est pas la seule propriété recherchée pour notre modèle, qui n'est pas conçu dans cet objectif. En introduisant volontairement du bruit dans nos caractéristiques physiologiques, le réseau de neurones a montré une forte chute de ses performances (-15% en moyenne) tandis que l'indice de stress montrait des performances plus stables (-8% en moyenne). Lorsque la caractéristique bruitée était retirée du calcul de l'indice de stress, la chute de performance était inférieure à 1% dans trois situations sur quatre.

II.C. Estimation de la Qualité d'un Signal de Rythme Cardiaque

Les résultats précédents ont montré la capacité de l'indice de stress à conserver ses performances lorsqu'une information pertinente est disponible au bon moment pour retirer une caractéristique qui présente un bruit trop important. Parmi les caractéristiques sélectionnées au Chapitre Cinq, trois étaient extraites du signal de rythme cardiaque. La mesure de la qualité du rythme cardiaque est un problème déjà abordé dans la littérature, mais la plupart des algorithmes existants s'appuient sur un apprentissage en boîte-noire et n'ont jamais été appliqués dans un contexte ambulatoire.

Pour combler cette lacune, nous avons mis au point une méthode boîte-blanche qui permet de mesurer la perte d'échantillons dans un signal de rythme cardiaque quelconque. L'Indice de Carence issu de cette méthode estime précisément la proportion d'échantillons manquants dans une fenêtre temporelle. A partir de cet indicateur, nous avons pu dégager un critère permettant de sélectionner les fenêtres temporelles sur lesquelles le signal de rythme cardiaque issu la mesure ECG d'une ceinture

thoracique (Zephyr BioHarness 3) ne présentait aucun échantillon manquant. Ce signal de référence nous a permis de mesurer la qualité du signal de rythme cardiaque issu de la mesure PPG du bracelet Empatica E4. En particulier, nous avons extrait de ces deux signaux les trois caractéristiques cardiaques sélectionnées au Chapitre Cinq, puis mesuré l'erreur sur les caractéristiques issues de la PPG.

Nos résultats ont montré que la mesure PPG fournissait une mesure précise de la moyenne du rythme cardiaque, une mesure peu précise de la dispersion du rythme cardiaque et une mesure très imprécise de la variabilité cardiaque haute fréquence. En sélectionnant les fenêtres temporelles sur la base de notre indice de carence, on a pu diminuer fortement le taux d'erreur mesuré pour chacune de ces caractéristiques. Cette diminution du taux d'erreur dépendait du seuil placé sur l'indice de carence. Si une tolérance Zéro sur l'indice de carence a presque supprimé l'erreur sur la moyenne et la dispersion du rythme cardiaque, la variabilité cardiaque haute fréquence cependant a conservé des valeurs d'erreur élevées. Ainsi, la mesure PPG semble générer des caractéristiques de variabilité cardiaque peu fiables pour la détection du stress en condition ambulatoire.

Couplé à un algorithme de retrait des valeurs aberrantes, l'indice de carence s'est affiché comme un prédicteur efficace de la qualité des caractéristiques cardiaques en condition ambulatoire. Couplé à notre indice de stress, il nous permet ainsi de proposer un modèle flexible et réactif à la qualité des mesures de rythme cardiaque.

II.D. Synthèse des Apports Méthodologiques

Au-delà d'un algorithme de détection du stress, ce manuscrit a également présenté des *méthodes* à travers lesquelles on a cherché à dépasser les discordances observées dans l'état de l'art entre les protocoles de modélisation du stress mental. Ces méthodes répondent à une contrainte commune : la transparence des processus et des calculs, afin que la validité de notre modèle de stress puisse être vérifiée sur de nouvelles bases de données.

On a développé, par exemple, une nouvelle méthode de sélection des caractéristiques : la sélection par estimateurs. Couplée à des critères statistiques sur la présence et l'absence d'un effet, cette méthode permet de sélectionner des caractéristiques *sensibles* au stress et *insensibles* aux facteurs confondants. Elle peut être adaptée à l'évaluation de n'importe quel concept cible, pourvu que l'on dispose d'une base de données permettant le contrôle de facteurs confondants.

On a appliqué, de même, un nouveau protocole de validation croisée : le processus *LOSOTO*. Celui-ci permet de mesurer la capacité d'un modèle à généraliser simultanément sur une nouvelle tâche et un nouveau sujet. A nouveau, l'application de ce processus nécessite une base de données multi-tâche et multi-sujet. L'inconvénient de ce processus est qu'il réduit fortement la quantité de données disponibles pour la validation du modèle en question.

On a proposé, enfin, un critère efficace pour sélectionner les segments de signal de rythme cardiaque qui ne présentent aucun défaut en condition ambulatoire. En plus de prédire les fenêtres sur lesquelles l'extraction de caractéristiques sera optimale, ce critère peut être utilisé dans la même démarche que celle proposée au Chapitre Sept : la détermination d'un signal de référence pour mesurer la qualité d'un autre signal de rythme cardiaque.

III. Perspectives de Recherche

Des perspectives de recherche immédiates se trouvent dans la recherche d'un seuil optimal sur l'indice de carence, qui permettrait un retrait efficace de chaque caractéristique lorsque l'information sur le rythme cardiaque se fait trop rare pour détecter correctement l'état de stress. Ce travail de recherche nécessite de finir le dépouillement de la base de données Semaine de Travail, qui contient 41 jours d'enregistrement obtenus sur 6 sujets différents. Chaque sujet a porté un à deux bracelets Empatica E4 et une ceinture BioHarness. La base de données comporte également deux vérités de terrain différentes sur le niveau de stress : des rapports verbaux en temps réel à l'initiative des participants, et une mesure rétrospective effectuée chaque soir.

Mais le potentiel de l'indice de stress ne se limite pas à la conservation de ses performances sur des données réelles. Par exemple, un atout majeur de ce détecteur est de pouvoir s'adapter à un ensemble de caractéristiques plus riche que le quatuor sélectionné dans ce manuscrit. Ce modèle pourrait ainsi intégrer de nouvelles caractéristiques sélectionnées à partir de nouvelles tâches expérimentales. Par ailleurs, le Chapitre 5 a mis en évidence des perspectives essentielles dans la recherche de nouvelles caractéristiques sensibles au stress émotionnel.

L'indice de stress est une mesure à plusieurs composantes : il indique l'état de stress du « point de vue » de chaque caractéristique. Sur la base de données Semaine de Travail, une analyse des valeurs de chaque indice simple pourrait permettre d'identifier des schémas de réponse propres à des situations particulières, plutôt que de rechercher une réponse générique au stress comme on l'a fait dans ce manuscrit. En effet, l'existence d'un stress physiologique polymorphe est une question toujours ouverte à laquelle ces travaux de thèse n'ont pas tenté de répondre.

Une observation plus subtile peut être faite sur la prévalence des facteurs confondants dans la vie quotidienne. En effet, certaines caractéristiques sont plus sensibles que d'autres à certains facteurs – voire même utiles seulement en présence de certains d'entre eux. Par exemple, l'autocorrélation du signal respiratoire pourrait n'être utile qu'en présence de parole. Le rythme cardiaque moyen, sur lequel nos modèles de stress se sont excessivement appuyés, n'est probablement utile qu'en l'absence totale d'effort physique. Lors de la conception d'un détecteur de stress, n'importe quel algorithme standard favoriserait une caractéristique ou l'autre en fonction des facteurs confondants présents dans la base de données – que ce soit en laboratoire ou en vie réelle. Or, il est impossible de prévoir la fréquence de ces facteurs dans la vie quotidienne de l'utilisateur final.

L'indice de stress, en revanche, pourrait facilement s'adapter aux facteurs confondants qu'il rencontre – pourvu qu'une information pertinente soit disponible pour les détecter. C'était la stratégie imaginée par Gjoreski et collaborateurs (2017, 2016), qui avaient ajouté des informations d'activité et de contexte à leur modèle physiologique élaboré en conditions contrôlées. De notre point de vue, cependant, cette stratégie était implémentée de manière imprécise : la sortie du détecteur physiologique était utilisée comme caractéristique en même temps que les données de contexte, sans en modifier le verdict – et la validité de ce verdict était impossible à vérifier. Un objectif à terme pour l'indice de stress serait de pouvoir modifier ce verdict en retirant simplement les caractéristiques déterminées comme inutiles, lorsqu'un facteur confondant est détecté par certaines données de contexte. Ces données pourraient être fournies, par exemple, sur la base d'un estimateur d'activité.

Annexes

Annexe 1. Stress Physique ou Stress Mental : l'Exemple de l'Activité Cardiovasculaire

I. Réponse Cardiaque à l'Effort Physique

Lorsque la demande métabolique s'emballe, le cortisol et l'adrénaline, en particulier, sont sécrétés pour mobiliser les ressources énergétiques de l'organisme ; à court terme, si la consommation de l'organisme en dioxygène dépasse un certain seuil ; ou à plus long terme, si l'exercice physique dure plusieurs dizaines de minutes (Hackney, 2006). Mais la réponse hormonale est relativement lente, et ces effets de seuil la rendent inefficace pour garantir l'homéostasie sur le court terme. Dès les premiers signes d'une demande croissante, certains automatismes du muscle cardiaque (ex : mécanisme de Frank-Starling) lui permettent d'augmenter sans délai son propre rendement (Martins-Pinge, 2011). Parallèlement, des récepteurs mécaniques et chimiques au sein du système vasculaire permettent l'établissement d'un contrôle global et réflexe de la pression sanguine par le système nerveux central (baroréflexe, chémoréflexe), médiatisé par le système nerveux autonome (Martins-Pinge, 2011; Taylor et al., 1999). Le rendement cardiaque (i.e. le débit de sang pompé par le cœur) est également modifié pour faire face à une augmentation de la demande en dioxygène.

En situation d'effort, celui-ci est proportionnel à la demande métabolique (mesurée en débit d'oxygène consommé) selon la formule de Fick (Butler et al., 2004) :

$$\dot{V}_{O_2} = HR * V_S * ([O_2]_a - [O_2]_v) \quad (0-1)$$

Ici, **HR** représente le rythme cardiaque (anglais : *heart rate*) ; **V_S** représente le volume d'éjection systolique (anglais : *stroke volume*), soit le volume éjecté par un battement du cœur ; si bien que **HR*V_S** mesure le rendement cardiaque. \dot{V}_{O_2} représente le débit d'oxygène consommé, et $([O_2]_a - [O_2]_v)$ le différentiel de concentration en dioxygène entre le sang artériel (**a**) et le sang veineux (**v**).

L'action du nerf parasympathique (nerf vague) est connue pour diminuer le rythme cardiaque, et celle du nerf sympathique pour l'augmenter. Cette effet chronotrope sur le muscle cardiaque est dû à l'action de deux neurotransmetteurs : l'acétylcholine pour les fibres nerveuses parasympathiques, et l'adrénaline pour les fibres nerveuses sympathiques (Taylor et al., 1999). L'action du nerf parasympathique est plus rapide que celle du nerf sympathique; si bien qu'une augmentation immédiate du rythme cardiaque ne peut être engendrée que par un retrait de l'activité parasympathique (Rowell, 1993). En cas de demande métabolique forte, l'activité sympathique prend le relais au-dessus d'environ 100 battements par minute, comme illustré sur la Figure 0.1.

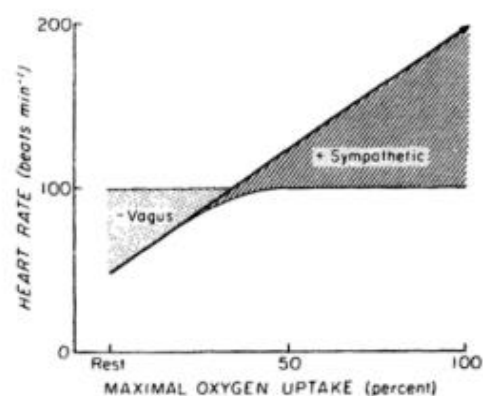
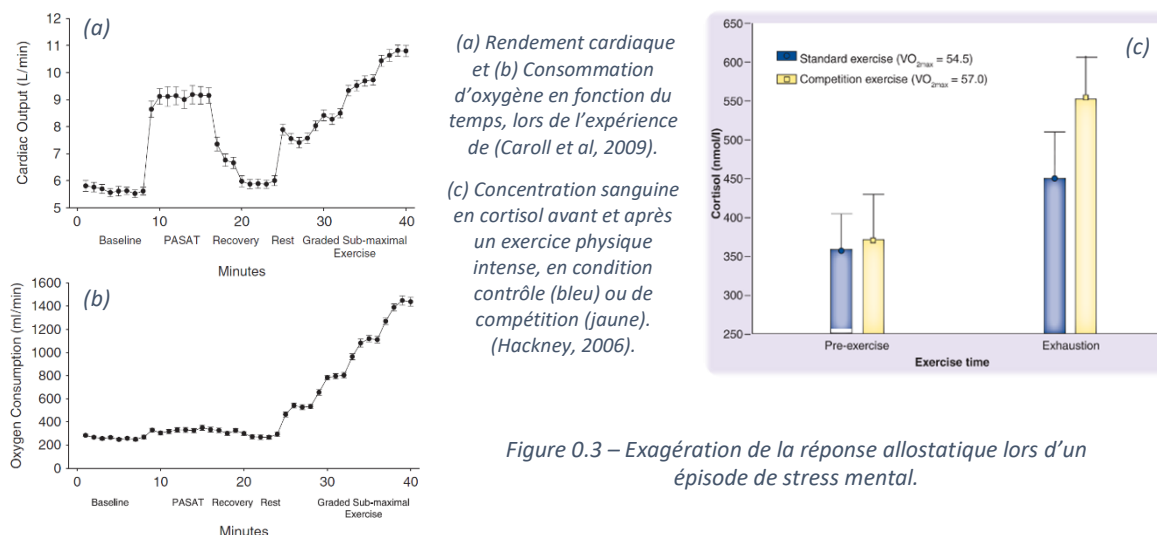


Figure 0.1 – Commande possible du rythme cardiaque en fonction de la demande métabolique en situation d'exercice (Rowell, 1993)

II. Effort Physique et Stress Mental

L'exposition chronique à l'exercice physique provoque une habitude de la réponse physiologique à la demande métabolique. En d'autres termes, une activité physique régulière diminue l'intensité du stress physique généré par l'effort (Hackney, 2006). Un exercice physique régulier a même été corrélé avec une diminution de la réponse physiologique en situation de stress mental (Rimmele et al., 2009). Le stress physique déclenché par l'exercice permet donc à l'organisme de s'habituer à la réponse allostatique. Pourquoi n'est-ce pas le cas pour le stress mental ? Une réponse possible à cette question se trouve dans l'exagération de la réponse allostatique dans le cas du stress mental : celle-ci véhicule en effet une offre énergétique bien au-delà des demandes de l'organisme stressé (Turner & Carroll, 1985).

Revenons sur l'exemple du rendement cardiaque qui, en situation d'exercice physique, est proportionnel à la demande métabolique. Carroll et collaborateurs ont ainsi tracé son évolution en fonction de la consommation en dioxygène (Carroll et al., 2009), lors d'une expérience constituée d'une tâche arithmétique (stress mental), enchaînée avec un exercice graduel sur vélo d'intérieur (stress physique). Ces données sont reproduites sur les graphiques (a) et (b) de la Figure 0.3. Sur la partie droite des deux courbes, le rendement cardiaque et la consommation en dioxygène évoluent de manière similaire durant l'exercice physique, évoluant graduellement sur une très grande amplitude. Durant la tâche arithmétique, la réaction cardiaque est comparable à un effort modéré, mais la variation de la consommation en O₂ est presque négligeable comparée à celle de l'exercice. Le stress mental induit ainsi une réponse physiologique qui mobilise les ressources énergétiques au-delà de la demande métabolique. D'un point de vue cardiovasculaire, les mécanismes de cette mobilisation sont similaires à ceux de l'effort physique (Sherwood et al., 1986). A la différence près qu'une réponse hémodynamique réflexe, au sein des tissus musculaires surperfusés en sang oxygéné, pourrait être liée à une installation chronique de l'hypertension.



L'exagération de la réponse allostatique fait ainsi une différence fondamentale entre le stress mental et le stress physique. Cette assertion reste vraie y compris lorsque du stress mental vient s'ajouter au stress physique. Le graphique (c) de la Figure 0.3 représente la concentration en cortisol avant et après l'effort, chez les mêmes sportifs placés ou non en situation de compétition (Hackney, 2006). On constate que la situation de compétition provoque la sécrétion de cortisol en plus grande quantité.

Annexe 2. Modèles Psychologiques du Stress

Dans les paragraphes qui suivent, on compare plusieurs modèles disponibles dans la littérature en psychologie du stress. Ces modèles ont servi de support théorique au Chapitre Deux - III.B.

I. Théorie transactionnelle

La théorie transactionnelle synthétise les formes émotionnelles du stress. En particulier, ce modèle rend compte des différentes manières de percevoir (et de réagir à) un même stressor. A rebours de la définition consensuelle proposée au Chapitre Deux - II, ce nouveau point de vue met à l'épreuve l'identité physiologique du stress mental.

I.A. Evaluation cognitive

I.A.1. Deux évaluations successives

Dans la section Chapitre Deux - II, on a défini l'état de stress (mental) comme une réponse physiologique (excessive) à un certain nombre de situations dites « stressantes ». Parallèlement, une situation est considérée comme « stressante » lorsqu'elle engendre cette réponse physiologique qui fait l'état de stress. Cette définition est récursive. D'après Lazarus et collaborateurs (1985), une telle situation se présente lorsque des variables supposées objectives (*ex.* : la *charge* de travail imposée par une situation) et d'autres subjectives (*ex.* : la *pression* liée au travail ressentie par une personne), recouvrent en fait un même concept – ici, le *stress* de la personne. On peut dépasser ce paradoxe considérant que la personne et son environnement forment un même système, et que seule compte la relation (ou « transaction ») qui lie le second à la première.

L'hypothèse de base est la suivante : l'individu *évalue* en permanence la situation dans laquelle il se trouve en fonction des implications que celle-ci peut avoir pour son bien-être. Ces transactions avec son environnement sont réalisées de manière inconsciente, au moyen de deux évaluations « cognitives⁸ qui se succèdent (Folkman et al., 1986).. (i) L'évaluation primaire concerne la *pertinence* de l'événement en cours par rapport à son bien-être, et la congruence de cet événement avec ses objectifs et ses croyances. Elle détermine une situation comme stressante en ce qu'elle s'oppose à deux types de situations non-stressantes⁹, à savoir : celles qui ne sont *pas pertinentes* pour le bien-être de l'individu, et celles qui constituent pour lui un *bénéfice* sans contrepartie. (ii) L'évaluation secondaire concerne la gestion concrète du stressor ; elle porte notamment sur le *potentiel de maîtrise* de l'individu face à la situation. Si la situation est déterminée comme *stressante*, l'évaluation secondaire entre en jeu pour différencier trois types de stress, en fonction du potentiel de maîtrise perçu (Lazarus & Folkman, 1987). (a) Le *préjudice*, dès lors qu'un dommage ou une perte est actuellement infligé au sujet ; (b) la *menace*, qui représente un préjudice potentiel à court ou à long terme ; ou (c) le *défi*, lorsque l'individu se focalise sur la maîtrise de la situation et sur les bénéfices du succès, plutôt que sur les dangers de l'échec.

⁸ Les recherches de Lazarus s'inscrivent dans le cadre de la théorie des évaluations cognitives, ou théorie multicomponentielle des émotions (Sander & Scherer, 2009).

⁹ Selon la version la plus récente de cette théorie, parue en 2001.

Le tableau ci-dessous résume ces catégories d'évaluation sous forme de table de vérité, où 1/0 signifie que l'implication est présente/absente dans la relation individu - environnement.

Tableau 0.1 – Caractérisation d'une situation selon la théorie transactionnelle

		Relation à l'environnement	Implications pour l'individu			
			Bonnes		Mauvaises	
			Actuelles	Potentielles	Potentielles	Actuelles
Évaluation Primaire / Secondaire	Non Pertinent	0	0	0	0	
	Bénéfice	1	0	0	0	
	Stress	Défi	0	1	1	0
		Menace	0	0	1	0
		Préjudice	0	0	0	1

1/0 signifie que l'implication est présente/absente dans la relation individu – environnement.

Les évaluations primaire et secondaire sont interdépendantes et s'effectuent en continu, si bien qu'un même événement peut revêtir successivement plusieurs étiquettes. Par exemple, un préjudice passé peut constituer une menace future ; et une menace actuelle peut évoluer en défi si le potentiel de maîtrise est réévalué à la hausse (Hazanov et al., 2009). L'évaluation primaire dessine un contour à la notion de *stresseur*. L'évaluation secondaire lui octroie des couleurs, c'est-à-dire une qualité intrinsèque, une perception catégorielle qui détermine ce qu'il convient de faire face à ce stresseur. En introduisant la notion d'évaluation cognitive, la théorie transactionnelle permet de compléter le modèle physiologique proposé au Chapitre 2.

I.A.2. Réponse émotionnelle

Un certain type d'évaluation déterminera le schéma d'action que l'individu devra mettre en place pour maîtriser la situation, dont le terme générique est « stratégie de coping » (de l'anglais *to cope* : surmonter, faire face). En d'autres termes, une stratégie de coping regroupe l'ensemble des actions entreprises par l'individu stressé pour diminuer son stress. Cela peut se produire soit par action sur le stresseur – c'est l'approche « centrée sur le problème » – soit par action sur la relation environnement/individu – c'est l'approche « centrée sur l'émotion » (Folkman & Lazarus, 1985).

L'approche transactionnelle propose ainsi un lien fondamental entre le stress et l'émotion, qui est vue comme le produit conscient de cette relation entre l'individu et son environnement. Pendant un épisode de stress, des émotions sont élicitées par l'évaluation secondaire à classer parmi le préjudice, la menace ou le défi. Le Tableau 0.2 ci-dessous fournit quelques exemples issus de (Folkman & Lazarus, 1985), qui catégorisent les émotions selon ces trois types d'évaluation. Les concepts de stress et d'émotion ne sont pas à confondre. L'évaluation non-stressante de *bénéfice*, par exemple, est également à l'origine d'émotions positives telles que la *joie* ou le *soulagement*.

Tableau 0.2 – Catégories de stress et émotions associées. Adapté de (Folkman & Lazarus, 1985)

Préjudice	Menace	Défi
Culpabilité	Inquiétude	Confiance
Dégoût, Déception	Anxiété	Espoir
Colère, Tristesse	Peur	Impatience

Une telle diversité des colorisations du stress par l'émotion ne peut être sans conséquence sur la physiologie de l'individu stressé. En particulier, des émotions différentes auront des corrélats différents en ce qui concerne l'activité nerveuse autonome, et les marqueurs couramment utilisés pour mesurer celle-ci (Kreibig, 2010). Il ne s'agit pas d'un élément confondant à proprement parler, puisque l'émotion est corrélative au stress. Les couleurs émotionnelles du stress en révèlent simplement la nature complexe et polymorphe. Cette complexité du stress peut-elle compromettre son identité physiologique, telle qu'on l'a présentée au Chapitre Deux - ?

I.B. Différentiation théorique de la réponse physiologique

Le lien entre la couleur émotionnelle du stress et sa réponse physiologique peut être étudié au travers des situations de *défi* et de *menace*. Une manière de symboliser les processus d'évaluation primaire et secondaire est de considérer que le sujet compare les *demandes* de l'environnement (niveau de risque) aux *ressources* dont il dispose pour y faire face (Harvey et al., 2010). Si les demandes excèdent les ressources, l'individu percevra la situation comme une menace. Si les deux quantités sont équivalentes, la situation sera évaluée comme un défi.

I.B.1. Réponse (in)adéquate

Tomaka, Blascovitch et collaborateurs (1993; 1997) ont ainsi traité la distinction menace / défi sur le plan physiologique, au moyen d'un protocole expérimental décliné en plusieurs versions. Sur la base d'un questionnaire confrontant les demandes et les ressources perçues, ou sur la base des conditions expérimentales, ils ont divisé leurs participants en un groupe montrant une évaluation de type « menace » et un groupe montrant une évaluation de type « défi ». De manière générale, les sujets du groupe « menace » obtenaient des performances moindres à la tâche et un sentiment de stress plus fort que ceux du groupe « défi ».

I.B.1.a) Réponse cardiovasculaire

Sur le plan physiologique, l'activité cardiovasculaire était mesurée en continu par cardiographie d'impédance. Tomaka et Palacios-Esquivel (1997) ont observé le schéma de réponse suivant :

- L'évaluation de type « défi » s'accompagnerait d'une mobilisation *efficace* des ressources de l'organisme. En particulier, on observerait une dilatation des vaisseaux périphériques (diminution de la résistance vasculaire) et une augmentation globale de l'activité cardiaque (contractilité du myocarde, rythme, rendement).
- L'évaluation de type « menace » s'accompagnerait d'une mobilisation *inefficace* des ressources de l'organisme. La réponse cardiovasculaire attendue serait présente, mais « désorganisée » : on observerait une vasomotricité inversée (augmentation de la résistance vasculaire périphérique), et une augmentation de l'activité cardiaque moins importante que dans le cas du défi.

Par la suite, ces mêmes auteurs ont lié les schémas de réponse psychologique et cardiovasculaire du « défi » et de la « menace » à la « croyance en un monde juste » (Tomaka & Blascovich, 1994). Ainsi, les corrélats physiologiques d'un épisode de stress seraient susceptibles de dépendre de la personnalité des individus.

I.B.1.b) Réponse neuroendocrinienne

Sur le modèle des expériences précédentes, (Harvey et al., 2010) ont simulé une réanimation avec des étudiants en médecine, en condition Stress Fort et Stress Faible selon les modalités de la tâche. La concentration en cortisol salivaire dépendait de la prédominance des demandes sur les ressources perçues ; mais seulement dans la condition Stress Fort, parmi les participants qui ont évalué la tâche comme une menace. Ces résultats suggèrent que le cortisol est surtout sécrété par l'organisme lorsque l'individu se sent menacé.

A ce sujet, Tomaka et Palacios-Esquivel (1997) ont établi un parallèle entre leur dualité défi / menace et la dualité effort / détresse proposée par Frankenhaeuser (1982). Ce modèle suggère que l'effort et la détresse seraient deux composantes complémentaires du stress et activeraient des voies neuroendocriniennes bien distinctes. Un *effort* sans détresse provoquerait seulement la sécrétion d'adrénaline, *via* l'activation du seul axe neuroendocrinien SAM. Une *détresse* sans effort provoquerait seulement la sécrétion de cortisol, *via* l'activation du seul axe neuroendocrinien HPA. En situation d'effort stressant, c'est-à-dire dans les activités de la vie courante, les deux mécanismes sont à l'œuvre (voir Chapitre Deux - II.B.2). Dans ce contexte, l'activité de l'axe HPA serait modulée par le *contrôle* perçu de l'individu sur la situation (Frankenhaeuser, 1982; Peters et al., 1998), d'où la différenciation défi / menace observée précédemment.

Par ailleurs, Tomaka et Palacios-Esquivel (1997) suggèrent que la différenciation défi / menace, ou encore effort / détresse, serait liée aux deux circuits cérébraux de la motivation (Fowles, 1988). L'*effort* serait initié par le *système d'approche* comportementale (*behavioral approach system*), lié à l'attrait d'une récompense ; tandis que la *détresse* serait initiée par le *système d'inhibition* comportementale (*behavioral inhibitory system*), lié à l'évitement d'une punition ou à une situation sans récompense. La réaction physiologique désorganisée de type « menace » serait liée à une coactivation de ces deux systèmes ; tandis qu'une réponse de type « défi » serait surtout liée au système d'approche comportementale.

I.B.2. Limites empiriques

Si ces auteurs ont pu vérifier cette hypothèse concernant l'activité cardiaque, leur étude échoue à la vérifier au niveau vasculaire et électrodermal (Tomaka & Palacios-Esquivel, 1997). Leur tâche arithmétique, qui était différente des précédentes (1994; 1993), n'a pas induit le schéma de réponse physiologique différencié prédit par Fowles (1988), entre « défi » et « menace ». Dans un autre protocole, des sujets féminins ont réalisé un discours improvisé sur un sujet polémique devant un examinateur masculin (Tomaka et al., 1999). A nouveau, la réponse cardiovasculaire observée fut à la fois conforme et différente de leurs prévisions. Contrairement aux attentes, le groupe « défi » avait un rythme cardiaque moins élevé que le groupe « menace ». Comme dans l'étude précédente (1997), mais contrairement aux autres (1994; 1993), les deux groupes ont montré une vasodilatation du système sanguin périphérique¹⁰.

Ainsi, bien qu'un schéma *relatif* de réponse physiologique différenciée ait été mis en évidence, soit une réponse *plus ou moins* adéquate à une mobilisation des ressources, on ne peut en retenir un modèle invariant pour chaque type de stress. Une telle différence ne se reflète pas toujours sur les

¹⁰ Remarque : (Tomaka et al., 1999) ont incriminé des différences dans les techniques de mesures utilisées entre les différents protocoles expérimentaux.

mesures périphériques, telles que le rythme cardiaque ou la résistance vasculaire périphérique. Cette limite illustre la nécessité de dégager une variation pertinente (comme la vasoconstriction) à l'échelle de plusieurs activités (arithmétique, discours, etc.), avant de lier celle-ci à une situation de stress. Une distinction explicite doit être faite entre le *stress* et l'*activité* qui l'accompagne. L'activité motrice n'étant pas seule responsable (voir Annexe 1), il faut aussi considérer l'activité cognitive comme source de bruit. Une étude des corrélats de l'effort mental est nécessaire pour définir, sur le plan physiologique, la notion d'*activité mentale non-stressante*.

II. Théorie de la Charge Cognitive

L'utilisation d'une tâche arithmétique pour induire une réponse physiologique n'est pas l'apanage de la recherche sur le stress émotionnel, pas plus que l'utilisation du discours. La qualité d'élocution a été prise pour indicateur d'un concept à la fois différent et semblable au stress, la *charge cognitive* (Yin & Chen, 2007). La littérature scientifique en matière de charge cognitive est aussi riche que celle en matière de stress, y compris l'étude de ses corrélats physiologiques. La notion d'*effort mental*, notamment, possède des points communs avec le stress. Pourtant, ces deux corpus de la littérature forment des ensembles quasi-disjoints de la recherche en psychologie cognitive. Avec un nouveau point de vue sur le stress psychologique, un bref aperçu du concept d'effort mental nous permettra (enfin) de redéfinir le stress *émotionnel* par contraste avec la charge mentale.

II.A. Terminologie de la charge cognitive

Si on appelle « mémoire à long terme » notre capacité (illimitée, inconsciente) à stocker des informations sur la durée, on appellera « mémoire de travail » notre capacité à stocker et manipuler temporairement un ensemble limité de ces informations qui se manifestent à la conscience pour

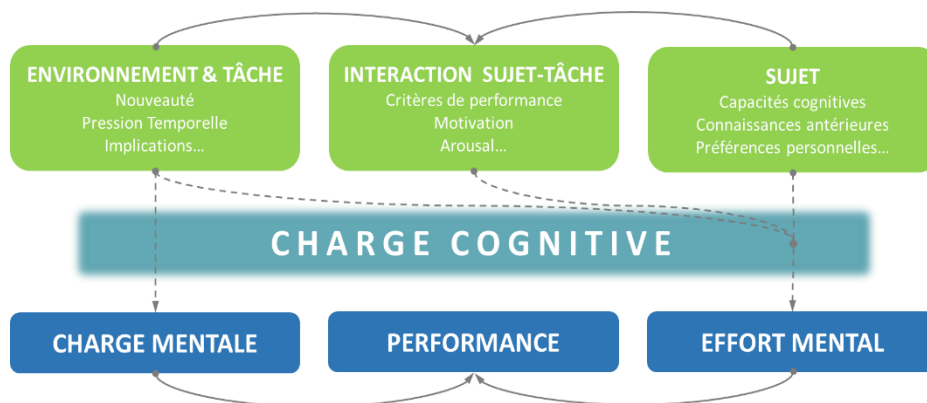


Figure 0.4 – Représentation de la charge cognitive : causes (en vert) et corrélats (en bleu).
Adaptée de (Paas & Merriënboer, 1994), complétée avec les informations du texte.

réaliser une tâche donnée (Sweller et al., 1998). On appelle alors « charge cognitive » la charge appliquée à la mémoire de travail, c'est-à-dire la quantité de ressources requises par l'individu pour réaliser la tâche. Un modèle synthétique de la charge cognitive a émergé des sciences de l'éducation, reprenant la terminologie couramment employée dans ce domaine de recherche. Dans ce modèle, la charge cognitive dépend (en vert dans la figure 1.5) : (i) de la complexité de la tâche (et le contexte dans lequel elle est réalisée) ; (ii) des ressources qui sont propres au sujet (capacités cognitives, connaissances antérieures, etc.) ; (iii) de l'interaction entre le sujet et la tâche. Ces facteurs

conditionnent la charge mentale, l'effort mental et a fortiori les performances à engager (en bleu dans la figure 1.5).

Par définition, les causes propres au sujet sont indépendantes de la tâche, (et vice-versa) ; et l'interaction entre ces deux entités exprime le vécu de la tâche par le sujet. Des facteurs relevant de ces trois catégories sont proposés en vert sur la Figure 0.4. Le stress partage ainsi l'essentiel des facteurs de charge cognitive, jusqu'à l'état d'éveil (*arousal*) de l'individu. Une conséquence de cette similarité se trouve dans l'utilisation des mêmes tâches standard pour induire l'un et l'autre, comme le test de Stroop (Fairclough & Houston, 2004), une tâche arithmétique (Paas et al., 1994), une tâche mémorielle (Backs & Seljos, 1994), ou une tâche de conduite (Mehler et al., 2009).

La *charge mentale* est la quantité des ressources cognitives exigée par la tâche et par les contraintes de l'environnement (bruit, température, etc.). Par définition, la charge mentale est indépendante du sujet et de son vécu sur la situation. Parallèlement, l'*effort mental* correspond aux ressources réellement déployées par le sujet pour répondre à ces exigences. Il dépend à la fois du sujet, de la tâche et de l'interaction entre les deux. Comme l'engagement d'un individu n'est pas toujours lié à la complexité d'une tâche, l'effort et la charge mentale peuvent être décorrélés entre eux (Paas & Van Merriënboer, 1994). La *performance* résulte du différentiel entre la charge mentale (« demandes » de la tâche en cours), l'effort mental et l'engagement. Par exemple, un effort supplémentaire sur une tâche difficile permettra d'obtenir de meilleures performances, et un effort insuffisant donnera des performances moindres. La *fatigue* également se traduit, si possible, par un effort compensatoire, et sinon, par une diminution des performances.

On identifie souvent la charge cognitive à la seule notion d'effort mental, car la charge mentale est inaccessible à la mesure. Comme le stress, on peut quantifier l'effort mental par des réponses verbales (Schmeck et al., 2015), ou par la réponse physiologique (Fairclough & Mulder, 2011). Les paramètres physiologiques suggérés pour mesurer l'effort mental sont très similaires à ceux du stress (voir Chapitre Trois - I).

II.B. Effort mental et stress

La littérature propose un nombre restreint d'approches qui distinguent explicitement le *stress* de l'*effort mental*, en particulier, les modèles dits « cognitifs-énergétiques ». Par exemple, Hockey (1997) a défini le stress comme l'inadéquation des ressources disponibles (le « budget d'effort ») aux demandes de la tâche en cours, provoquée par un élément perturbant. L'ajustement de ce décalage peut se faire de deux manières : soit par dévaluation des objectifs (diminution des demandes), soit par un « effort compensatoire » (augmentation des ressources). Cet effort compensatoire aurait un « coût » psychophysiologique caractérisé par un sentiment d'anxiété ou de fatigue, une activité sympathique dominante sur le parasympathique et une sécrétion accrue d'adrénaline et de cortisol.

De son côté, Gaillard (2001; 1994) a considéré non pas l'insuffisance des ressources, mais les stratégies de mobilisation de celles-ci (Staal, 2004). En particulier, le stress était vu comme une stratégie de mobilisation inefficace (à l'instar de la réaction-type *menace* proposée par Tomaka (1993)). Par contraste, la charge mentale était vue comme une stratégie de mobilisation des ressources adaptée à la situation. Cette réponse peut être assortie d'émotions positives (Gaillard, 2001) – à l'instar de la réaction-type *défi* proposée par Tomaka (1993).

Ces deux conceptions du stress héritent de la dualité effort/détresse proposée par Frankenhaeuser (1982). Elles défendent l'existence de deux schémas distincts de régulation des ressources, qui ajustent l'effort à la demande pour assurer la performance. Dans le premier schéma, l'effort est proportionnel à la demande objective (*charge mentale*) ; il s'accompagne d'une activation du seul axe SAM ; et cesse avec l'activité concernée par cette demande. Dans le second schéma, l'effort n'est pas proportionnel à la demande objective ; il s'accompagne d'une coactivation des axes SAM et HPA ; et persiste après l'activité. Le stress (*émotionnel*) se situerait dans ce deuxième schéma et la répétition d'une telle réponse persistante serait à l'origine du stress chronique. Cette association du stress avec une réaction physiologique excessive est analogue au raisonnement proposé au Chapitre Deux - III.A.

II.C. Charge affective

L'identification du stress à des expériences seulement négatives (réponse-type « menace ») ou également positives (réponse-type « défi ») ne fait pas consensus dans la littérature, comme la montre la définition de Lazarus et Folkman (1985). Ainsi, on peut considérer l'(in)existence d'un « bon stress » selon la définition qu'on donne au phénomène – une mobilisation générique des ressources de l'organisme, ou seulement sa version « inadéquate ».

On peut, en revanche, considérer la situation « d'effort faible », pour laquelle des ressources sont immédiatement disponibles pour répondre à la demande environnementale. Selon Hockey (1997), cette situation-là donnera systématiquement lieu au premier schéma de réponse développé plus haut, qui correspond à l'effort-sans-détresse proposé par Frankenhaeuser (1982), à savoir, une activation de l'axe SAM et une prédominance modérée de l'activité sympathique sur le parasympathique. Celle-ci correspond à une réponse-réflexe de l'organisme (que Fairclough & Mulder (2011) ont appelé « réaction de défense »), à une demande cognitive quelconque (charge mentale), pourvu que des ressources soient immédiatement disponibles pour y faire face.

En théorie, il est ainsi possible de distinguer une activité cognitive stressante d'une activité non-stressante sur la base d'implications physiologiques différentes. Pour reprendre la terminologie de la charge cognitive, nous appellerons *charge mentale* la réponse psychophysologique contingente au traitement cognitif de la situation en cours. Par contraste, nous appellerons *charge affective* la réponse psychophysologique qui dévie de cette mobilisation strictement nécessaire. Le *stress émotionnel*, c'est-à-dire notre concept-cible, est inclus dans celui de charge affective.

Cette approche n'est pas seulement nécessaire pour nommer les deux schémas de réaction que nous avons décrits plus haut : elle est aussi cohérente avec les définitions précédentes du stress psychologique. Selon l'approche transactionnelle, le stress psychologique apparaît lorsque la situation est pertinente pour le bien-être de l'individu, et sollicite « trop » de ses ressources. Ainsi, une tâche peu complexe a peu de chances d'être stressante. Par ailleurs, l'exposition quotidienne à de la charge mentale ne suffit pas à induire un stress chronique selon le modèle Demande-Contrôle de Karasek & Theorell (1990), ou le déséquilibre Effort-Récompense de Siegrist & Peter (1994). La distinction charge affective / charge mentale est donc pertinente aussi bien à l'échelle micro- que macro-analytique.

Annexe 3. Stratégies de Neutralisation de l'Effet du Sujet sur les Corrélats du Stress

On appelle *effet du sujet* les différences interindividuelles dans la réaction physiologique au stress. Dans le Chapitre Trois - , on montre que cet effet du sujet peut être induit par des paramètres propres à chaque individu (biologie, comportement). Cette annexe synthétise les stratégies qui permettent de prendre en compte ce type d'effet du sujet sur les caractéristiques physiologiques.

Dans la littérature, on peut dégager trois tendances dans la manière de traiter cet effet. Ces approches sont résumées dans le Tableau 0.3 ci-dessous.

Tableau 0.3 : Stratégies de neutralisation de l'effet Sujet.

Modèle		Multi-utilisateur générique	Multi-utilisateur calibré	Mono-utilisateur
(i) Construction du Modèle	Protocole	Un groupe expérimental est enregistré en situation(s) de Stress et de Non-Stress	" "	" "
	Apprentissage	Sur toutes les données disponibles	Sur toutes les données disponibles, normalisées par individu	Sur chaque individu du groupe
(ii) Application du Modèle	Protocole	/	L'utilisateur final est enregistré en situation(s) de Repos ⁽¹⁾	L'utilisateur final est enregistré en situation(s) de Stress et de Non-Stress
	Apprentissage	/	Calibration sur les données de l'utilisateur final	Le modèle est appris sur l'utilisateur final
Principal Avantage		Aucun protocole pour l'utilisateur final	Neutralisation partielle de l'effet Sujet	Neutralisation optimale de l'effet Sujet.
Principal Inconvénient		Aucune neutralisation de l'effet Sujet	Protocole léger pour l'utilisateur final	Protocole lourd pour l'utilisateur final

" " : identique à la case de gauche. / : Identique aux cases du dessus.

(1) C'est une des réalisations possibles de cette approche (voir dans le texte)

Dans chacun des cas, un modèle est appris sur un groupe expérimental en situation de stress et de non stress. Dans ce contexte, l'effet du sujet peut intervenir sur deux niveaux successifs. (i) Lors de la construction du modèle, on optimise les paramètres à partir d'un groupe de participants qui n'est pas forcément représentatif de toute une population. (ii) Lors de l'application du modèle, on calcule ses caractéristiques physiologiques sur un individu distinct du groupe expérimental : *l'utilisateur final*. De manière générale, on peut douter de l'efficacité d'un modèle de stress sur un individu distinct du groupe d'apprentissage (*validité externe*).

On ne cherche pas à neutraliser l'effet du sujet lorsqu'on apprend un modèle sur les valeurs brutes de ses caractéristiques. Cette stratégie « générique » est rarement assumée telle quelle dans la littérature, ou seulement pour quelques caractéristiques (Salazar-Ramirez et al., 2018), car elle suppose que le comportement des caractéristiques est universel.

La stratégie opposée consiste à apprendre un modèle différent pour chaque individu puis à calculer la performance moyenne pour valider l'algorithme d'apprentissage. Le même algorithme est appliqué pour chaque nouvel utilisateur du modèle. Ce type d'approche « mono-utilisateur » implique de placer l'utilisateur final en situation de *stress* et d'activité *non-stressante*, afin d'obtenir des données spécifiques à cette personne dans ces conditions. On neutralise ainsi l'effet sujet au niveau de l'élaboration (i) et au niveau de l'application (ii) de la mesure de stress. Ce type de modèle peut être plus efficace plus efficace qu'un modèle multi-utilisateur (Hernandez et al., 2011). Mais son application est un fardeau pour l'utilisateur final qui doit subir un protocole expérimental standard ; ce qui rend improbable l'utilisation d'une telle stratégie sur un « consommateur » lambda.

Entre ces deux extrêmes, un compromis consiste à *normaliser* les caractéristiques à partir de mesures effectuées sur la personne : il en résulte un modèle « multi-utilisateur calibré » introduit dans le Tableau 0.3. Par défaut, on réalise un centrage – réduction des valeurs f de la caractéristique à partir d'une moyenne de référence μ_r et d'un écart-type σ_r calculés pour chaque individu :

$$f_{norm} = \frac{f - \mu_r}{\sigma_r} \quad (0-2)$$

La calibration est effectuée sur le groupe expérimental d'une part, et sur l'utilisateur final d'autre part, de sorte que les valeurs f_{norm} de chaque individu soient réparties dans un espace commun centré sur Zéro. D'autres calculs sont possibles, comme la réduction min-max. Les valeurs de référence peuvent être l'ensemble des mesures réalisées sur la personne, ou bien des mesures effectuées en situation de repos. L'avantage du repos est qu'il propose des conditions de mesure standard et facilement répliquables entre les individus, ainsi la normalisation reste indépendante des activités de la personne.

Cette dernière stratégie permet de venir à bout du problème de *répartition* des données dans l'espace des caractéristiques, sans imposer un protocole lourd à l'utilisateur final (quelques minutes de repos, éventuellement répétées sur des jours différents). Lors de l'apprentissage du modèle, sa validité externe peut être augmentée par une validation croisée de type LOSO (voir Chapitre Trois - III.A). Dans ce dernier cas, les paramètres du modèle ne sont pas optimisés pour l'utilisateur final. Ainsi, la performance mesurée pour le modèle peut être moins élevée que dans un modèle modèle mono-utilisateur. Cette problématique est traitée au Chapitre Trois - section III.B.

Annexe 4. Modèles élaborés en conditions contrôlées

Tableau 0.4 : Synthèse de l'état de l'art en matière de détection du stress sur des données expérimentales enregistrées en conditions contrôlées

	(Healey & Picard, 2005)	(Zhai & Barreto, 2006b)	(Shi et al., 2010)	(Setz et al., 2010)	(Sierra et al., 2011)	(Sun et al., 2012)	(Sun et al., 2012)	(Villarejo et al., 2012)	(Choi et al., 2012)	(Melillo et al., 2013)	(Wijsman et al., 2013)	(Mohino-Herranz et al., 2015)	(Sandulescu et al., 2015)	(Hovsepian et al., 2015)	(Gjoreski et al., 2016)
Stresseur	Conduite en voiture	Test de Stroop	TSST, Tâche arithm., CPT	Tâche arithm.	Préparation d'oral	Test de Stroop, Tâche arithm.	Test de Stroop, Tâche arithm.	Nervosité simulée, Calcul, Lecture rapide	Stroop, TSST, Mémoire, Dessin, Attention	Examen verbal	Tâche arithm., Puzzle logique, Tâche mémoire	Tetris & Tâche arithm.	Tâche arithm., TSST	TSST, Tâche arithm., CPT	Tâche arithm.
Mesures incluses	ECG, AED, Resp, EMG	AED, PPG, Diam, Pupille, Temp. Peau	ECG, AED, Resp, Temp. Peau	AED	AED, ECG	AED, ECG, Accéléro.	AED, ECG, Accéléro.	AED	PPG, Resp, AED	ECG	ECG, AED, Resp., EMG	ECG, Impéd. thorax	AED, PPG	ECG, Resp.	AED, PPG, Temp. Peau
Algorithme d'apprentissage	Analyse Discriminante	Séparateur à Vastes Marges	Séparateur à Vastes Marges	Analyse Discriminante	Réseau neurones flou	Arbre de Décision	Séparateur à Vastes Marges	Weka BayesNet	Régression Logistique	Arbre de Décision	Régression Logistique	Perceptron (1 couche cachée)	Séparateur à Vastes Marges	Séparateur à Vastes Marges	Forêt Aléatoire
Précision	100,0%	90,1%	68,0%	82,8%	98,9%	92,4%	81,0%	90,9%	81,0%	82,1%	74,5%	67,7%	80,9%	95,3%	83,0%
C0. Modèle interprétable	1	0	0	1	0	1	0	0	1	0	1	0	0	0	0
C1. Tâche contrôlée	0	0	1	1	0	0	0	0	0	0	1	1	1	0	1
C2. Entraînement multi-tâche	0	0	1	0	0	1	1	1	1	0	1	1	1	1	0
C2_{bis}. Validation multi-tâche	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C3. Entraînement multi-utilisateur	1	1	1	1	1	1	1	0	1	1	0	1	0	1	1
C3_{bis}. Validation multi-utilisateur	0	0	1	1	0	0	1	0	1	0	0	1	0	1	1
Score (/6)	1	1	5	4	1	2	3	1	3	1	3	4	3	3	4

Le score est calculé selon la formule suivante (Chapitre Trois - III.B.2) : $Score = 2 * C1 + C2 + C2_{bis} + C3 + C3_{bis}$

Annexe 5. Extraction des Caractéristiques

I. Méthode générale

Soit un signal physiologique brut exprimé en fonction du temps : $\varphi(t)$, $t \in [0 ; t_{fin}]$; où $t=0$ correspond au début de la condition expérimentale ($e \in \{Rc, Rc, C, S\} \times \{D2, SECPT, TSST, TA\}$), et $t=t_{fin}$ correspond à la fin. Dans le reste de ce chapitre, les signaux bruts dont on dispose sont :

- i) L'intervalle inter-battements **IBI(t)**, qui prend la valeur de l'intervalle entre un battement cardiaque à l'instant t et celui de l'instant $t-1$. Ce signal est échantillonné de manière irrégulière.
- ii) La conductance de la peau **SC(t)**. Ce signal possède deux composantes largement connues : la composante *tonique*, à variations lentes, et la composante *phasique*, à variations rapides. Ce signal est échantillonné à 1KHz.
- iii) Le signal respiratoire **Resp(t)**, mesuré par pléthysmographie d'inductance. L'information pertinente est contenue dans les variations de ce signal, qui reflète le volume instantané de la cage thoracique. Ce signal est échantillonné à 1KHz.

Une caractéristique physiologique correspond à un traitement sur le signal $\varphi(t)$ qui (a) recouvre une certaine *fenêtre* de ce signal, et (b) cherche à quantifier un phénomène physiologique observable sur cette fenêtre. Par exemple, la moyenne du rythme cardiaque. On appellera L la largeur d'une fenêtre de signal et τ l'intervalle temporel couvert par cette fenêtre entre un instant t et l'instant $t+L$. L'extraction de caractéristiques est un procédé en trois étapes :

1. Traitement du signal brut : le signal φ est traité de sorte à en extraire un signal ou plusieurs signaux d'intérêt qui serviront de base au calcul des caractéristiques.
2. Découpage du signal prétraité en K fenêtres τ_k de temps de largeur L , où $K = \lfloor \frac{t_{fin}}{L} \rfloor$.
3. Calcul des caractéristiques sur chaque fenêtre τ_k .

Dans l'optique d'un transfert en vie réelle, notre procédé d'extraction de caractéristiques avait pour contrainte d'être automatisable, exécutable en temps réel et nécessitant peu de ressources en calcul. Parallèlement à nos travaux, ce transfert a pu être réalisé par transcription du code Matlab en langage C (*via MatlabCoder™*), compilation de l'algorithme et test sur une application smartphone.

II. Signal Cardiaque

II.A. Traitement du Signal Brut

Après l'expérience, les signaux ECG ont été traités de sorte à en extraire le signal d'intervalle inter-battements (**IBI**). Les battements cardiaques ont été repérés au moyen d'une procédure standard de reconnaissance des pics, conçue pour le signal ECG : l'algorithme de Pan et Tompkins (1985). Celui-ci traite le signal en quatre étapes successives. (i) Le signal est *filtré*, passe-haut puis passe-bas, dans la bande de fréquences [5 ; 15]Hz, afin de supprimer la ligne de base, le bruit de fond et diminuer les artéfacts d'origine musculaire. (ii) Le signal est *dérivé* à l'aide d'un filtre dérivateur, puis (iii) le signal est *mis au carré*, afin de mettre en relief les complexes QRS qui marquent chaque battement cardiaque. (iv) Le signal est finalement *lissé* (moyenne glissante de 150ms) pour s'affranchir du bruit

résiduel après mise au carré. A la suite de ce traitement, un critère de décision est mis en œuvre pour déterminer les pulsations du signal résultant qui correspondent réellement au complexe QRS. Cet algorithme a été entièrement implémenté sous MATLAB par H. Sedghamiz (2014) et utilisé tel quel sur nos signaux ECG. Pour finir, l'intervalle temporel entre deux complexes QRS était considéré comme échantillon IBI s'il ne dépassait pas 3 secondes.

II.B. Caractéristiques Extraites

II.B.1. Caractéristiques standard

Le rythme cardiaque instantané **HR(t)** est la fonction inverse de l'IBI à chaque instant t . Le rythme cardiaque moyen fait partie des caractéristiques standard utilisées dans la détection du stress ; avec les paramètres de variabilité cardiaque (**HRV**), dont le lien avec le stress a été démontré au Chapitre Trois - II.A. Le nom de ces caractéristiques et leur lien avec le stress est rappelé dans le Tableau 0.5 ci-dessous.

Tableau 0.5 : Évolution attendue des caractéristiques standard du signal cardiaque

Caractéristique	Évolution attendue
Moyenne (HR)	↗
RMSSD (IBI)	↘
Puissance HF (IBI)	↘
Puissance LF (IBI)	↗
Rapport LF/HF (IBI)	↗

Pour mémoire, **RMSSD** signifie : *Root Mean Square of the Successive Differences*. **LF** signifie *Low Frequencies* : c'est l'acronyme standard pour désigner les fréquences du signal cardiaque comprises entre 0,04Hz et 0,15Hz. **HF** signifie *High Frequencies* : c'est l'acronyme standard pour désigner les fréquences du signal cardiaque comprises entre 0,15Hz et 0,5Hz. Par « puissance LF » ou « puissance HF », on entend la puissance spectrale dans ces deux bandes de fréquence. Le rapport LF/HF est le rapport de ces deux puissances spectrales.

II.B.2. Caractéristiques supplémentaires

On a également calculé la RMSSD sur le signal HR, afin de comparer cette caractéristique à celle issue de l'IBI. On a calculé l'écart-type du signal de rythme cardiaque, qui reflète l'amplitude globale de ses variations sur la fenêtre temporelle. Enfin, il existe deux moyens de calculer la puissance spectrale dans une bande de fréquence : on peut chercher à normaliser ou non par la puissance totale du signal.

II.C. Méthode d'Extraction

Comme le signal IBI utilisé ici provient d'une instrumentation standard dans un environnement où l'activité physique était faible, les artéfacts étaient rares (souvent absents) dans la plupart des enregistrements. Les signaux IBI ne comportaient pas de valeurs extrêmes. On a fait le choix de ne pas repérer ou corriger les artéfacts pour ne pas s'éloigner davantage d'un signal observable en vie réelle.

Pour le calcul de la RMSSD, cependant, les différences successives entre les échantillons IBI ont été triées selon leur caractère « effectivement successif », c'est-à-dire ne résultant pas d'un pic manqué sur l'électrocardiogramme. On a considéré que deux échantillons étaient « effectivement successifs » lorsqu'ils étaient séparés de moins de deux secondes, ce qui correspond à 30 battements par minute sur le signal de rythme cardiaque. On a ainsi obtenu un signal **Sdiff**, donné pour chaque instant d'échantillonnage t_n par la formule suivante :

$$Sdiff(\phi(t_n)) = \begin{cases} \phi(t_n) - \phi(t_{n-1}) & \text{si } t_n - t_{n-1} < 2s \\ \emptyset & \text{sinon} \end{cases} \quad (0-3)$$

La puissance spectrale a été extraite à partir de la densité spectrale de puissance (DSP). La DSP d'un signal est habituellement donnée par le carré de sa transformée de Fourier (normalisé par la largeur L de la fenêtre de temps τ). Mais la transformée de Fourier suppose un échantillonnage régulier. Pour un signal d'échantillonnage à fréquence variable, une estimation de la DSP peut être obtenue par la méthode de Lomb-Scargle (Lomb, 1976; Scargle, 1982) qui utilise un algorithme de type moindres-carrés. Cette méthode est efficacement implémentée dans la fonction native *plomb* de la toolbox Signal Processing sur Matlab. La DSP a été ainsi calculée sur le signal IBI privé de sa composante

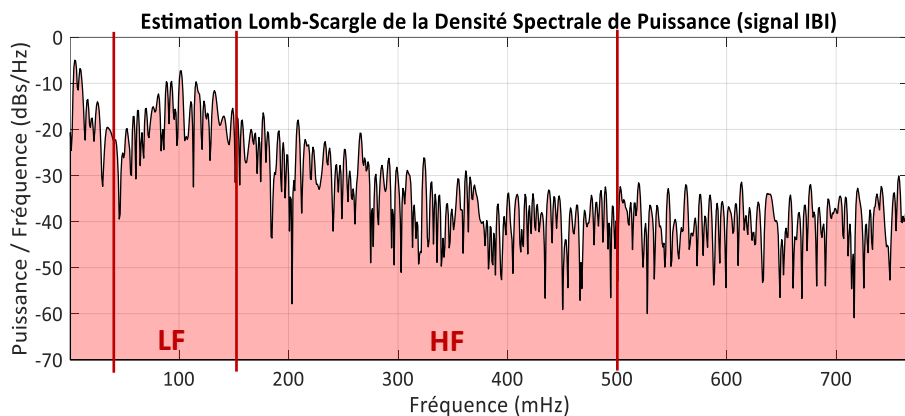


Figure 0.6 : Puissances spectrales LF et HF mesurées sur un périodogramme de l'intervalle inter-battements

continue, sur toute la fenêtre temporelle τ . Enfin, la puissance LF et la puissance HF ont été extraites en intégrant la DSP sur l'intervalle de fréquence qui convient, via la méthode des trapèzes. Par la suite, ces puissances ont pu être normalisées par l'aire totale sous la courbe DSP. La Figure 0.6 ci-dessous représente la densité spectrale, à l'issue de la fonction *plomb*, d'un signal IBI enregistré pendant la tâche TSST τ . On y a matérialisé les limites de fréquence LF et HF afin d'illustrer l'extraction des puissances spectrales correspondantes.

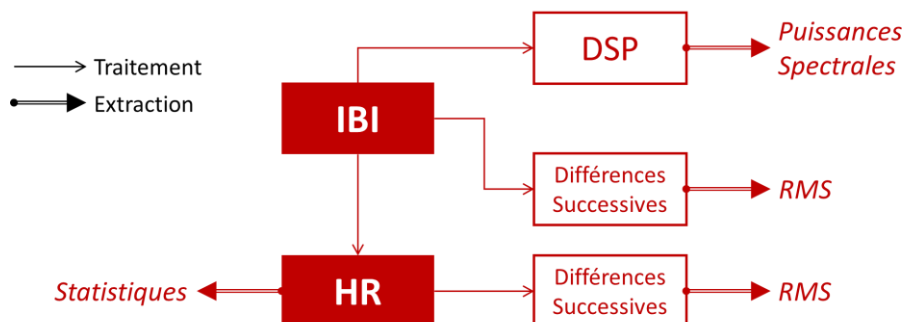


Figure 0.5 : Procédé d'extraction de caractéristiques sur le signal cardiaque

On constate notamment la présence du pic dit « Mayer-Hering », à 100mHz, dont l'amplitude est sensée évoluer en situation de stress. Pour finir, la chaîne complète des traitements sur le signal cardiaque est résumée dans la Figure 0.5.

III. Caractéristiques du Signal Electrodermal

III.A. Caractéristiques Extraites

III.A.1. Caractéristiques standard

Comme introduit au Chapitre Trois - II.B, on considère deux composantes dans l'activité électrodermale. Dans le cas de la conductance, la composante *tonique* est le niveau de base du signal (appellation standard : *skin conductance level*, **SCL**), à laquelle s'ajoutent des variations rapides et transitoires de la conductance : les réponses électrodermales, ou *skin conductance responses* (**SCRs**). Celles-ci forment la composante dite *phasique* du signal et traduisent l'excitation des glandes sudoripares par les fibres nerveuses efférentes. Sur une fenêtre temporelle τ , le niveau moyen du SCL comme le nombre total (i.e. la fréquence globale) des SCRs sont souvent considérés comme un étalon-or de l'activité nerveuse sympathique sur la peau. Le Tableau 0.6 rappelle leur évolution attendue en situation de stress.

Tableau 0.6 – Évolution attendue des caractéristiques électrodermales standard en situation de stress.

Caractéristique	Évolution attendue
Moyenne (SCL)	↗
Fréquence des SCRs	↗

III.A.2. Caractéristiques supplémentaires

Dans leur tentative de discriminer le stress de la charge mentale, cependant, Setz et collaborateurs (2010) ont trouvé d'autres résultats. Sur les 10 modèles de stress correspondant à autant de protocoles d'apprentissage, le SCL moyen n'a pas été sélectionné une seule fois, tandis que la fréquence totale des SCRs obtenait un score de 4/10. Ces propriétés *globales*, facilement calculables du signal électrodermal ont été délaissées par les algorithmes au profit de propriétés *locales*. Comme la fréquence *instantanée* des réponses électrodermales, dont le 85^e percentile sur la fenêtre temporelle τ a été sélectionnée 7 fois sur 10. De même, l'amplitude médiane des SCRs a été sélectionnée 9 fois sur 10 ; tandis que pour la composante tonique, la *dérivée* obtenait un meilleur score (5/10) que SCL moyen. Ainsi, une analyse fine et individuelle de chaque réponse électrodermale apparaît prometteuse dans le cas où la comparaison des caractéristiques ne se limite pas au signal de repos.

III.B. Méthode d'extraction

III.B.1. Stratégie générale

On a d'abord traité le signal brut de conductance cutanée (**SC**) afin de retirer les pics qui ne pouvaient prétendre à l'étiquette de réponse électrodermale (ce traitement est détaillé dans le paragraphe suivant). A partir du signal corrigé de ses artéfacts (**SC_{traité}**), on pouvait ensuite extraire les

variations lentes de la conductance à l'aide d'un filtre passe-bas de fréquence de coupure 0,1Hz (les passe-bas utilisés dans les paragraphes à venir sont des filtres Butterworth d'ordre 2). Bien qu'il ne s'agisse pas exactement d'un « niveau de base », on appellera **SCL** ce signal qui reflète les variations de SC au travers des réponses électrodermales. De même, on a conservé les variations rapides du signal $SC_{\text{traité}}$ à l'aide d'un filtre passe-bas de fréquence de coupure 1,0Hz. On obtenait ainsi un signal

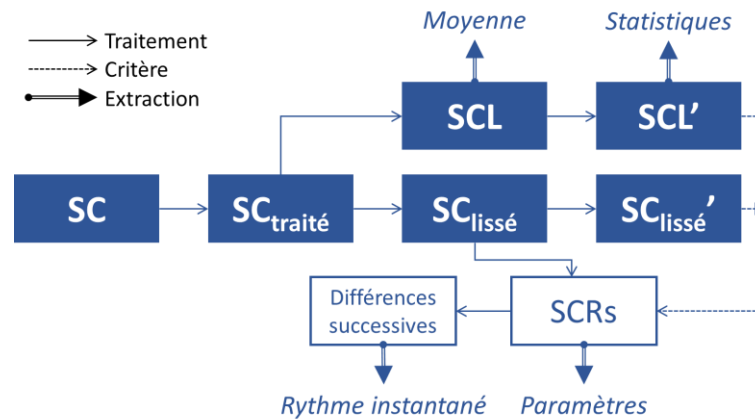


Figure 0.7 : Procédé d'extraction de caractéristiques sur le signal électrodermal

« rapide » $SC_{\text{lissé}}$ qui épousait fidèlement le signal original en réduisant drastiquement le bruit, afin que celui-ci soit propre à la détection de pics qui a permis de localiser chaque candidat à l'étiquette de « **SCR** » (réponse électrodermale). Par la suite, on a calculé les dérivées temporelles des signaux SCL et $SC_{\text{lissé}}$ (**SCL'** et **$SC_{\text{lissé}}'$**) ; ces dérivées ont servi de critère pour discriminer les SCRs et repérer dans le temps le début la fin de chacune d'entre elles. Pour finir, l'intervalle temporel entre les réponses électrodermales successives a été calculé pour constituer une mesure du *rythme instantané* des SCRs (« *scrr* » dans le Tableau 4.2). Les caractéristiques extraites de ces signaux sont des statistiques : moyenne, RMS, écart-type et 85^e percentile. Le procédé complet est résumé dans le schéma de la Figure 0.7.

III.B.2. Correction d'artéfacts

Des pics de pression positive ou négative de grande amplitude et de variation rapide étaient visibles dans le signal brut **SC**. Ces pics sont susceptibles d'être automatiquement traités comme des réponses électrodermales, polluant ainsi le calcul de caractéristiques comme l'amplitude médiane des SCRs. Afin de détecter ces pics de manière automatique, le signal original **SC** était filtré passe-bas à 0,5Hz par un filtre Butterworth d'ordre 2. Avec cette fréquence de coupure, le signal filtré doit épouser la composante phasique de la courbe SC, à condition que celle-ci varie dans les temps caractéristiques d'une réponse électrodermale (une seconde au minimum). On a considéré comme artéfact toute portion du signal brut qui déviait du signal filtré de plus de 0,2 μS . Pour corriger le signal sur cet

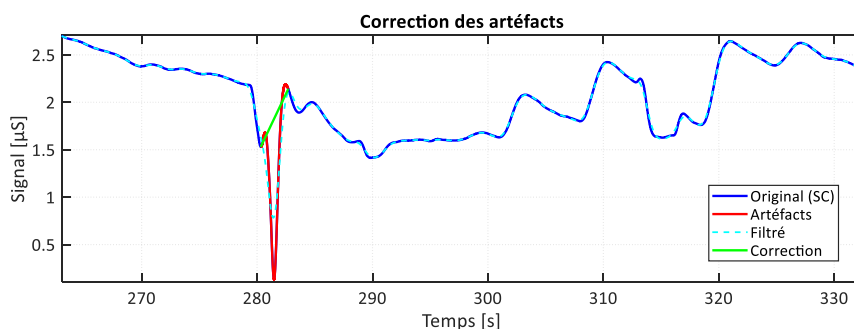


Figure 0.8 : Détection et correction d'un artéfacts sur le signal de conductance de la peau

intervalle temporel, la portion considérée comme artéfact a été interpolée linéairement. Le procédé est illustré dans la Figure 0.8.

III.B.3. Caractérisation des réponses électrodermales

L'algorithme implémenté pour détecter les SCRs utilise notamment des critères de dimensions pour sélectionner un pic, et des critères de pente pour déterminer ces dimensions. Tous les pics positifs

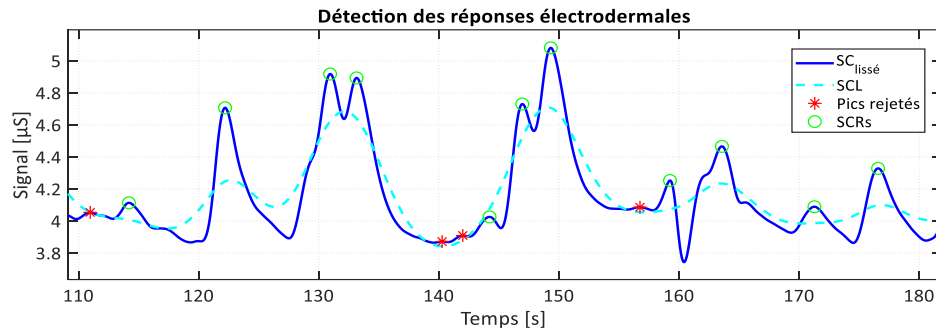


Figure 0.9 : Détection et sélection des réponses électrodermales sur le signal de conductance lissé.

sont d'abord repérés sur le signal $SC_{lissé}$. Puis, selon les recommandations de Green et collaborateurs (2014), on trie les pics repérés selon les critères suivants : un temps de montée compris entre 0,5s et 5s ; une amplitude minimale $0,02\mu S$; et un temps de récupération maximal de 10s. Une illustration de cette sélection des pics sur est donnée dans la Figure 0.9.

Les caractéristiques individuelles de chaque pic ont ensuite été calculées de la manière suivante :

- Pour le calcul du rythme instantané (scr_r dans le Tableau 4.2), on repère dans le temps chaque SCR par le moment de son sommet, soit de la dérivée du signal rapide : $SC_{lissé}'(t_{sommet}) = 0\mu S/s$
- Le temps de montée (scr_{Rise} dans le Tableau 4.2) est l'intervalle temporel séparant le début de la réponse et le sommet du pic. Le début de la réponse est le premier instant où la dérivée du signal rapide dépasse celle du signal lent¹¹ de $0,02\mu S/s$: $SC_{lissé}'(t_{début}) - SCL'(t_{début}) = 0,02\mu S/s$.
- Le temps de récupération (scr_{Rec}) est l'intervalle temporel séparant le sommet du pic de la fin de la réponse. La fin de la réponse est le dernier instant où la dérivée du signal lent dépasse celle du signal rapide de $0,01\mu S/s$: $SC_{lissé}'(t_{fin}) - SCL'(t_{fin}) = -0,01\mu S/s$.
- Enfin, l'amplitude d'une réponse (scr_{Amp}) est calculée à partir des valeurs du signal aux trois instants définis plus haut : $scr_{Amp} = SC_{lissé}(t_{sommet}) - \frac{1}{2}(SC_{lissé}(t_{début}) + SC_{lissé}(t_{fin}))$.

IV. Signal Respiratoire

Le signal respiratoire estime les variations du volume thoracique en fonction du temps. Il oscille lentement autour d'une valeur arbitraire et sur une amplitude qui n'est pas calibrée pour refléter un réel volume d'air inspiré. On se restreindra donc ici à une estimation globale de la fréquence, de l'amplitude et de la régularité des variations du signal dans la fenêtre de temps τ .

¹¹ Ce critère est plus conservatif que celui de Green et collaborateurs (2014), qui proposent le même seuil de pente ($0,02\mu S/s$) mais sans retrancher la dérivée du signal lent. Nous avons ajouté cette contrainte car en situation de stress, le niveau de base (SCL) peut évoluer sur de grandes amplitudes et dépasser lui-même la pente absolue de $0,02\mu S/s$.

IV.A. Extraction d'une caractéristique standard

Une augmentation de la fréquence respiratoire a depuis longtemps été associée à une situation de stress (Tulen, Moleman, van Steenis, et al., 1989). On peut estimer celle-ci en cherchant la fréquence correspondant à l'amplitude maximale sur la DSP du signal respiratoire, pourvu que ce maximum se situe dans les fréquences naturelles des cycles respiratoires humains (Shouldice et al., 2010). La DSP du signal a été calculée à partir de la transformation de Fourier rapide (*fft*), puis son maximum a été recherché dans l'intervalle $[0,1 ; 2]$ Hz.

IV.B. Extraction de caractéristiques supplémentaires

On peut extraire d'autres caractéristiques potentiellement pertinentes à partir du signal brut. L'amplitude de la respiration, par exemple, est symptomatique d'un besoin accru de l'organisme en

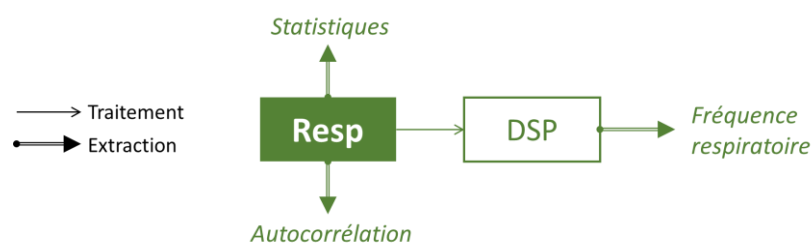


Figure 0.10 : Procédé d'extraction de caractéristiques sur le signal respiratoire

dioxygène. Celle-ci peut être estimée de manière statistique, par l'écart-type et l'amplitude totale des variations du signal respiratoire. Par ailleurs, le stress peut être amené à réguler ou à déréguler les cycles respiratoires (Gandhi et al., 2015), soit pour traduire une agitation, soit pour traduire une mobilisation des ressources vers une irrigation continue des tissus en O_2 . La régularité d'un signal peut être mesurée à partir de sa fonction d'autocorrélation, avec un délai proche de Zéro. On a donc calculé la fonction d'autocorrélation normalisée du signal respiratoire, et intégré la courbe entre les délais de 0s et 1s pour obtenir un indicateur de régularité (respAutocorr dans le Tableau 4.2). Le procédé complet d'extraction des caractéristiques respiratoires est illustré dans la Figure 0.10.

Annexe 6. Tests Statistiques

Dans cette annexe, on détaille les calculs et certains choix de paramètres directement utilisés dans le Chapitre Cinq - pour mesurer l'absence ou la présence d'un *effet* (i.e. médiane $\neq 0$) dans un échantillon de valeurs. Les notations adoptées dans cette partie correspondent à celles utilisées dans la littérature standard. Ainsi certains symboles que l'on trouvera ici (e.g., α) ne correspondent pas aux symboles du Chapitre Cinq -.

I. Le test des rangs signés de Wilcoxon

I.A. Statistique de test

Soit un échantillon de *données* $x_i, i \in \{1..N\}$, dont les valeurs sont théoriquement centrées sur 0. On souhaite mesurer le déséquilibre effectif de ces valeurs autour de 0.

On classe d'abord les exemples de l'échantillon selon leur valeur absolue ; on obtient ainsi un *rang* $r(|x_i|)$ pour chacun des exemples x_i de l'échantillon. A partir de ces rangs, on calcule la statistique de test W en sommant les rangs de toutes les valeurs positives :

$$W = \sum_{i=1}^N \text{sgn}(x_i) * r(|x_i|) \quad (0-4)$$

Où $\text{sgn}(x_i)$ vaut 1 si $x_i > 0$, et 0 sinon (Shieh et al., 2007).

I.B. Score Z et Taille d'Effet

Lorsque N est suffisamment grand, la statistique W suit une loi de probabilité normale, dont la moyenne μ_w et l'écart-type σ_w ne dépendent que de N (la « taille d'échantillon »). On peut ainsi normaliser W sous forme d'un score Z qui dépend de W et de N selon l'équation (0-5) ci-dessous. R est un terme correctif appliqué dans le cas où des mesures étaient égales lors de la phase de classement.

$$Z(W, N) = \frac{W - \mu_w(N)}{\sigma_w(N)/\sqrt{N}} = \frac{W - N(N + 1)/4}{\sqrt{[2N(N + 1)(2N + 1) - R]/24}} \quad (0-5)$$

De manière générale, la valeur absolue de Z augmente avec la proportion d'exemples de même signe au sein de l'échantillon. Mais elle augmente également avec la taille d'échantillon : d'après l'équation (0-5), $Z = O(N^{1/2})$. Le score Z affranchi de l'influence de la N donne la *taille d'effet* ϑ , que l'on utilise tout au long du Chapitre Cinq -. Celle-ci est calculée par l'équation (0-6) ci-dessous (Fritz et al., 2012)¹².

¹² Remarque : comme elle est la somme de valeurs bornées, la valeur de ϑ est elle-même bornée dans un intervalle centré sur 0. Les deux bornes de cet intervalle correspondent au cas limite où toutes les valeurs ont le même signe. La valeur maximale de $|\vartheta|$ est quasi indépendante de N , ainsi pour $N \in \{4 .. 80\}$: $-0,9 \leq \vartheta \leq 0,9$. Cette propriété distingue ϑ des tailles d'effet issues des tests paramétriques usuels ; mais à notre connaissance, la littérature scientifique ne considère pas cette distinction.

$$\theta(W) = \frac{Z(W, N)}{\sqrt{N}} \quad (0-6)$$

I.C. Démontrer la présence d'un effet

Soit un échantillon de données $\{x_i\}$: on cherche à montrer la *présence d'un effet* à l'intérieur de cet échantillon. Cela revient à rejeter l'hypothèse nulle H_0 , qui considère que la médiane de $\{x_i\}$ est nulle, au profit d'une hypothèse alternative H_A , qui considère que la médiane de $\{x_i\}$ n'est pas nulle.

Lorsqu'on projette ces hypothèses sur la caractéristique W du test de Wilcoxon, réduite au score Z de l'équation (0-5), on peut les écrire de la manière suivante :

$$\begin{cases} H_0: Z(W, N) = 0 \\ H_A: Z(W, N) \neq 0 \end{cases} \quad (0-7)$$

On considère que l'on peut rejeter l'hypothèse nulle H_0 (ce qui revient à dire que H_A est vraie) lorsque la valeur absolue de Z est supérieure à une valeur seuil que l'on donne dans l'équation (0-8) ci-dessous.

$$|Z(W, N)| > Z_{1-\frac{\alpha}{2}} = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \quad (0-8)$$

Ici:

- α est un paramètre que l'on aura choisi à l'avance. Ici, α représente le risque d'erreur de type I, c'est-à-dire le risque de rejeter à tort l'hypothèse nulle H_0 au profit de l'hypothèse alternative H_A .
- Φ est la densité cumulée de la loi Normale $\mathcal{N}(0,1)$: pour $x \in \mathbb{R}$, $\Phi(x)$ est l'aire sous la courbe de la distribution gaussienne de moyenne 0 et de variance 1 entre $-\infty$ et x .

On peut aussi raisonner directement sur la probabilité d'erreur, lorsqu'on projette la relation précédente (0-8) dans l'espace de α , à l'aide de la fonction Φ . On appelle généralement p l'image de Z par Φ :

$$p = \Phi(|Z(W, N)|) < \frac{\alpha}{2} \quad (0-9)$$

A partir de la valeur de Z mesurée dans l'équation (0-5), on peut ainsi directement calculer la probabilité p de rejeter à tort l'hypothèse nulle, c'est-à-dire la probabilité de se tromper lorsqu'on affirme qu'un *effet* est bien *présent* dans l'échantillon de données. On utilise généralement le seuil $\alpha=0,05$ ¹³ pour affirmer que l'effet en question est *statistiquement significatif*. Pour $\alpha=0,05$: $Z_{1-\alpha/2} \approx 1,96$. Ainsi, on rejettera l'hypothèse nulle si le score Z calculé est supérieur à +1,96 ou inférieur à -1,96.

Ce test dit « bilatéral » permet de démontrer la présence d'un effet *positif ou négatif* entre deux conditions. Lorsqu'on ne connaît pas la variation attendue de f , une valeur suffisamment faible de p nous permet de démontrer *qu'il existe une différence* entre les valeurs mesurées dans ces deux conditions. Par exemple, une modification du rythme cardiaque entre les conditions Stress et Contrôle.

¹³ Remarque : dans notre convention de calcul, dans le cas où $\alpha=0,05$ l'hypothèse nulle est rejetée pour H_0 : $p < 0,25$. On utilisera quand même la formule : « $p < 0,05$ » pour donner directement la valeur seuil α . En effet lorsqu'on effectue seulement un test statistique (celui de la section I.C), il est d'usage de multiplier p par 2 et de comparer directement p à α .

II. Le test d'équivalence

II.A. Démontrer l'absence d'un effet

En revanche, l'absence d'effet significatif ne permet pas de montrer que *cette différence n'existe pas*. Elle indique simplement qu'on manque de mesures pour prouver cette différence. Ainsi on ne pourra pas montrer, par exemple dans le Chapitre Cinq - , que la condition Contrôle n'a pas généré de stress, ou que le rythme cardiaque ne dépend pas du contexte.

Il faut pour cela transformer le calcul précédent en *test d'équivalence* (Lakens et al., 2018). Il s'agit d'une variante du test bilatéral d'égalité que l'on a décrit plus haut. Parmi les différentes manières utilisées pour implémenter un test d'équivalence, le « double test unilatéral » (en anglais : *Two One-Sided Test*, ou **TOST**) est l'une des plus fréquentes. Sa logique consiste à démontrer que la médiane de notre échantillon n'est *ni inférieure, ni supérieure* à un certain nombre ϵ qui est « proche » de 0. Ce test a déjà fait ses preuves en sciences cliniques et biologiques ; il est utilisé dans un nombre croissant d'études impliquant les tests statistiques paramétriques.

A notre connaissance, il n'existe pas de manière standard d'implémenter le TOST sur la base d'un test non-paramétrique. Certaines méthodes proposées utilisent des hypothèses sur la distribution des données de l'échantillon (Wellek, 1996). D'autres ont appliqué le test de Wilcoxon unilatéral en retranchant la valeur $\pm\epsilon$ aux valeurs de l'échantillon (Mara & Cribbie, 2012). Ces méthodes sont incompatibles avec notre objectif, qui est de dégager un critère d'absence d'effet pour une caractéristique quelconque, sans hypothèse sur sa distribution ni même sur ses valeurs. Pour cette raison, on appliquera le test de Wilcoxon unilatéral directement sur la valeur $Z(W,N)$, en raisonnant sur la taille d'effet ϑ . On transpose ainsi les principes du TOST généralement appliqués à la statistique T et à la taille d'effet d issues d'un test de Student (Lakens, 2017). Dans les calculs qui vont suivre, on suppose (comme dans le premier paragraphe) que $Z(W,N)$ est distribuée normalement suivant $\mathcal{N}(\mu_w, \sigma_w)$, ce qui est une approximation de la valeur réelle de Z pour une caractéristique quelconque.

II.B. TOST : Principes

Reprenons les hypothèses de l'équation (0-7), qui concernent le test bilatéral classique, exprimées cette fois en taille d'effet. On montrait l'existence d'un effet lorsqu'on rejetait l'hypothèse nulle c'est-à-dire une taille d'effet proche de 0. Rejeter l'hypothèse nulle H_0 revenait à valider l'une *ou* l'autre des hypothèses alternatives H_A^+ ou H_A^- , exprimées dans l'équation suivante :

$$\begin{cases} H_0: & \theta(W) = 0 \\ H_A: & \underbrace{[\theta(W) < 0]}_{H_A^-} \text{ OU } \underbrace{[\theta(W) > 0]}_{H_A^+} \end{cases} \quad (0-10)$$

On validait H_A^+ *ou* H_A^- en vérifiant que le score Z correspondant à la taille d'effet ϑ , était inférieur *ou* supérieur à un seuil qui dépendait de α : $[(\theta(W) * \sqrt{N}) < -Z_{1-\alpha/2}] \text{ ou } [(\theta(W) * \sqrt{N}) > Z_{1-\alpha/2}]$.

A l'inverse, pour montrer *l'absence* d'un effet, on voudra *rejeter* l'hypothèse que la taille d'effet soit inférieure *ou* supérieure à un certain seuil, que l'on aura fixé à l'avance. Soit un seuil fixé sur la taille d'effet : $\vartheta_{eq} > 0$. On souhaite rejeter l'hypothèse nulle H_0 , qui considère : $|\vartheta| > \vartheta_{eq}$. Ceci revient à valider l'hypothèse H_A , qui considère : $|\vartheta| < \vartheta_{eq}$, c'est-à-dire : « ϑ est proche de 0 ». Ces hypothèses ont chacune deux formes simples, exprimées dans l'équation suivante :

$$\begin{cases} H_0: \underbrace{[\theta(W) < -\theta_{eq}]}_{H_0^-} \text{ OU } \underbrace{[\theta(W) > \theta_{eq}]}_{H_0^+} \\ H_A: \underbrace{[\theta(W) > -\theta_{eq}]}_{H_A^-} \text{ ET } \underbrace{[\theta(W) < \theta_{eq}]}_{H_A^+} \end{cases} \quad (0-11)$$

Où $\{-\vartheta_{eq}; +\vartheta_{eq}\}$ sont les **limites d'équivalence**, exprimées en unité de taille d'effet¹⁴.

Ainsi, montrer « l'équivalence » consiste à montrer que la taille d'effet de notre échantillon est comprise dans un intervalle d'équivalence : $[-\vartheta_{eq}; +\vartheta_{eq}]$, que l'on aura déterminé à l'avance (voir comment dans la section III). La Figure 0.11 ci-contre est issue de (Lakens et al., 2018). Elle illustre comparativement un test bilatéral (décrit en section I.C) et un test d'équivalence. La zone où ϑ (« effect size ») s'inscrit dans l'hypothèse nulle (« H_0 ») est représentée en couleur. La zone où ϑ s'inscrit dans l'hypothèse alternative (« H_1 ») est représentée en blanc. $\{\Delta_L; \Delta_U\}$ représentent des limites d'équivalence arbitraires.

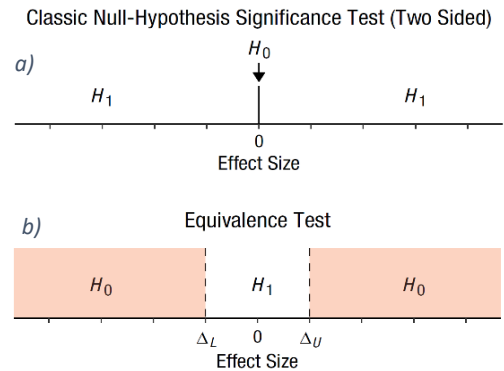


Figure 0.11 : Illustration d'un test bilatéral (a) et d'un test d'équivalence (b). Depuis (Lakens, 2018)

Pour montrer l'équivalence en pratique, il faut donc rejeter les deux formes simples de l'hypothèse nulle :

- a. D'une part, l'hypothèse $H_0^+ : (\theta(W, N) - \theta_{eq}) > 0$, avec une probabilité d'erreur inférieure à α .
- b. D'autre part, l'hypothèse $H_0^- : (\theta(W, N) + \theta_{eq}) < 0$, avec une probabilité d'erreur inférieure à α .

Cela revient à montrer cette double relation sur le score Z, adaptée de Lakens (2017) :

$$\begin{cases} Z^+ = (Z(W, N) - \theta_{eq} * \sqrt{N}) < (-Z_{1-\alpha}) \\ Z^- = (Z(W, N) + \theta_{eq} * \sqrt{N}) > (+Z_{1-\alpha}) \end{cases} \quad (0-12)$$

Dans cette formule, comme dans l'équation (0-8) pour le test bilatéral, les seuils $Z_{1-\alpha}$ et ϑ_{eq} sont fixés à l'avance et indépendants de la taille d'échantillon N. On pose généralement : $Z_{1-\alpha} \approx 1,64$ pour $\alpha = 0,05$. De même, la valeur de ϑ_{eq} peut être fixée à l'avance si l'on veut montrer que la taille d'effet ϑ a 90% de chances de se situer dans l'intervalle $[-0.5; +0.5]$. Mais dans cette approche subjective, la probabilité de montrer une équivalence dépendra fortement de la taille de l'échantillon comme le montre le facteur \sqrt{N} dans l'équation (0-12). Selon (Lakens et al., 2018), une limite d'équivalence $\vartheta_{eq}=0,5$ serait susceptible de rejeter la présence d'un effet dans 25% à 33% des cas seulement. Cette limite doit évoluer en fonction de la *puissance statistique* du test, c'est-à-dire sa probabilité de valider l'hypothèse alternative lorsque celle-ci est vraie.

Par ailleurs, pour le test d'équivalence comme pour le test bilatéral, la probabilité de démontrer l'absence ou la présence d'effet est fortement dépendante de la taille de l'échantillon N, car dans les deux cas $Z = O(N^{1/2})$ (voir I.B). Dans le Chapitre Cinq -, il est question de *démontrer* puis de *comparer* l'absence ou la présence d'un effet pour des échantillons de taille variable ($0 < N \leq 80$). Pour que les traitements statistiques appliqués à nos données soient homogènes entre différentes tailles d'échantillon, nous devons adapter d'adapter ϑ_{eq} et α en fonction du nombre d'échantillons.

¹⁴ Dans le cas général, elles ne sont pas nécessairement symétriques par rapport à Zéro : on aura alors deux bornes indépendantes, θ_{inf} et θ_{sup}

III. Gestion d'une taille d'échantillon quelconque

III.A. Erreur de type II et puissance statistique

Pour un test statistique quelconque, on appelle « erreur de type II » (β) la probabilité de rejeter à tort l'hypothèse alternative. A l'inverse, la probabilité de montrer la présence d'un effet lorsque cet effet existe (test bilatéral), ou son absence lorsqu'il n'existe pas (test d'équivalence) est appelée « puissance statistique » du test. Cette probabilité, logiquement égale à $1-\beta$, dépend fortement des paramètres évoqués dans le paragraphe précédent. La relation entre N , α et β , ϑ et éventuellement ϑ_{eq} est connue pour le test bilatéral comme pour le test d'équivalence (Chow et al., 2002; Julious, 2004). La puissance statistique d'un test diminue lorsqu'on diminue le seuil α sur l'erreur de type I : moins on a de chances de rejeter l'hypothèse nulle, plus on a de chances de rejeter l'hypothèse alternative. La puissance statistique augmente également la taille de l'échantillon (N). Enfin :

- Pour le test bilatéral, la puissance statistique augmente avec la taille d'effet mesurée $|\vartheta|$ (pourvu que H_A soit vraie, ce qui implique : $|\vartheta| > 0$).
- Pour le test d'équivalence, la puissance statistique augmente lorsque $|\vartheta|$ diminue et lorsque la limite d'équivalence ϑ_{eq} augmente (pourvu que H_A soit vraie, ce qui implique : $|\vartheta| < \vartheta_{eq}$).

Pour chacun des deux tests, on peut calculer le nombre N d'échantillons nécessaires pour rejeter l'hypothèse nulle, avec un seuil d'erreur α et une puissance statistique $1-\beta$ (Chow et al., 2002; Julious, 2004). Dans la présente analyse, ce nombre est donné par les formules suivantes¹⁵ :

$$\left\{ \begin{array}{l} \text{Test bilatéral :} \\ \text{Test d'équivalence :} \end{array} \right. \quad \begin{array}{l} N(\theta) = \frac{\left(Z_{1-\frac{\alpha}{2}} + Z_{1-\beta} \right)^2}{\theta^2} \\ N(\theta, \theta_{eq}) = \frac{\left(Z_{1-\alpha} + Z_{1-\frac{\beta}{2}} \right)^2}{(\theta_{eq} - |\theta|)^2} \end{array} \quad (0-13)$$

Avec toujours : $|\vartheta| > 0$ pour le test bilatéral, $0 \leq |\vartheta| < \vartheta_{eq}$ pour le test d'équivalence.

III.B. Adaptation des seuils

Pour construire un test statistique valide, on cherche souvent une puissance statistique d'au moins 0.8 – soit 80% de chances de démontrer l'absence ou la présence d'un effet (Julious, 2004). Comme notre base de données (Chapitre Quatre -) comporte un faible nombre d'exemples, on ne cherchera pas à augmenter ce chiffre. Notre objectif ici est que la puissance statistique reste stable, même lorsque N varie fortement entre deux caractéristiques ($0 < N \leq 80$). A partir de l'équation ci-dessus, on souhaite ainsi trouver des *ordres de grandeur* sur la valeur de N qui détermineront les futures valeurs de α ; de sorte qu'à taille d'effet constante, la puissance statistique reste proche de 80%.

¹⁵ A nouveau, ces formules adaptées de (Chow et al., 2002) sont des approximations. Le calcul de la puissance statistique réelle d'un test non-paramétrique, en fonction du nombre N d'exemples, doit faire des hypothèses sur la distribution de l'échantillon soumis au test (Shieh et al., 2007).

La Figure 0.12 ci-dessous présente une estimation du nombre N présenté dans les formules de l'équation (0-13), en fonction des tailles d'effet, pour différentes valeurs de α , en supposant une puissance statistique égale à 80%. Le graphique de gauche affiche N en fonction de ϑ pour le test bilatéral. Le graphique de droite affiche N fonction de la limite d'équivalence, lorsque $\vartheta=0$ (c'est-à-dire en l'absence totale d'effet). Pour éviter la confusion sur ces graphiques entre ϑ_{eq} et ϑ , on a temporairement renommé « Δ_{eq} » cette limite d'équivalence.

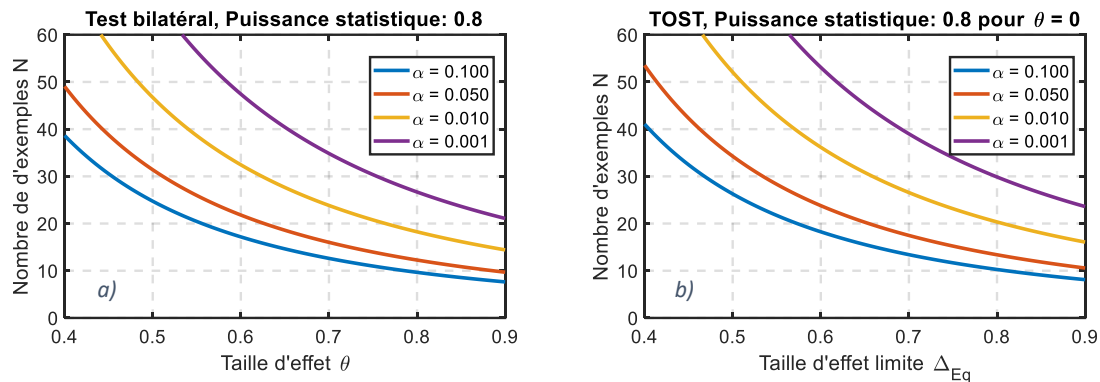


Figure 0.12 : Estimation du nombre d'exemples standard pour démontrer la présence d'un effet en fonction de la taille d'effet ϑ (a), ou l'absence d'effet en fonction de la limite d'équivalence Δ_{eq} lorsque $\vartheta=0$ (b).

Sur le graphique de gauche¹⁶, on peut lire qu'il faut environ $N \sim 50$ échantillons pour montrer l'existence d'un effet : (i) avec une probabilité d'erreur de 5% ($\alpha=0,05$), (ii) lorsque la taille d'effet mesurée ϑ est de l'ordre de 0,4. Mais pour une taille d'effet ϑ de l'ordre de 0,6, avec le même seuil α , le nombre d'échantillons descend à $N \sim 20$. Parallèlement lorsque $N \sim 50$, cette même taille d'effet ($\vartheta \sim 0,6$) nous permet également de montrer la présence d'un effet avec $\alpha=0,001$.

Les chiffres sont relativement similaires lorsqu'on cherche une valeur « idéale » de Δ_{eq} . Pour démontrer l'absence d'effet, on suppose désormais une taille d'effet nulle : $\vartheta=0$. Sur le graphique de droite, on peut lire que lorsque la limite d'équivalence Δ_{eq} est de l'ordre de 0,4, on peut montrer l'absence d'un effet : (i) pour $\alpha=0,05$, (ii) avec $N \sim 50$ échantillons. Mais lorsque Δ_{eq} est de l'ordre de 0,6, avec le même seuil α , le nombre d'échantillons descend à $N \sim 20$. Parallèlement lorsque $N \sim 50$, cette même limite ($\Delta_{eq} \sim 0,6$) nous permet de montrer la présence d'un effet avec $\alpha=0,001$.

Afin de garantir une puissance statistique proche des 80%, on peut ainsi fixer la limite d'équivalence Δ_{eq} à **0.6** pour montrer l'absence d'un effet ; et prévoir, à l'inverse, que le test bilatéral montrera la présence d'un effet pour une taille d'effet ϑ mesurée à **0.6**. Cela permet de construire les intervalles de taille d'échantillon N qui étalonneront les seuils α – c'est-à-dire les valeurs de p_{seuil} dans le Chapitre Cinq - II.B. Cet étalonnage est proposé dans le Tableau 0.7 ci-dessous.

Tableau 0.7 : Etalonnement de α en fonction du nombre N d'exemples disponibles

Valeur de N	$0 < N \leq 15$	$15 < N < 25$	$25 < N \leq 45$	$N > 45$
Valeur de α ¹⁷	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.001$

¹⁶ Malgré nos approximations, cette estimation de $N(\theta)$ pour $\{\alpha=0,5 ; \beta=0,2\}$ est cohérente avec celles de (Shieh et al., 2007). Cependant, l'estimation ne prend pas en compte que la taille d'effet est ici bornée.

¹⁷ La valeur $\alpha = 0,1$ est rarement utilisée dans la littérature scientifique. Elle est proposée ici pour que le critère de sélection d'une caractéristique existe pour les petites valeurs de N . Les conclusions du test pourront toujours être rejetées sur la base du nombre effectif d'exemples disponibles.

Annexe 7. Résultats de la Sélection des Caractéristiques

a) Sélection des caractéristiques sur les tâches valides

Les tableaux qui suivent complètent ceux du Chapitre Cinq - . Pour commencer, le Tableau 0.8 ci-dessous présente les tailles d'effet mesurées pour nos 32 caractéristiques pour les tâches {TSST + TA}.

Tableau 0.8 : Tailles d'effet mesurées pour les 32 caractéristiques de l'ensemble de départ sur les tâches TSST et TA

TSST + TA	Tâche	Contrainte	Pression	Activité	Contexte
Attendu	*** ϑ_T [N _T]	*** ϑ_K [N _K]	*** ϑ_P [N _P]	ooo ϑ_A [N _A]	ooo ϑ_E [N _E]
<i>ibiHF</i>	ooo +0.08 [40]	ooo +0.00 [40]	*__ +0.35 [40]	o__ -0.26 [40]	ooo -0.14 [40]
<i>ibiHFnorm</i>	*** -0.55 [40]	ooo +0.05 [40]	ooo -0.08 [40]	*__ -0.37 [40]	ooo +0.23 [40]
<i>ibiLF/HF</i>	*** +0.61 [40]	o__ -0.28 [40]	ooo -0.03 [40]	*** +0.46 [40]	ooo -0.23 [40]
<i>ibiLF</i>	*** +0.51 [40]	ooo -0.17 [40]	ooo +0.17 [40]	*** +0.42 [40]	*__ -0.35 [40]
<i>ibiLFnorm</i>	*** +0.41 [40]	*__ -0.34 [40]	o__ -0.25 [40]	*** +0.67 [40]	ooo +0.02 [40]
<i>hrMean</i>	*** +0.85 [40]	*** +0.71 [40]	*** +0.72 [40]	*** +0.66 [40]	ooo +0.13 [40]
<i>hrRmssd</i>	*** +0.74 [40]	*** +0.51 [40]	*** +0.51 [40]	*__ +0.30 [40]	ooo -0.18 [40]
<i>ibiRmssd</i>	ooo -0.14 [40]	ooo -0.18 [40]	ooo +0.16 [40]	ooo -0.16 [40]	ooo -0.17 [40]
<i>hrStd</i>	*** +0.86 [40]	*** +0.72 [40]	*** +0.78 [40]	*__ +0.32 [40]	*** -0.46 [40]
<i>respAutocorr</i>	*** +0.81 [22]	*** +0.76 [22]	*** +0.65 [22]	*** +0.42 [22]	ooo +0.20 [22]
<i>respRange</i>	*** +0.72 [22]	ooo +0.20 [22]	*__ +0.38 [22]	*** +0.73 [22]	ooo -0.16 [22]
<i>respRate</i>	*** -0.83 [22]	ooo -0.22 [22]	ooo -0.22 [22]	*** -0.64 [22]	ooo +0.16 [22]
<i>respStd</i>	*** +0.81 [22]	*__ +0.36 [22]	*** +0.48 [22]	*** +0.80 [22]	ooo -0.20 [22]
<i>sclDerMean</i>	*** -0.44 [40]	ooo -0.15 [40]	o__ -0.30 [40]	ooo -0.03 [40]	*** +0.47 [40]
<i>sclDerRMS</i>	*** +0.77 [40]	*__ +0.34 [40]	*** +0.46 [40]	*** +0.46 [40]	*__ -0.35 [40]
<i>sclMean</i>	*** +0.68 [40]	ooo +0.09 [40]	o__ +0.24 [40]	*** +0.74 [40]	ooo -0.10 [40]
<i>scrAmpP50</i>	*** +0.75 [40]	ooo +0.16 [40]	o__ +0.23 [40]	*** +0.64 [40]	o__ -0.24 [40]
<i>scrAmpP85</i>	*** +0.51 [40]	ooo -0.07 [40]	ooo +0.19 [40]	*__ +0.35 [40]	ooo -0.19 [40]
<i>scrAmpStd</i>	*** +0.49 [40]	ooo -0.11 [40]	ooo +0.13 [40]	*__ +0.33 [40]	o__ -0.28 [40]
<i>scrArea</i>	*** +0.67 [40]	ooo +0.03 [40]	ooo +0.09 [40]	*** +0.75 [40]	ooo -0.05 [40]
<i>scrFreq</i>	*** +0.74 [40]	ooo -0.01 [40]	ooo -0.03 [40]	*** +0.82 [40]	ooo +0.09 [40]
<i>scrGreatFreq</i>	*** +0.56 [40]	ooo -0.13 [40]	ooo -0.09 [40]	*** +0.71 [40]	o__ -0.25 [40]
<i>scrMildFreq</i>	*** +0.59 [40]	ooo +0.23 [40]	ooo +0.18 [40]	*** +0.63 [40]	ooo +0.04 [40]
<i>scrRecovMean</i>	ooo +0.02 [38]	ooo -0.17 [39]	o__ -0.22 [35]	ooo +0.04 [37]	ooo +0.06 [36]
<i>scrRecovStd</i>	o__ +0.30 [38]	ooo -0.05 [39]	ooo -0.14 [35]	o__ +0.23 [37]	ooo +0.14 [36]
<i>scrRiseMean</i>	*__ -0.35 [38]	*** +0.42 [39]	ooo -0.09 [35]	ooo -0.10 [37]	*__ +0.39 [36]
<i>scrRiseStd</i>	o__ +0.25 [38]	*__ +0.31 [39]	ooo +0.06 [35]	ooo -0.08 [37]	ooo +0.10 [36]
<i>scrrP50</i>	*** +0.59 [36]	ooo -0.09 [38]	ooo +0.12 [32]	*** +0.68 [34]	ooo -0.09 [34]
<i>scrrP85</i>	*** +0.68 [36]	ooo +0.03 [38]	o__ +0.28 [32]	*** +0.56 [34]	o__ -0.23 [34]
<i>scrrStd</i>	*** +0.62 [36]	ooo +0.06 [38]	*__ +0.31 [32]	o__ +0.28 [34]	ooo -0.13 [34]
<i>scrStrongFreq</i>	*** +0.71 [40]	ooo -0.12 [40]	ooo -0.10 [40]	*** +0.80 [40]	ooo +0.01 [40]
<i>scrWeakFreq</i>	ooo -0.11 [40]	ooo +0.17 [40]	ooo -0.12 [40]	ooo -0.02 [40]	*__ +0.36 [40]

Les caractéristiques en vert ont été sélectionnées sur les critères proposés en section II.C, traduits dans la 1^{ère} ligne Attendu. Les caractéristiques en rouge correspondent aux « standards » qui n'ont pas été sélectionnées. Pour plus de précisions sur les caractéristiques, voir Tableau 4.2. Dans chaque cellule : Critère validé | Taille d'effet | [Nombre d'exemples].
 '***': forte présence d'effet – '*__': faible présence d'effet – 'o__': faible absence d'effet – 'ooo': forte absence d'effet.

b) Congruence inter-tâches

Au Chapitre Cinq - III.B.2, on distingue par tâche les tailles d'effet mesurées pour les caractéristiques pré-sélectionnées sur les tâches TSST et TA. Pour chaque tâche $i \in \{TSST, TA, D2, SECPT\}$, les tableaux ci-dessous détaillent les tailles d'effet et le nombre de sujets relatifs aux symboles Tableau 5.7. Un tableau correspond à un estimateur.

Tableau 0.9 : Effet Tâche pour chaque tâche du protocole expérimental sur les caractéristiques pré-sélectionnées

Effet Tâche	TSST	TA	D2	SECPT
EVAS	*** +0.69 [20]	*** +0.68 [20]	*** +0.47 [20]	o__ +0.29 [20]
<i>hrMean</i>	*** +0.88 [20]	*** +0.79 [20]	*** +0.65 [19]	*** +0.68 [19]
<i>hrStd</i>	*** +0.87 [20]	*** +0.85 [20]	o__ -0.24 [19]	*** +0.80 [19]
<i>hrRmssd</i>	*** +0.68 [20]	*** +0.80 [20]	o__ +0.23 [19]	*** +0.60 [19]
<i>respAutocorr</i>	*** +0.80 [11]	*** +0.86 [11]	*** -0.56 [11]	ooo +0.19 [11]
<i>respStd</i>	*** +0.86 [11]	*** +0.72 [11]	*__ +0.40 [11]	o__ +0.24 [11]
<i>sclDerRMS</i>	*** +0.81 [20]	*** +0.73 [20]	*** +0.68 [20]	*** +0.83 [19]

Dans chaque cellule : Critère validé | Taille d'effet | [Nombre d'exemples].

'***': forte présence d'effet – '*__': faible présence d'effet – 'o__': faible absence d'effet – 'ooo': forte absence d'effet.

Tableau 0.10 : Effet Contrainte pour chaque tâche du protocole expérimental sur les caractéristiques pré-sélectionnées

Effet Contrainte	TSST	TA	D2	SECPT
EVAS	*** +0.83 [20]	*** +0.82 [20]	*** +0.53 [20]	*** +0.58 [20]
<i>hrMean</i>	*** +0.86 [20]	*** +0.49 [20]	ooo +0.13 [20]	*** +0.84 [19]
<i>hrStd</i>	*** +0.83 [20]	*** +0.59 [20]	ooo +0.09 [20]	o__ +0.25 [19]
<i>hrRmssd</i>	*** +0.47 [20]	*** +0.57 [20]	ooo +0.05 [20]	o__ -0.30 [19]
<i>respAutocorr</i>	*** +0.86 [11]	*** +0.70 [11]	ooo -0.13 [11]	*__ +0.46 [11]
<i>respStd</i>	*** +0.64 [11]	ooo +0.03 [11]	*__ -0.43 [11]	ooo +0.13 [11]
<i>sclDerRMS</i>	*** +0.47 [20]	ooo +0.18 [20]	ooo +0.04 [20]	*__ +0.38 [19]

Dans chaque cellule : Critère validé | Taille d'effet | [Nombre d'exemples].

'***': forte présence d'effet – '*__': faible présence d'effet – 'o__': faible absence d'effet – 'ooo': forte absence d'effet.

Tableau 0.11 : Effet Pression pour chaque tâche du protocole expérimental sur les caractéristiques pré-sélectionnées

Effet Pression	TSST	TA	D2	SECPT
EVAS	*** +0.68 [20]	*** +0.67 [20]	o__ +0.30 [20]	*** +0.60 [20]
<i>hrMean</i>	*** +0.88 [20]	*__ +0.41 [20]	*__ +0.41 [19]	*** +0.78 [19]
<i>hrStd</i>	*** +0.85 [20]	*** +0.68 [20]	*__ +0.36 [19]	*** +0.62 [19]
<i>hrRmssd</i>	*** +0.49 [20]	*** +0.56 [20]	ooo +0.22 [19]	o__ +0.23 [19]
<i>respAutocorr</i>	*** +0.72 [11]	*** +0.62 [11]	ooo -0.11 [11]	ooo +0.16 [11]
<i>respStd</i>	*** +0.86 [11]	ooo +0.08 [11]	*__ +0.48 [11]	*__ +0.46 [11]
<i>sclDerRMS</i>	*** +0.57 [20]	*__ +0.35 [20]	ooo +0.19 [20]	*** +0.52 [19]

Dans chaque cellule : Critère validé | Taille d'effet | [Nombre d'exemples].

'***': forte présence d'effet – '*__': faible présence d'effet – 'o__': faible absence d'effet – 'ooo': forte absence d'effet.

Tableau 0.12 : Effet Activité pour chaque tâche du protocole expérimental sur les caractéristiques pré-sélectionnées

Effet Activité	TSST	TA	D2	SECPT
EVAS	ooo -0.17 [20]	ooo -0.16 [20]	*** +0.65 [20]	*** -0.51 [20]
<i>hrMean</i>	*** +0.68 [20]	*** +0.63 [20]	*** +0.53 [20]	ooo -0.11 [20]
<i>hrStd</i>	ooo +0.13 [20]	*** +0.48 [20]	*** -0.55 [20]	ooo -0.14 [20]
<i>hrRmssd</i>	ooo +0.21 [20]	* _ +0.39 [20]	o _ -0.29 [20]	* _ +0.33 [20]
<i>respAutocorr</i>	* _ +0.40 [11]	* _ +0.48 [11]	o _ -0.29 [11]	ooo +0.21 [11]
<i>respStd</i>	*** +0.78 [11]	*** +0.86 [11]	o _ +0.27 [11]	* _ -0.40 [11]
<i>sclDerRMS</i>	*** +0.44 [20]	*** +0.48 [20]	*** +0.48 [20]	* _ +0.33 [20]

Dans chaque cellule : Critère validé | Taille d'effet | [Nombre d'exemples].

'***': forte présence d'effet – '* _': faible présence d'effet – 'o _': faible absence d'effet – 'ooo': forte absence d'effet.

c) Caractéristiques standard rejetées par nos critères

Au Chapitre Cinq - III.C, on examine enfin les résultats des caractéristiques « standard » non sélectionnées dans la section III.B. Les tableaux ci-dessous détaillent les tailles d'effet et le nombre de sujets relatifs aux symboles du Tableau 5.8. Les résultats sont regroupés par estimateur : Tâche (T), Contrainte (K), Pression (P), Activité (A) et Contexte (E).

Tableau 0.13 : Effet Tâche pour les tâches expérimentales TSST et TA sur les caractéristiques standard non sélectionnées

Effet Tâche	TSST	TA	TSST + TA
Attendu	*** ϑ [N]	*** ϑ [N]	*** ϑ [N]
<i>ibiLF</i>	* _ +0.40 [20]	*** +0.62 [20]	*** +0.51 [40]
<i>ibiLFnorm</i>	* _ +0.33 [20]	*** +0.46 [20]	*** +0.41 [40]
<i>ibiHF</i>	o _ -0.25 [20]	*** +0.46 [20]	ooo +0.08 [40]
<i>ibiHFnorm</i>	*** -0.61 [20]	*** -0.46 [20]	*** -0.55 [40]
<i>ibiLF/HF</i>	*** +0.58 [20]	*** +0.65 [20]	*** +0.61 [40]
<i>ibiRmssd</i>	* _ -0.37 [20]	ooo +0.11 [20]	ooo -0.14 [40]
<i>respRate</i>	*** -0.88 [11]	*** -0.80 [11]	*** -0.83 [22]
<i>sclMean</i>	*** +0.65 [20]	*** +0.73 [20]	*** +0.68 [40]
<i>scrAmpP50</i>	*** +0.76 [20]	*** +0.73 [20]	*** +0.75 [40]
<i>scrFreq</i>	*** +0.79 [20]	*** +0.68 [20]	*** +0.74 [40]
<i>scrP85</i>	*** +0.70 [18]	*** +0.64 [18]	*** +0.68 [36]

Dans chaque cellule : Critère validé | Taille d'effet | [Nombre d'exemples].

'***': forte présence d'effet – '* _': faible présence d'effet – 'o _': faible absence d'effet – 'ooo': forte absence d'effet.

Tableau 0.14 : Effet Pression pour les tâches expérimentales TSST et TA sur les caractéristiques standard non sélectionnées

Effet Pression	TSST	TA	TSST + TA
Attendu	*** ϑ [N]	*** ϑ [N]	*** ϑ [N]
<i>ibiLF</i>	ooo +0.17 [20]	ooo +0.16 [20]	ooo +0.17 [40]
<i>ibiLFnorm</i>	ooo -0.21 [20]	* _ -0.33 [20]	o _ -0.25 [40]
<i>ibiHF</i>	ooo +0.15 [20]	*** +0.48 [20]	* _ +0.35 [40]
<i>ibiHFnorm</i>	ooo -0.23 [20]	ooo +0.09 [20]	ooo -0.08 [40]
<i>ibiLF/HF</i>	o _ +0.26 [20]	o _ -0.25 [20]	ooo -0.03 [40]
<i>ibiRmssd</i>	ooo +0.03 [20]	* _ +0.37 [20]	ooo +0.16 [40]

Effet Pression	TSST	TA	TSST + TA
Attendu	*** ϑ [N]	*** ϑ [N]	*** ϑ [N]
respRate	*__ -0.40 [11]	ooo -0.05 [11]	ooo -0.22 [22]
sclMean	ooo +0.19 [20]	o__ +0.29 [20]	o__ +0.24 [40]
scrAmpP50	ooo +0.10 [20]	*__ +0.38 [20]	o__ +0.23 [40]
scrFreq	ooo -0.01 [20]	ooo -0.03 [20]	ooo -0.03 [40]
scrP85	*__ +0.31 [16]	o__ +0.21 [16]	o__ +0.28 [32]

Dans chaque cellule : Critère validé | Taille d'effet | [Nombre d'exemples].

'***': forte présence d'effet – '*__': faible présence d'effet – 'o__': faible absence d'effet – 'ooo': forte absence d'effet.

Tableau 0.15 : Effet Contrainte pour les tâches expérimentales TSST et TA sur les caractéristiques standard non sélectionnées

Effet Contrainte	TSST	TA	TSST + TA
Attendu	*** ϑ [N]	*** ϑ [N]	*** ϑ [N]
ibiLF	o__ -0.24 [20]	ooo -0.15 [20]	ooo -0.17 [40]
ibiLFnorm	*__ -0.30 [20]	*__ -0.34 [20]	*__ -0.34 [40]
ibiHF	ooo -0.17 [20]	ooo +0.16 [20]	ooo +0.00 [40]
ibiHFnorm	o__ -0.23 [20]	o__ +0.28 [20]	ooo +0.05 [40]
ibiLF/HF	ooo -0.02 [20]	*** -0.48 [20]	o__ -0.28 [40]
ibiRmssd	o__ -0.28 [20]	ooo -0.06 [20]	ooo -0.18 [40]
respRate	*__ -0.36 [11]	ooo -0.07 [11]	ooo -0.22 [22]
sclMean	ooo +0.05 [20]	ooo +0.13 [20]	ooo +0.09 [40]
scrAmpP50	ooo -0.02 [20]	*__ +0.31 [20]	ooo +0.16 [40]
scrFreq	ooo -0.02 [20]	ooo +0.01 [20]	ooo -0.01 [40]
scrP85	ooo -0.03 [19]	ooo +0.06 [19]	ooo +0.03 [38]

Dans chaque cellule : Critère validé | Taille d'effet | [Nombre d'exemples].

'***': forte présence d'effet – '*__': faible présence d'effet – 'o__': faible absence d'effet – 'ooo': forte absence d'effet.

Tableau 0.16 : Effet Activité pour les tâches expérimentales TSST et TA sur les caractéristiques standard non sélectionnées

Effet Activité	TSST	TA	TSST + TA
Attendu	ooo ϑ [N]	ooo ϑ [N]	ooo ϑ [N]
ibiLF	ooo +0.19 [20]	*** +0.61 [20]	*** +0.42 [40]
ibiLFnorm	*** +0.58 [20]	*** +0.74 [20]	*** +0.67 [40]
ibiHF	*__ -0.36 [20]	ooo -0.13 [20]	o__ -0.26 [40]
ibiHFnorm	*__ -0.33 [20]	*__ -0.41 [20]	*__ -0.37 [40]
ibiLF/HF	*__ +0.33 [20]	*** +0.58 [20]	*** +0.46 [40]
ibiRmssd	ooo -0.21 [20]	ooo -0.10 [20]	ooo -0.16 [40]
respRate	*__ -0.46 [11]	*** -0.84 [11]	*** -0.64 [22]
sclMean	*** +0.75 [20]	*** +0.73 [20]	*** +0.74 [40]
scrAmpP50	*** +0.71 [20]	*** +0.58 [20]	*** +0.64 [40]
scrFreq	*** +0.84 [20]	*** +0.79 [20]	*** +0.82 [40]
scrP85	*** +0.63 [17]	*** +0.52 [17]	*** +0.56 [34]

Dans chaque cellule : Critère validé | Taille d'effet | [Nombre d'exemples].

'***': forte présence d'effet – '*__': faible présence d'effet – 'o__': faible absence d'effet – 'ooo': forte absence d'effet.

Annexe 8. Pertinence des Estimateurs : Compléments d'Analyses

a) Problème adressé

Dans le Chapitre Cinq - II, on a proposé cinq *estimateurs* (Tâche **T**, Activité **A**, Pression **P**, Contexte **E**, Contrainte **K**) pour calculer la « part du stress » et celle des facteurs confondants dans la réponse d'une caractéristique f , au travers de nos quatre conditions expérimentales (**c** : tâche contrôle, **s** : tâche stressante, **rc** : repos avant contrôle, **rs** : repos avant stress).

$$\begin{cases} T = f_s - f_{rs} \\ A = f_c - f_{rc} \\ K = f_s - f_c \\ E = f_{rs} - f_{rc} \\ P = T - A = K - E \end{cases} \quad (0-14)$$

Le calcul de ces estimateurs est justifié par un modèle de l'évolution de f au travers des différentes conditions expérimentales (voir Chapitre Cinq - I.B). L'objectif de cette annexe est (i) de vérifier si les hypothèses de ce modèle sont bien cohérentes avec résultats présentés au Chapitre Cinq - III.B ; et (ii) de vérifier si les estimateurs de la « part du stress » (**P**, **K**) ne sont pas biaisés dans leur estimation.

b) Rappel du cadre théorique : des contributions cibles et contributions parasites

On redonne ci-dessous l'équation (5-4), qui exprime les cinq estimateurs en fonction des contributions de l'*environnement* (ε), de l'*activité* (α) et de la *pression* (π) dans chacune des conditions (S, C, Rc, Rs). Les composantes en **noir** proviennent du scénario dit « idéal ». Les composantes en **gris** proviennent du scénario dit « réaliste », où l'on considère que l'on ne peut garantir l'absence de charge affective dans les conditions de *non-stress* (C, Rc, Rs).

$$\begin{cases} T = \alpha + (\pi_s - \pi_{rs}) & = \alpha + \pi_t \\ A = \alpha + (\pi_c - \pi_{rc}) & = \alpha + \pi_a \\ K = (\pi_s - \pi_c) + (\varepsilon_{rs} - \varepsilon_{rc}) & = \pi_k + \varepsilon_e \\ E = (\varepsilon_{rs} - \varepsilon_{rc}) + (\pi_{rs} - \pi_{rc}) & = \varepsilon_e + \pi_e \\ P = (\pi_s - \pi_c) + (\pi_{rs} - \pi_{rc}) & = \pi_k - \pi_e = \pi_t - \pi_a \end{cases}$$

A droite de l'équation, on trouve des termes composites, comme la *pression* liée aux stressseurs (π_k), la *pression* liée à l'Activité (π_a), la *pression* liée au Contexte (π_e) ou la contribution *spontanée* du Contexte (ε_e). On peut exprimer chaque estimateur $Z \in \{T, A, K, E, P\}$ comme la somme d'une contribution « cible » : $x \in \{\pi_t, \alpha, \pi_k, \varepsilon_e\}$; et d'une contribution « parasite » : $y \in \{\alpha, \pi_a, \varepsilon_e, \pi_e\}$, la cible d'un estimateur pouvant être le parasite d'un autre (α, ε_e). Lorsque y est presque nulle (*i.e.*, $|x| \gg |y|$), on peut dire que Z est une mesure de x . Ainsi :

- L'estimateur Contrainte (**K**) mesure la *pression* liée aux stressseurs (π_k) lorsque $\varepsilon_e \approx 0$, c'est-à-dire lorsque la différence de contexte n'a pas d'influence *spontanée* sur la caractéristique f .
- L'estimateur Pression (**P**) mesure la *pression* liée aux stressseurs (π_k) lorsque $\pi_e \approx 0$, c'est-à-dire lorsque la différence de contexte n'a pas d'influence *affective* sur la personne.
- L'estimateur Activité (**A**) mesure la contribution de l'*activité* (α) lorsque $\pi_a \approx 0$, c'est-à-dire lorsque la charge affective varie peu entre Repos et Contrôle.
- Enfin, l'estimateur Contexte (**E**) mesure la contribution *spontanée* du contexte (ε_e) lorsque $\pi_e \approx 0$.

c) Relation théorique entre les tailles d'effets

En pratique, on ne peut pas accéder directement à la valeur des termes π_e et ε_e ; ni même montrer qu'ils sont faibles. En particulier : $\mathbf{E} = 0$ ne se signifie pas que \mathbf{P} et \mathbf{K} ne sont pas *biaisés* par π_e et ε_e . De même, on ne peut accéder à la valeur du terme résiduel π_e dans l'expression de \mathbf{A} . Pour montrer que l'approximation : $\mathbf{Z} \approx \mathbf{x}$ est vraie, il faut trouver un autre moyen pour appuyer l'hypothèse : $\gamma \approx 0$.

Soit un estimateur \mathbf{Z}_1 et une contribution γ quelconque, dont dépend l'estimateur Z_1 . Lorsque les valeurs de γ sont *non-nulles* et *de même signe* entre plusieurs tâches et plusieurs participants, la contribution γ participe à l'effet Z_1 mesuré par sa *taille d'effet* : ϑ_{Z_1} . Lorsque ce même terme γ est partagé par un autre estimateur \mathbf{Z}_2 , alors il participe également à l'effet Z_2 mesuré par ϑ_{Z_2} .

A l'échelle de plusieurs caractéristiques, γ introduit ainsi un lien de *corrélacion* entre les tailles d'effet ϑ_{Z_1} et ϑ_{Z_2} . Ainsi, l'existence ou non d'un lien de corrélacion entre deux tailles d'effet ϑ_{Z_1} et ϑ_{Z_2} permet de faire les hypothèses suivantes.

- (i) Si ϑ_{Z_1} et ϑ_{Z_2} sont *corrélées* entre elles, alors la composante γ commune à \mathbf{Z}_1 et \mathbf{Z}_2 est non-nulle et son signe est homogène entre les participants et les tâches. Si γ est une composante parasite de l'estimateur Z_1 , alors Z_1 est biaisé dans l'estimation de sa composante cible.
- (ii) Si ϑ_{Z_1} et ϑ_{Z_2} ne sont *pas corrélées* entre elles, alors soit la composante commune à \mathbf{Z}_1 et \mathbf{Z}_2 est nulle, soit son signe n'est pas homogène entre les participants et les tâches. A l'échelle interindividuelle, cette composante peut être considérée comme un « bruit » centré sur Zéro.

d) Relations empiriques entre les tailles d'effet

La Figure 0.13 (ci-après) représente la taille d'effet ϑ calculée pour un certain estimateur, en fonction de la taille d'effet ϑ d'un autre estimateur, pour les 32 caractéristiques extraites dans le Chapitre Quatre - II.B. Les tailles d'effet ϑ sont calculées pour l'ensemble des participants disponibles, dans la synthèse des tâches sélectionnées en section III.A : {TSST + TA}.

Le graphique (a) affiche l'effet Pression (ϑ_P) en fonction de l'effet Activité (ϑ_A). Les caractéristiques sélectionnées sont encadrées en noir. La « bande de rejet » (en rouge) correspond aux valeurs de l'effet Pression qui ne correspondent pas au critère de sélection proposé en II.C.2 : $|\vartheta_P| > 0,3$. Parmi les caractéristiques sélectionnées, on observe qu'un effet Pression fort et positif ($\vartheta_P > 0,3$) accompagne un effet Activité fort et positif ($\vartheta_A > 0,3$). En revanche, on n'observe pas de relation claire entre ϑ_A et ϑ_P pour les caractéristiques situées dans la bande de rejet ($|\vartheta_P| < 0,3$). Ainsi, le lien observé entre l'effet Pression et l'effet Activité semble concerner seulement les caractéristiques *sensibles au stress*.

De son côté, le graphique (b) affiche l'effet Tâche (ϑ_T) en fonction de l'effet Activité (ϑ_A). Comme dans le graphique (a), les caractéristiques sélectionnées (encadrées en noir) montrent de fortes valeurs de ϑ_T et de fortes valeurs de ϑ_A . Cela signifie que ces caractéristiques ont globalement varié *dans le même sens* durant la condition Stress et durant la condition Contrôle d'une même tâche. Sur le graphique (b), la zone grise repère les points (ϑ_A, ϑ_T) pour lesquelles l'effet Activité *prédomine* sur l'effet Tâche, c'est-à-dire que les points tels que la valeur ϑ_A est « plus forte » que la valeur ϑ_T et de même signe. On constate que le critère de sensibilité au stress ($|\vartheta_{P,K}| > 0,3$) a suffi à sélectionner des caractéristiques en-dehors de cette zone, et cela même si ϑ_A et ϑ_T n'apparaissent pas dans ce critère.

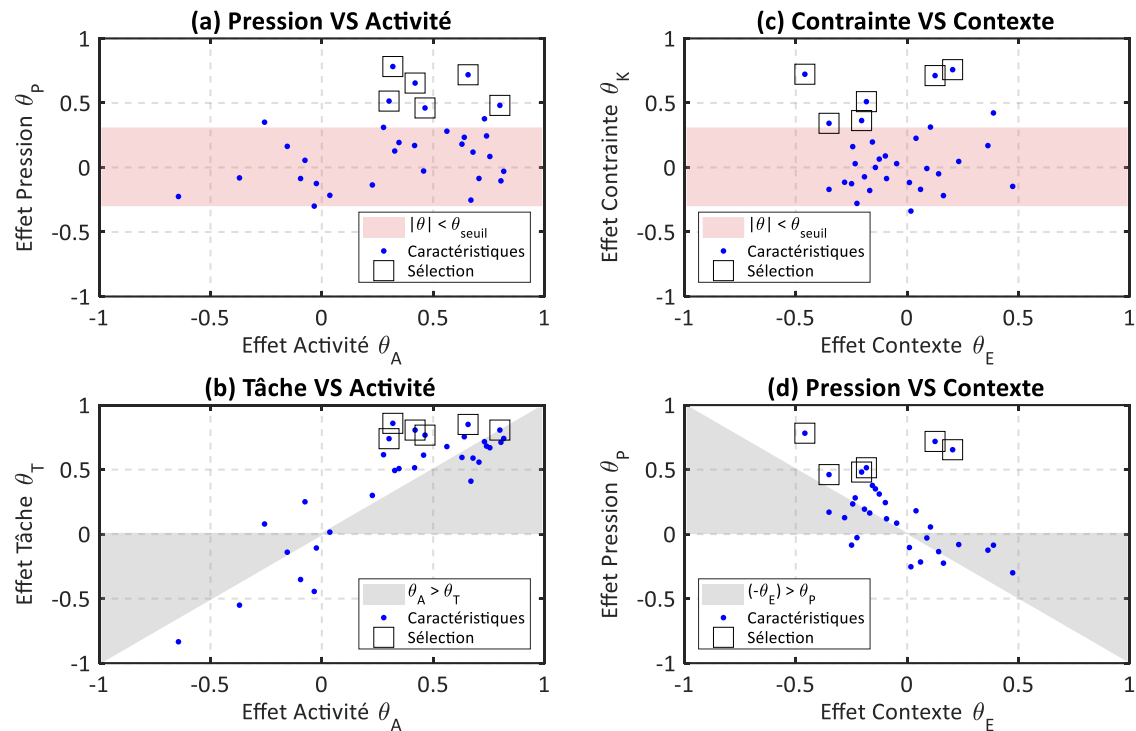


Figure 0.13 : Lien empirique entre les effets Tâche (ϑ_T), Activité (ϑ_A), Contrainte (ϑ_K), Contexte (ϑ_E) et Pression (ϑ_P).

Ainsi, des liens « empiriques » apparaissent entre les tailles d'effet de nos différents estimateurs. Sur le graphique (a), le lien observé entre l'effet Pression (ϑ_P) et l'effet Activité (ϑ_A) concerne uniquement les caractéristiques déterminées comme *sensibles* au stress. Sur le graphique (b), le lien observé entre l'effet Tâche (ϑ_T) et l'effet Activité (ϑ_A) peut être étendu à *toutes* les caractéristiques considérées dans cette étude : il prend la forme d'une *corrélation* nette entre ϑ_T et ϑ_A ($R=0,72$, $p<0,001$). Un phénomène similaire peut être observé sur le graphique (d), qui montre une *corrélation* négative ($R=-0,50$, $p=0,004$) entre l'effet Pression (ϑ_P) et l'effet Contexte (ϑ_E).

e) Ces relations ne sont pas seulement liées au mode de calcul des estimateurs

Pour expliquer ces résultats, on pourrait observer que nos cinq estimateurs (**T**, **A**, **K**, **E**, **P**) dépendent l'un de l'autre car ils sont calculés sur seulement quatre mesures f_j ($j \in \{rc, rs, c, s\}$) pour une tâche et un sujet donnés (voir équation (0-14) en début d'annexe). Ces relations « dans le calcul » (*i.e.*, à l'échelle d'une tâche et d'un sujet) ont pu se traduire « dans les effets » (*i.e.*, à l'échelle de plusieurs tâches et plusieurs sujets), parmi les résultats présentés dans le Chapitre Cinq - III.B. Par exemple, on peut calculer ainsi les estimateurs Tâche (**T**) et Pression (**P**).

$$\begin{cases} T = P + A \\ P = K - E \end{cases} \quad (0-15)$$

Dans chaque cas, l'estimateur d'intérêt : $\mathbf{Z} \in \{\mathbf{T}, \mathbf{P}\}$ peut s'exprimer comme la somme d'un estimateur de la *contribution du stress* : $\mathbf{X} \in \{\mathbf{P}, \mathbf{K}\}$, et d'un estimateur de la *contribution d'un facteur confondant* : $\mathbf{Y} \in \{+\mathbf{A}, -\mathbf{E}\}$. Lorsqu'on regarde les graphiques (b) et (d) de la Figure 0.13 à la lumière de l'équation (0-15), on peut être tenté d'avancer l'explication suivante : le calcul de l'estimateur $\mathbf{Z} \in \{\mathbf{T}, \mathbf{P}\}$, à partir de l'estimateur $\mathbf{Y} \in \{+\mathbf{A}, -\mathbf{E}\}$, entraîne une dépendance entre les effets ϑ_z et ϑ_y .

Mais de manière équivalente à l'équation (0-15), on pourrait aussi écrire : $\mathbf{P} = \mathbf{T} - \mathbf{A}$. Or, le graphique (a) de la Figure 0.13 tend à montrer que ϑ_P ne dépend pas systématiquement de ϑ_A ($R=0,22$,

$p=0,23$). De même, on peut écrire : $\mathbf{K} = \mathbf{P} + \mathbf{E}$. Or, le graphique (c) n'affiche aucune corrélation entre ϑ_K et ϑ_E ($R=0,03$, $p=0,87$). Ainsi, le mode de calcul des estimateurs (équation (0-14)) n'explique pas la valeur des tailles d'effet présentées dans le Chapitre Cinq - III.B. En revanche, ces résultats sont cohérents avec le modèle des contributions développé au Chapitre Cinq - II

f) Ces relations montrent un biais dans la valeur de l'estimateur Pression

Le Tableau 0.17 ci-dessous affiche les *corrélations* mesurées sur chaque graphique de la Figure 0.13 et les *contributions* de l'équation (5-4) qui sont communes entre les deux estimateurs concernés.

Tableau 0.17 : Corrélations (Spearman) mesurées entre les tailles d'effet de la Figure 0.13 et contribution commune aux estimateurs correspondants.

N° graphique & Tailles d'effet	Contribution commune	Corrélation (R, p)
a) Pression (P) et Activité (A)	$\pi_a = \pi_c - \pi_{rc}$	R=0,22 ; p=0,23
b) Tâche (T) et Activité (A)	α	R=0,72 ; p<0,001
c) Contrainte (K) et Contexte (E)	$\varepsilon_e = \varepsilon_{rs} - \varepsilon_{rc}$	R=0,03 ; p=0,87
d) Pression (P) et Contexte (E)	$\pi_e = \pi_{rs} - \pi_{rc}$	R=-0,50 , p=0,004

Les conclusions que l'on peut tirer de chaque ligne de ce tableau sont les suivantes.

- La composante parasite $\pi_a = \pi_c - \pi_{rc}$ peut être considérée comme un bruit centré sur Zéro. A l'échelle interindividuelle, l'estimateur Activité (A) n'est donc pas biaisé dans sa mesure de la contribution de l'effet *activité* (α).
- La contribution de l'*activité* (α) est non-nulle et cohérente pour la majorité des caractéristiques considérées dans ce chapitre. L'introduction de cette composante dans notre modèle était donc pertinente et confirme l'importance de considérer un *effet de l'activité*.
- La composante $\varepsilon_e = \varepsilon_{rs} - \varepsilon_{rc}$ peut être considérée comme un bruit centré sur Zéro. A l'échelle interindividuelle, l'estimateur Contrainte (K) n'est pas biaisé dans sa mesure de π_k .
- En revanche, la composante parasite $\pi_e = \pi_{rs} - \pi_{rc}$ est non-nulle et cohérente pour la majorité des caractéristiques considérées dans ce chapitre. A l'échelle interindividuelle, l'estimateur Pression (P) est donc biaisé dans sa mesure de la composante π_k (*pression* liée aux stressseurs). De même, l'estimateur Contexte (E) est biaisé dans sa mesure de la composante ε_e .

Ainsi, la majorité des valeurs obtenues pour les estimateurs Tâche (T), Activité (A), Contrainte (K), Contexte (E) et Pression (P) viennent confirmer le modèle par contributions ε , α , π présenté au Chapitre Cinq - I.B. Ces trois contributions semblent être indépendantes entre elles, comme le montrent la décorrélation des effets Activité et Pression, ou Contrainte et Contexte, sur les graphiques (a) et (c) de la Figure 0.13.

D'après nos conjectures, le graphique (d) de la Figure 0.13 révèle un *effet du contexte* sur la charge affective des participants en condition de repos. Cet effet peut être lié à notre protocole expérimental. Au Chapitre Cinq - III.A, Tableau 5.4 et le Tableau 5.5 ont montré un effet du contexte sur les évaluations psychologiques (EVAS, STAI) pour les tâches D2, TSST et TA. L'explication se trouvait dans la condition Contrôle SECPT_C, qui était la seule à montrer un effet *destressant* sur nos participants. Cette charge affective non-stressante a pu persister dans les conditions Repos (Rc) qui précédaient les autres tâches Contrôle, lorsque celles-ci étaient effectuées à la suite de SECPT_C. A l'échelle de plusieurs

individus, cette charge affective a pu impacter valeur de l'estimateur Contexte (**E**) et de l'estimateur Pression (**P**) calculés sur nos caractéristiques physiologiques.

Cette mesure concrète d'un *effet du contexte* sur les caractéristiques semble ainsi donner raison au scénario « réaliste » sur la distribution de ces composantes, plutôt qu'au scénario « idéal » qui considèrerait que la charge affective était nulle dans les conditions de *non-stress*. Si notre étude s'était limitée à un simple calcul de contraste, comme celui proposé par notre estimateur Pression, nous aurions sélectionné nos caractéristiques sur la base de l'effet Pression sans savoir que celui-ci était partiellement déterminé par la succession des tâches Contrôle durant l'expérience.

On peut tirer de ces résultats les trois conclusions suivantes.

- i. Le modèle des contributions proposé au Chapitre Cinq - I.B est cohérent.
- ii. Il était pertinent de valider un critère par un autre (*i.e.*, l'estimateur **P** par l'estimateur **K**) dans notre sélection des caractéristiques sensibles au stress.
- iii. Il est utile de considérer *tous* les effets mesurables dans une base de données (*e.g.*, l'effet Tâche, l'effet Activité, l'effet Contexte) ; et pas seulement ceux qui servent de critère à la sélection des caractéristiques (*e.g.*, l'effet Pression et l'effet Contrainte).

Annexe 9. Apprentissage en boîte blanche

Cette annexe compare les méthodes de l'apprentissage statistique, sur le plan théorique, pour l'estimation du stress à partir d'une base de données restreinte.

a) Fondements de l'approche bayésienne

◆ Probabilités conditionnelles et loi de Bayes

Pour reprendre les termes du Chapitre Six - , on note \mathcal{F} l'ensemble des caractéristiques identifiées comme des corrélats du stress. Soit f une de ces caractéristiques ; et f_τ la valeur mesurée à un instant τ quelconque. Certaines valeurs de f ont été mesurées plus fréquemment en situation de stress qu'en situation de non-stress (et vice-versa). En d'autres termes : la probabilité \mathbf{P} associée à l'événement « on a mesuré f_τ » ($[f = f_\tau]$), est différente selon que la personne se trouve en situation de stress (\mathbf{S}) ou en situation de non-stress (\mathbf{NS}). Par abstraction, on considère une classe quelconque y – comme correspondant à une situation particulière : Stress ou Non Stress ($y \in \{S, NS\}$), ou encore Stress, Contrôle ou Repos ($y \in \{S, C, R\}$).

Dans le cas où f prend des valeurs discrètes, la probabilité de mesurer la valeur f_τ dans la situation Y correspond à la fréquence où l'on a mesuré f_τ dans la situation Y (Ng & Jordan, 2002).

$$\Pr([f = f_\tau] | [y = Y]) = \frac{\text{Nb}([f = f_\tau] \cap [y = Y])}{\text{Nb}([y = Y])} \quad (0-16)$$

Ici, « $\Pr(a | b)$ » signifie « probabilité de l'événement a , sachant la situation b » ; « $a \cap b$ » symbolise la survenue de l'événement a dans la situation b ; et « $\text{Nb}(c)$ » symbolise le nombre d'exemples pour l'événement ou la situation c . Lorsque f prend des valeurs continues, on postule une loi de probabilité (ex : gaussienne) pour chacune des classes et on estime ses paramètres (ex : moyenne, écart-type) à partir des données disponibles.

On s'intéresse maintenant à un ensemble \mathcal{F} composé de K caractéristiques physiologiques : f^k . A chaque instant de mesure τ , on dispose ainsi d'un ensemble de K mesures : f_τ^k . A partir des valeurs mesurées pour chaque caractéristique, un modèle génératif cherche à décrire la probabilité de mesurer chaque ensemble de valeurs $\{f_\tau^k\}$ pour chaque situation. Pour chaque classe Y , cela revient à estimer une seule loi de probabilité dite *jointe* :

$$P_Y(\tau) = \Pr\left(\bigcap_{f^k \in \mathcal{F}} [f^k = f_\tau^k] | [y = Y]\right) \quad (0-17)$$

Où l'opérateur « $\bigcap_{f^k \in \mathcal{F}} \dots$ » symbolise la mesure d'une valeur f_τ^k pour chaque caractéristique f^k . Ici, $P_Y(\tau)$ exprime la probabilité de mesurer simultanément les valeurs $\{f_\tau^k\}$ dans la situation Y .

A partir de l'information précédente, un classifieur bayésien peut estimer la probabilité de la classe y à partir des mesures $\{f_\tau^k\}$ ¹⁸. Cette probabilité s'exprime ainsi :

¹⁸ Remarque : certains modèles (ex : Markov caché) incluent la probabilité de chaque classe à l'instant précédent. Ces modèles (qui utilisent des paramètres supplémentaires) n'ont pas été retenus ; notamment car nos classes Stress et Non-Stress ne représentent pas une évolution temporelle de l'état de stress. En effet, les sessions Stress et Contrôle de notre campagne expérimentale ne sont pas contiguës dans le temps.

$$P_{\tau}(Y) = \Pr \left([y = Y] \mid \bigcap_{f^k \in \mathcal{F}} [f^k = f_{\tau}^k] \right) \quad (0-18)$$

Pour la classe Y, on passe de la *vraisemblance* $P_Y(\tau)$ exprimée dans l'équation (0-17), à la *probabilité* $P_{\tau}(Y)$ exprimée dans l'équation (0-18), en invoquant la loi de Bayes.

$$P_{\tau}(Y) = \frac{P_Y(\tau) * \Pi_Y}{\sum_y (P_Y(\tau) * \Pi_y)} \quad (0-19)$$

Ici, Π_y représente la prévalence (ou *prior*) de la classe y, c'est-à-dire la fréquence à laquelle on rencontre l'état y. Le numérateur exprime la probabilité marginale de mesurer les valeurs $\{f_{\tau}^k\}$.

◆ *Fonction Discriminante et Classification*

Supposons dans un premier temps que notre problème possède deux classes : Stress ($y=S$) et Non-Stress ($y=NS$). Avec les notations ci-dessus, affirmer : « l'individu est stressé » revient à dire que la probabilité de Stress est la plus grande $P_{\tau}(S) > P_{\tau}(NS)$. Pour poser ce verdict, on peut calculer une *fonction discriminante* λ dont le signe (+ ou -) détermine la classe (S ou NS) en sortie du modèle.

$$\lambda_{\tau} = \ln \left(\frac{P_{\tau}(S)}{P_{\tau}(NS)} \right) \quad (0-20)$$

Formellement, $\lambda : \mathcal{F} \rightarrow \mathbb{R}$ est une projection de l'espace des caractéristique \mathcal{F} (à K dimensions) sur l'axe des réels \mathbb{R} (à 1 dimension). A travers l'expression : $\lambda_{\tau}=0$, on définit dans l'espace \mathcal{F} une « frontière de décision » qui sépare les deux classes S et NS. Pour un classifieur quelconque, l'efficacité de l'apprentissage dépend de sa capacité à définir correctement cette frontière de décision.

b) Méthodes génératives

A partir de l'équation (0-19), on peut exprimer la fonction discriminante λ de la manière suivante.

$$\lambda_{\tau}^{Gen} = \ln \left(\frac{P_{\tau}(S)}{P_{\tau}(NS)} \right) = \ln \left(\frac{P_S(\tau)}{P_{NS}(\tau)} \right) + \ln \left(\frac{\Pi_S}{\Pi_{NS}} \right) \quad (0-21)$$

Un travail d'apprentissage peut être réalisé sur les quatre termes qui déterminent λ^{Gen} . Les deux priors (Π_S et Π_{NS}) correspondent aux proportions globales de Stress et de Non-Stress dans la base de données. Pour les lois jointes ($P_S(\tau)$ et $P_{NS}(\tau)$), on formule des hypothèses simplificatrices qui déterminent le type de classifieur utilisé. Si on suppose que les caractéristiques sont indépendantes en situation de Stress (idem pour le Non-Stress), on utilise un modèle Bayésien Naïf. Dans ce cas, la probabilité jointe $P_S(\tau)$ (idem pour $P_{NS}(\tau)$) est le simple produit des $P(f^k | y=S)$. Si on suppose seulement que les données ont une distribution gaussienne, on utilise l'Analyse Discriminante Quadratique. Si on suppose que ces gaussiennes sont semblables entre les classes (S et NS), on utilise l'Analyse Discriminante Linéaire, plus économe en paramètres. Dans chaque cas, le calcul des paramètres de $P_S(\tau)$ et l'expression de λ_{τ}^{Gen} possèdent une solution unique et connue analytiquement.

Lorsqu'on construit un classifieur à N classes, où $N \geq 3$, la probabilité jointe $P_n(\tau)$ est calculée indépendamment pour chaque classe $n \in \{1..N\}$, puis les posteriors $P_{\tau}(y=n)$ sont déduits de la loi de Bayes à partir du prior Π_n de chaque classe $n \in \{1..N\}$. Dans le cas général, la classification est effectuée en comparant directement les posteriors $P_{\tau}(y=n)$ de chaque classe $n \in \{1..N\}$.

◆ *Limites théoriques*

L'analyse discriminante et l'algorithme bayésien naïf sont des méthodes dites « génératives » (Ng & Jordan, 2002), ou « informatives » (Rubinstein et al., 1997), car elles proposent un modèle statistique des mesures pour chaque classe. D'autres techniques comme la régression logistique sont dites « discriminantes » car elles se focalisent sur la frontière de décision. Lorsque beaucoup d'exemples sont disponibles, un classifieur génératif serait moins efficace que son « équivalent » discriminant, mais il serait plus efficace avec peu d'exemples (Ng & Jordan, 2002). Cette propriété tend à favoriser les approches génératives pour les bases de données restreintes, comme c'est le cas lorsqu'elle est extraite d'une campagne expérimentale.

Mais dans le cas de mesures physiologiques, il faut aussi considérer la robustesse du modèle. Le classifieur génératif fait des hypothèses sur les entrées (distributions gaussiennes, caractéristiques indépendantes, etc.). Ses capacités de généralisation diminueront si les hypothèses ne correspondent pas au comportement réel des données. Par exemple, l'hypothèse gaussienne ne tient pas nécessairement pour les caractéristiques utilisées dans le Chapitre Six - . Il en est de même pour l'hypothèse d'indépendance, car ces quatre caractéristiques sont corrélées entre elles (voir Chapitre Cinq - III.D, Tableau 5.9). Par contraste, une régression logistique postule seulement l'allure de la fonction discriminante λ (voir ci-dessous) ; ce qui la rend plus souple sur la distribution réelle des données (Rubinstein et al., 1997).

Par ailleurs, la plus forte inconnue d'un modèle de stress ne se trouve pas dans le comportement des caractéristiques. Dans l'équation (0-21), le rapport Π_S/Π_{NS} quantifie l'équilibre entre les classes S et NS. Dans la base de données du Chapitre Six - , ce terme vaut 1 si l'on considère comme Non-Stress seulement la classe Contrôle ; et environ 1/2 si l'on considère aussi la classe Repos. Dans la vie réelle, ce terme dépend des activités quotidiennes de la personne : il est imprévisible et n'a aucune raison de correspondre aux ordres de grandeurs proposés dans notre base de données.

Que ce soit par une mauvaise estimation des priors ou par de mauvaises hypothèses, une approche générative aura donc plus de chances d'introduire un biais statistique lorsque le modèle appris en laboratoire sera mis en œuvre sur le terrain¹⁹. Ces arguments justifient l'utilisation de la régression logistique pour dimensionner notre indice de stress dans le Chapitre Six - .

c) Régression Logistique

◆ *2 classes*

Telle qu'on l'a introduite dans le Chapitre Six - , la régression logistique à deux classes consiste à estimer directement la fonction discriminante λ^{RL} , en supposant que celle-ci est une combinaison linéaire des entrées f_{τ}^k .

$$\lambda_{\tau}^{RL} = \beta_0 + \sum_{k=1}^K \beta_k * f_{\tau}^k \quad (0-22)$$

¹⁹ Remarque : il existe un certain nombre de méthodes qui permettraient d'aborder ce problème ; comme l'apprentissage semi-supervisé (Bernardo et al., 2007), qui s'accommode de données non-étiquetées ; ou l'apprentissage par renforcement (Kaelbling et al., 1996), prévu pour évoluer avec de nouvelles mesures étiquetées. Pour être efficaces, ces méthodes nécessitent l'accès à des données expérimentales en vie réelle. Au stade actuel de cette étude, l'option n'est pas encore disponible.

Le travail d'apprentissage consiste à estimer les $K+1$ paramètres β_k de l'équation (0-22), de sorte à maximiser le nombre d'exemples bien classés dans chaque catégorie. On utilise pour cela un algorithme itératif, comme une descente de gradient sur l'entropie croisée entre la classe réelle et la probabilité estimée de stress : $P_\tau(S)$. Dans le problème à deux classes, les probabilités de Stress et de Non-Stress sont complémentaires à chaque instant τ : $P_\tau(S) + P_\tau(NS) = 1$. En injectant cette relation dans l'équation (0-20), on peut exprimer la probabilité de Stress $P_\tau(S)$ à partir de la fonction discriminante λ . On obtient ainsi l'équation (6-3) proposée au Chapitre Six - .

$$P_\tau(S) = \frac{1}{1 + \exp(-\lambda_\tau^{RL})} = \mathcal{L}(\lambda_\tau^{RL}) \quad (0-23)$$

[[Lien de retour : Chapitre Six - I.B.1](#)]

◆ *N classes*

Pour une régression logistique à N classes (ou *multinomiale*), on détermine une classe de référence. Dans le Chapitre Six - , on choisira la classe Stress ; dans le cas général, on sélectionne arbitrairement la classe : « $n=N$ ». L'approche standard consiste alors à diviser le problème à N classes en $N-1$ problèmes à deux classes. On estime ainsi $N-1$ fonctions discriminantes $\lambda^{RL}(n)$, qui expriment le rapport des probabilités de chaque classe n sur celle de la classe N , en recherchant une combinaison linéaire des entrées f_τ^k .

$$\lambda_\tau^{RL}(n) = \ln\left(\frac{P_\tau(n)}{P_\tau(N)}\right) = \beta_0^n + \sum_{k=1}^K \beta_k^n * f_\tau^k \quad (0-24)$$

De cette manière, on cherche à insérer un hyperplan entre chaque paire de classes (n, N) dans l'espace \mathcal{F} des caractéristiques. Cette hypothèse permet d'exprimer la probabilité de chaque classe $n \in \{1..N-1\}$ en fonction de la probabilité de la classe N :

$$P_\tau(n) = \exp\left(\beta_0^n + \sum_{k=1}^K \beta_k^n * f_\tau^k\right) * P_\tau(N) \quad (0-25)$$

On peut alors considérer le fait que la somme de toutes les probabilités fait 1. Cela nous donne une expression pour la probabilité de la classe N :

$$\begin{aligned} P_\tau(N) &= 1 - \sum_{n=1}^{N-1} P_\tau(n) \\ P_\tau(N) &= 1 - P_\tau(N) * \sum_{n=1}^{N-1} \exp\left(\beta_0^n + \sum_{k=1}^K \beta_k^n * f_\tau^k\right) \\ P_\tau(N) &= \frac{1}{1 + \sum_{n=1}^{N-1} \exp\left(\beta_0^n + \sum_{k=1}^K \beta_k^n * f_\tau^k\right)} \end{aligned} \quad (0-26)$$

Et le reste des probabilités peut être déduit de cette formule. Pour un problème à trois classes (Repos **R**, Contrôle **C** et Stress **S**), on obtient :

$$P_\tau(S) = \frac{1}{1 + \exp\left(\beta_0^R + \sum_{k=1}^K \beta_k^R * f_\tau^k\right) + \exp\left(\beta_0^C + \sum_{k=1}^K \beta_k^C * f_\tau^k\right)} \quad (0-27)$$

Où les $\{\beta_k^{R1}\}_{k \in 1..K}$ sont les coefficients calculés dans le problème Stress VS Repos ; et les $\{\beta_k^C\}_{k \in 1..K}$ sont les coefficients calculés dans le problème Stress VS Contrôle.

[[Lien de retour : Chapitre Six - III.C.1](#)]

d) Réseau de neurones à 1 unité dans la couche cachée

◆ Expression

Dans le Chapitre Six - , on appelle \mathcal{N}^1 le réseau de neurones artificiel constitué d'une seule unité dans sa couche cachée et de deux unités dans sa couche de sortie. Le neurone de la couche cachée possède $K+1$ paramètres : β_k^{CC} , $k \in \{0..K\}$; et sa fonction de transfert est la tangente hyperbolique \mathcal{T} . Chaque neurone n de la couche de sortie possède deux paramètres β_j^{CS-n} , $j \in \{0,1\}$; et leur fonction de transfert est une fonction Softmax appliquée à l'ensemble de la couche de sortie. Si on appelle \mathcal{N}_τ^1 la probabilité de stress en sortie de \mathcal{N}^1 , cette quantité s'écrit :

$$\begin{aligned} \mathcal{N}_\tau^1 &= \frac{\exp\left(\beta_0^{CS1} + \beta_1^{CS1} * \mathcal{T}\left(\beta_0^{CC} + \sum_{k=1}^K \beta_k^{CC} * f_\tau^k\right)\right)}{\exp\left(\beta_0^{CS1} + \beta_1^{CS1} * \mathcal{T}\left(\beta_0^{CC} + \sum_{k=1}^K \beta_k^{CC} * f_\tau^k\right)\right) + \exp\left(\beta_0^{CS2} + \beta_1^{CS2} * \mathcal{T}\left(\beta_0^{CC} + \sum_{k=1}^K \beta_k^{CC} * f_\tau^k\right)\right)} \\ &= \mathcal{L}\left(\beta_0^{CS} + \beta_1^{CS} * \mathcal{T}\left(\beta_0^{CC} + \sum_{k=1}^K \beta_k^{CC} * f_\tau^k\right)\right), \quad \beta_j^{CS} = \beta_j^{CS1} - \beta_j^{CS2} \text{ pour } j \in \{0,1\} \end{aligned} \quad (0-28)$$

Où \mathcal{L} est la fonction logistique et les coefficients β_j^{CS} expriment la différence des paramètres de la couche de sortie. L'utilisation de la fonction Softmax pour un problème à deux classes équivaut à une régression logistique sur la sortie du neurone de la couche cachée.

◆ Frontière de décision

Avec l'expression précédente, on peut facilement montrer que la frontière de décision entre les classes Stress et Contrôle ($\mathcal{N}_\tau^1 = 1/2$) est linéaire dans l'espace des caractéristiques.

$$\begin{aligned} \mathcal{N}_\tau^1 &= \frac{1}{2} \\ \beta_0^{CS} + \beta_1^{CS} * \mathcal{T}\left(\beta_0^{CC} + \sum_{k=1}^K \beta_k^{CC} * f_\tau^k\right) &= 0 \\ \mathcal{T}\left(\beta_0^{CC} + \sum_{k=1}^K \beta_k^{CC} * f_\tau^k\right) &= -\frac{\beta_0^{CS}}{\beta_1^{CS}}, \quad \beta_1^{CS} \neq 0 \\ \beta_0^{CC} + \sum_{k=1}^K \beta_k^{CC} * f_\tau^k &= \mathcal{T}^{-1}\left(-\frac{\beta_0^{CS}}{\beta_1^{CS}}\right), \quad |\beta_0^{CS}| < |\beta_1^{CS}| \\ \underbrace{\beta_0^{CC} + \mathcal{T}^{-1}\left(\frac{\beta_0^{CS}}{\beta_1^{CS}}\right)}_{\beta_0'} + \sum_{k=1}^K \beta_k^{CC} * f_\tau^k &= 0 \end{aligned} \quad (0-29)$$

La fonction arctangente hyperbolique (\mathcal{T}^{-1}) est impaire, d'où l'expression de β_0' . Avec ce terme comme intercept, la dernière ligne est une équation linéaire du point de vue des mesures f_τ^k . A chaque étape de la validation croisée dans le Chapitre Six - , on s'est assuré que le modèle linéaire produit par les coefficients précédents prédisait les mêmes classes que le modèle \mathcal{N}^1 . Ainsi, la frontière de décision du réseau \mathcal{N}^1 est un hyperplan dans l'espace \mathcal{F} , au même titre que la frontière d'une régression logistique à deux classes (les paramètres de \mathcal{N}^1 sont simplement déterminés de manière plus

élaborée). Cette continuité de forme entre la régression logistique et les réseaux de neurones justifie le choix de cette paire d'algorithmes dans notre comparaison de modèles.

◆ *Vers la frontière de décision idéale*

Un réseau de neurones tel que ceux présentés dans le le Chapitre Six - peut théoriquement résoudre n'importe quel problème de classification, pourvu qu'on lui fournisse suffisamment d'unités dans la couche cachée (Cybenko, 1989). En effet, un nombre arbitrairement grand de paramètres permet à un tel modèle de tracer une frontière de décision arbitrairement complexe dans l'espace \mathcal{F} des caractéristiques fournies au modèle. Cette qualité « d'approximateur universel » explique en partie le succès des réseaux de neurones artificiels dans les problèmes standard de classification. On peut dès lors se demander combien d'unités dans la couche cachée sont nécessaires pour modéliser la frontière « idéale » pour l'état de stress dans notre espace de caractéristiques. Cette question est adressée au travers des modèles \mathcal{N}^x proposés dans le Chapitre Six - .

[[Lien de retour : Chapitre Six - III.C.3](#)]

e) *Mesure des performances d'un détecteur de stress*

Une fois plusieurs modèles appris sur une base d'apprentissage commune, on cherche à comparer leurs performances sur la base de test. Reprenons notre problème à deux classes : à chaque exemple d'une certaine classe $\mathbf{y} \in \{\mathbf{S}, \mathbf{NS}\}$, un détecteur de stress doit attribuer une étiquette $\hat{\mathbf{y}} \in \{\mathbf{S}, \mathbf{NS}\}$.

Le taux de *vrais positifs* ($\mathbf{VP}^{\%}$) correspond à la proportion d'exemples \mathbf{S} correctement étiquetés par le modèle. Son complémentaire est le taux de faux négatifs ($\mathbf{FN}^{\%}$), la proportion d'exemples \mathbf{S} mal étiquetés par le modèle. Symétriquement, le taux de *vrais négatifs* ($\mathbf{VN}^{\%}$) correspond la proportion d'exemples \mathbf{NS} correctement étiquetés par le modèle. Son complémentaire est le taux de faux positifs ($\mathbf{FP}^{\%}$), la proportion d'exemples \mathbf{NS} mal étiquetés par le modèle. Ces quatre taux s'expriment de la manière suivante, où les probabilités $\Pr(\mathbf{X}|\mathbf{Y})$ sont calculées sur le modèle de l'équation (0-16).

$$\begin{aligned} \mathbf{VP}^{\%} &= \Pr\left(\frac{[\hat{\mathbf{y}} = \mathbf{S}] | [\mathbf{y} = \mathbf{S}]}{\text{Vrai Positif}}\right) \\ \mathbf{FP}^{\%} &= \Pr\left(\frac{[\hat{\mathbf{y}} = \mathbf{S}] | [\mathbf{y} = \mathbf{NS}]}{\text{Faux Positif}}\right) = 1 - \mathbf{VN}^{\%} \\ \mathbf{VN}^{\%} &= \Pr\left(\frac{[\hat{\mathbf{y}} = \mathbf{NS}] | [\mathbf{y} = \mathbf{NS}]}{\text{Vrai Négatif}}\right) \\ \mathbf{FN}^{\%} &= \Pr\left(\frac{[\hat{\mathbf{y}} = \mathbf{NS}] | [\mathbf{y} = \mathbf{S}]}{\text{Faux Négatif}}\right) = 1 - \mathbf{VP}^{\%} \end{aligned} \quad (0-30)$$

Un algorithme d'apprentissage est généralement calibré de sorte à maximiser la *précision* du modèle ($\mathbf{Prc}^{\%}$), c'est à-dire sa capacité à étiqueter correctement chaque classe du modèle. Cette quantité est un compromis entre le taux de faux positifs et le taux de vrais positifs. Afin de prévenir un éventuel déséquilibre des classes, on peut simplement calculer la moyenne de ces deux taux.

$$\mathbf{Prc}^{\%} = \frac{1}{2} * (\mathbf{VP}^{\%} + \mathbf{VN}^{\%}) \quad (0-31)$$

Cette formule a généré les performances affichées au Chapitre Six - .

[[Lien de retour : Chapitre Six - IV.A.1](#)]

Annexe 10. Interpolation de l'Indice Simple

a) Expression des coefficients

Au paragraphe II.C du Chapitre Six - , on dispose de deux points (X_{Ref}, Y_{Ref}) et (X_{Cer}, Y_{Cer}) que l'on souhaite interpoler avec la courbe d'équation :

$$y = \frac{1}{1 + \exp(a * x + b)} \tag{0-32}$$

Où x représente une valeur calibrée (théoriquement « proche » de Zéro) d'une caractéristique quelconque et (a, b) sont les paramètres de l'indice de stress appris sur notre base de données. Formellement, on recherche une expression des paramètres (a, b) telle que cette courbe passe exactement par les points (X_{Ref}, Y_{Ref}) et (X_{Cer}, Y_{Cer}) . La reproduction (ci-dessous) de la Figure 6.3 donne l'allure souhaitée, avec : $X_{Ref} = f_{Ref}$, $X_{Cer} = f_{Cer}$, $Y_{Ref} = S_{Ref}$ et $Y_{Cer} = S_{Cer}$.

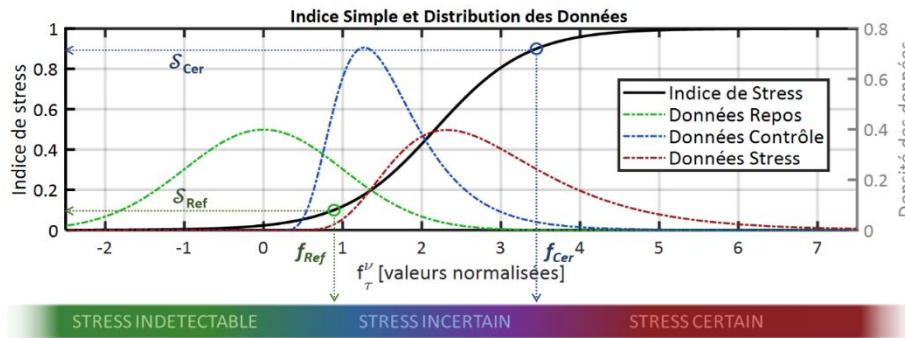


Figure 6.3 : Dimensions de l'indice de Stress.

On commence par observer que la fonction logistique \mathcal{L} est inversible. On peut exprimer sa réciproque \mathcal{L}^{-1} en fonction de x, a et b .

$$\forall y \in]0; 1[, \quad \mathcal{L}^{-1}(y) = -\ln\left(\frac{1}{y} - 1\right) \tag{0-33}$$

$$\exists ! x \in \mathbb{R}, \quad \mathcal{L}^{-1}(y) = -a * x - b$$

On peut alors écrire le classique système de 2 équations à 2 inconnues qui relie deux points d'une droite (ici : $(X_{Ref}, \mathcal{L}^{-1}(Y_{Ref}))$ et $(X_{Cer}, \mathcal{L}^{-1}(Y_{Cer}))$) aux paramètres de cette droite (ici a, b) :

$$\begin{cases} a * X_{Ref} + b = \ln\left(\frac{1}{Y_{Ref}} - 1\right) \\ a * X_{Cer} + b = \ln\left(\frac{1}{Y_{Cer}} - 1\right) \end{cases} \tag{0-34}$$

Ce système possède une solution unique et nos deux paramètres héritent des formules suivantes :

$$\begin{cases} a = \frac{\ln\left(\frac{1}{Y_{Cer}} - 1\right) - \ln\left(\frac{1}{Y_{Ref}} - 1\right)}{X_{Cer} - X_{Ref}} \\ b = \ln\left(\frac{1}{Y_{Ref}} - 1\right) - a * X_{Ref} \end{cases} \tag{0-35}$$

Au Chapitre Six - , ces deux formules ont servi de base pour calculer (au signe près) les hyperparamètres (β_0, β_1) de l'indice \mathcal{M} à chaque nouvel apprentissage des seuils f_{Ref} et f_{Cer} .

b) Application & rôle de la classe Repos

En pratique, chaque point $(f_{Ref}, \mathcal{S}_{Ref})$ et $(f_{Cer}, \mathcal{S}_{Cer})$ est déterminé par une régression logistique différente : un détecteur Stress/Repos pour $(f_{Ref}, \mathcal{S}_{Ref})$ et un détecteur Stress/Contrôle pour $(f_{Cer}, \mathcal{S}_{Cer})$. Dans cette démarche, on aurait pu s'appuyer simplement sur une régression logistique Stress/Contrôle. Quelle est l'utilité de la classe Repos dans le processus d'apprentissage ?

Une illustration du problème est proposée dans la Figure 0.14. Sur les graphiques du bas (b, d), celle-ci montre la probabilité de stress estimée par les régressions logistiques (1) et (2) proposées plus haut, à partir des données représentées sur les graphiques (a, c). On peut observer la construction de l'indice \mathcal{S} par interpolation sur les points $(f_{Ref}, \mathcal{S}_{Ref})$ et $(f_{Cer}, \mathcal{S}_{Cer})$, chacun marqué par un rond noir.

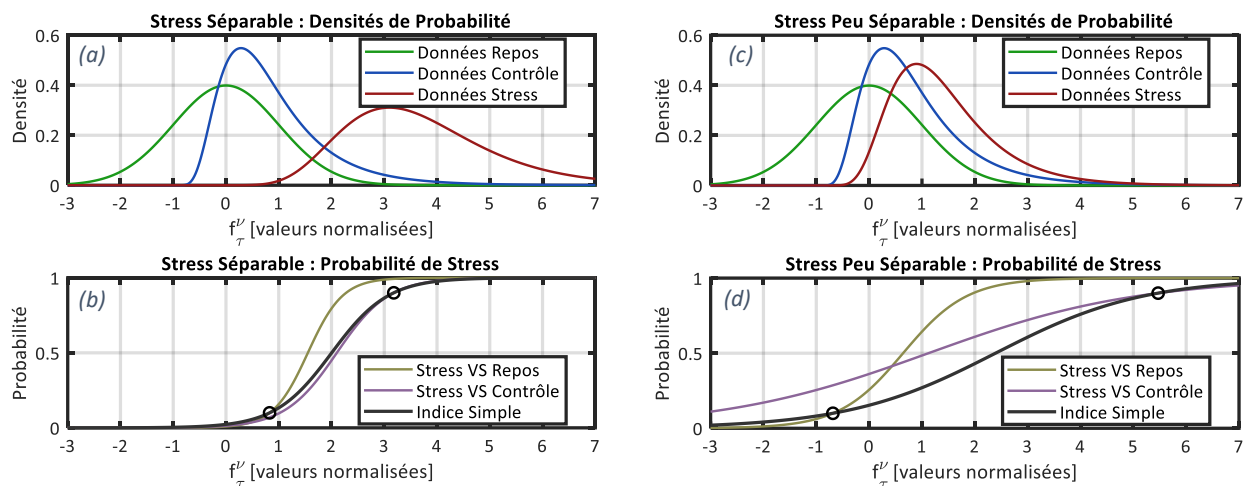


Figure 0.14 : Construction de l'indice simple sur données simulées

En haut (a, c) : Densité de probabilité de chaque classe.

En bas (b, d) : Probabilité de stress estimée par régression logistique (Stress VS Repos, Stress VS Contrôle) et par l'indice de stress.

A gauche (a, b) : Cas où la classe Stress est séparable de la classe Contrôle.

A droite (c, d) : Cas où la classe Stress est peu séparable de la classe Contrôle.

Seule la densité probabilité du Stress change entre les graphiques (a) et (c). On peut ainsi distinguer deux cas de figure. (i) Lorsque les classes Stress et Contrôle sont « bien séparables » (graphique (a)), il y a peu de différence entre la courbe de la régression Stress/Contrôle et la courbe de l'indice \mathcal{S} (graphique (b)). (ii) Lorsque la classe stress est « peu séparable » de la classe Contrôle (graphique (c)), la courbe de régression Stress/Contrôle a tendance à s'étirer dans l'espace des valeurs de f (graphique (d)). Cet effet est moins important pour l'indice \mathcal{S} , dont la dispersion dans les valeurs négatives de f est limitée par le point de référence : $(f_{Ref}, \mathcal{S}_{Ref})$.

Quelle que soit la caractéristique envisagée, c'est le cas « peu séparable » qui est concrètement observé dans notre base de données. Utilisée telle quelle, la courbe Stress/Contrôle visible sur le graphique (d) générera *spontanément* des valeurs élevées, même en l'absence d'activité mentale ou motrice. Cette courbe prédit en effet de fortes probabilités de stress sur un domaine où la classe Repos est prédominante ($f_{\tau}^v < 0,5$). De telles valeurs se répercuteront en proportion sur la valeur de l'indice multiple \mathcal{M} .

Ainsi, l'utilisation d'un détecteur Stress/Repos permet de s'assurer que l'indice simple \mathcal{S} présente des valeurs faibles en situation de Repos. Cette stratégie permet de diminuer le bruit porté sur l'indice multiple \mathcal{M} par les variations spontanées de la caractéristique f .

c) Contribution d'une caractéristique sur la sortie de l'indice multiple

Dans le Chapitre Six -, on définit l'indice \mathcal{M} comme une moyenne des indices simples \mathcal{S} de chaque caractéristique. Si K est le nombre de caractéristiques utilisées pour l'indice \mathcal{M} , la contribution de chacune d'entre elles sur la sortie vaut $1/K$ fois la valeur de l'indice \mathcal{S} .

Afin de représenter cette valeur, on peut approximer l'indice \mathcal{S} avec une fonction linéaire par morceaux L représentée par l'équation suivante.

$$\mathcal{S} \approx L, \quad L = \begin{cases} 0, & \text{si } f \in]-\infty; f_{Ref}[\\ \frac{f - f_{Ref}}{f_{Cer} - f_{Ref}}, & \text{si } f \in [f_{Ref}; f_{Cer}] \\ 1, & \text{si } f \in]f_{Cer}; +\infty[\end{cases} \quad (0-36)$$

La contribution de chaque caractéristique sur \mathcal{M} vaut donc 0 ou $1/K$ en-dehors des bornes f_{Ref} et f_{Cer} . Même pour des valeurs extrêmes de f , celle-ci est donc limitée lorsque K est suffisamment grand.

A l'intérieur des bornes f_{Ref} et f_{Cer} , la contribution de f peut être considérée comme inversement proportionnelle à l'amplitude de l'intervalle : $f_{Cer} - f_{Ref}$. Dans le paragraphe précédent, on a montré que cet intervalle augmentait lorsque la séparabilité des classes diminuait du point de vue de la caractéristique f . On peut alors conclure sur les propriétés suivantes.

- i) *La contribution des caractéristiques sur l'indice multiple est bornée dans l'intervalle : $]0; 1/K[$, dont les bornes correspondent aux valeurs seuils f_{Ref} et f_{Cer} .*
- ii) *Dans l'intervalle $[f_{Ref}; f_{Cer}]$, la contribution de la caractéristique f augmente avec le pouvoir discriminant de f entre les classes Stress et Contrôle.*

Annexe 11. Algorithmes d'apprentissage

Cette annexe rassemble les algorithmes d'apprentissage proposés au Chapitre Six - , section III. Il s'agit de pseudo-code synthétisant les algorithmes sur Matlab qui ont permis de générer les résultats présentés dans la section IV. Pour que ces algorithmes suffisent à leur compréhension, certaines variables et fonctions sont explicitées dans un glossaire en fin d'annexe.

a) Protocole de validation croisée

L'algorithme de validation croisée proposé dans le paragraphe III.B.2 est donné ci-dessous.

Algorithme 1 : Calcul des performances dans un protocole de validation croisée LOSOTO

```

Fonction Validation_Croisée (Modèle, Base_Données)
  # Initialiser le vecteur des prédictions
  Prédications = []
  Pour chaque Tâche dans Base_Données
    Pour chaque Sujet dans Base_Données
      # Diviser la base
      Base_Apprentissage, Base_Test = ...
          ... Diviser_LOSOTO (Base_Données, Sujet, Tâche)
      # Apprendre le modèle
      Modèle_Appris = Apprendre (Modèle, Base_Apprentissage)
      # Prédire la sortie du modèle
      Prédict_tmp = Prédire (Modèle_Appris, Mesures (Base_Test))
      # Ajouter au vecteur des prédictions
      Prédications = Ajouter (Prédications, Prédict_tmp)
    Fin
  Fin
  # Calculer la performance du modèle à partir des classes réelles
  Perf = Calculer_Précision (Prédications, Classes (Base_Données))
  Renvoyer Perf, Prédications

```

Ici, la variable Prédications peut contenir des valeurs continues jusqu'à l'évaluation finale des performances, qui utilise un seuil sur cette variable. Dans le cas de l'indice \mathcal{M} , on peut utiliser les valeurs prédites pour choisir ce seuil de détection.

[\[Lien de retour : Chapitre Six - III.B.2.c \]](#)

b) Apprentissage d'un Réseau de Neurones

Pour l'apprentissage d'un réseau de neurones, la base d'apprentissage est divisée en deux ensembles distincts : une base d'entraînement et une base de validation. A chaque itération, les paramètres du modèle sont optimisés sur la base d'entraînement ; puis le modèle est appliqué sur la base de validation, pour vérifier que ses capacités de généralisation n'ont pas diminué depuis les itérations précédentes. Dans notre analyse, la base de validation des réseaux de neurones a été sélectionnée avec un découpage *LOSO*. Ce découpage est proposé pour chaque sujet, puis on conserve le modèle qui possède le meilleur pouvoir discriminant sur la base d'apprentissage initiale. Ici, le pouvoir discriminant correspond à l'entropie croisée sur les valeurs prédites par le modèle.

Algorithme 2 : Apprentissage d'un réseau de neurones dans l'expérience 1

```

Fonction Apprendre (Modèle, Base_Apprentissage)
  # Initialiser le vecteur d'erreur et la liste des modèles appris
  Modèles = []
  Erreurs = []
  Pour chaque Sujet dans Base_Apprentissage
    # Diviser la base d'apprentissage : Entraînement + Validation
    Base_Entraînement, Base_Validation = ...
    ... Diviser_LOSO (Base_Apprentissage, Sujet)
    # Apprendre le modèle courant
    Modèle_tmp = Entraîner (Modèle, Base_Entraînement, Base_Validation)
    # Prédire la sortie du modèle courant sur la base d'apprentissage
    Prédications = Prédire (Modèle_tmp, Mesures (Base_Apprentissage))
    # Calculer la performance du modèle à partir des classes réelles
    Erreur_tmp = Entropie_Croisée (Prédications, Classes (Base_Apprentissage))
    # Stocker l'erreur et le modèle courants
    Erreurs = Ajouter (Erreurs, Erreur_tmp)
    Modèles = Ajouter (Modèles, Modèle_tmp)
  Fin
  # Conserver le modèle le plus adapté à la base d'apprentissage
  Modèle_Final = Modèles [Indice_Minimum (Erreurs)]
  Renvoyer Modèle_Final

```

Les paramètres du `Modèle_Final` dépendent de leur initialisation dans la fonction `Entraîner`. Celle-ci est aléatoire. Afin de contrôler la part d'aléatoire dans la mesure des performances, on a appris et validé plusieurs fois chaque réseau de neurones avec le protocole défini dans l'Algorithme 1.

Algorithme 3 : Mesure des performances d'un réseau de neurones dans l'expérience 1

```

Fonction Mesurer_Performances (Modèle, Base_Données)
  # Initialiser le vecteur des précisions mesurées
  Perfs = []
  # Nombre d'itérations = deux fois le nombre de paramètres
  Nb_Itérations = [ 2*Nb_Paramètres (Modèle) Si Type (Modèle) == Réseau_Neurones ...
    ... Sinon 1 ]
  Pour chaque Itération dans 1 .. Nb_Itérations
    # Mesurer la performance en validation croisée pour l'itération
    # courante (Voir Algorithme 1)
    Perf_tmp = Validation_Croisée (Modèle, Base_Données)
    # Stocker la performance mesurée
    Perfs = Ajouter (Perfs, Perf_tmp)
  Fin
  Renvoyer Moyenne (Perfs)

```

Les performances affichées dans le Tableau 6.4 correspondent à la performance moyenne obtenue pour chacun des modèles – et à l'intervalle de confiance à 95% calculé à partir de la taille d'échantillon.

[\[Lien de retour : Chapitre Six - III.C.2 \]](#)

c) Mesure de performances dans un environnement dégradé

On peut reprendre certains éléments des algorithmes 1 et 3 pour mesurer les performances d'un algorithme d'apprentissage sur des données bruitées. L'Algorithme 4 ci-dessous propose de répliquer 50 fois chaque base de test en lui appliquant une fonction de bruit, au fur et à mesure du protocole de validation LOSOTO. La fonction de bruit en question peut être appliquée à une seule ou à toutes les caractéristiques (l'information est absente de l'algorithme).

Algorithme 4 : Mesure des performances d'un modèle dans un environnement dégradé (expérience 2)

```

Fonction Mesurer_Robustesse(Modèle, Base_Données, Bruit)
  Nb_Réplifications = 50
  # Initialiser le vecteur des précisions mesurées
  Perfs = []
  # Procédé de validation croisée (voir Algorithme 1) adapté à l'introduction de bruit
  Pour chaque Tâche dans Base_Données
    Pour chaque Sujet dans Base_Données
      # Diviser la base en prélevant une base de test
      Base_Apprentissage, Base_Test = ...
        ... Diviser_LOSOTO(Base_Données, Sujet, Tâche)
      # Répliquer la base de test
      Base_Test = Répliquer(Base_Test, Nb_Réplifications)
      # Bruiter la base de test
      Base_Test = Fonction_Bruit(Base_Test)
      # Apprendre le modèle
      Modèle_Appris = Apprendre(Modèle, Base_Apprentissage)
      # Calculer la sortie du modèle sur tous les réplicats bruités
      Pour chaque Réplicat dans Base_Test
        # Calculer la sortie du modèle
        Prédications = Prédire(Modèle_Appris, Mesures(Réplicat))
        # Ajouter au vecteur des prédictions correspondant à ce réplicat
        Perf_tmp = Calculer_Précision(Prédications, Classes(Réplicat))
        Perfs = Ajouter(Perfs, Perf_tmp)
      Fin
    Fin
  Fin
  Renvoyer Moyenne(Perfs)

```

Les résultats affichés dans la Figure 6.8 correspondent à la performance moyenne obtenue pour chacun des modèles – et à l'intervalle de confiance à 95% calculé à partir de la taille d'échantillon.

[\[Lien de retour : Chapitre Six - III.D.2 \]](#)

d) Glossaire

◆ Variables

- `Modèle` contient toutes les informations liées au modèle que l'on cherche à entraîner : type, valeurs des paramètres, algorithme d'apprentissage, etc.
- `Base_Données` contient toutes les mesures de la base de données, ainsi que toutes les informations relatives à ces mesures (Sujet, Tâche, Classe).

- `Base_Apprentissage` contient toutes les données utilisées pour apprendre un modèle.
- `Base_Entraînement` et `Base_Validation` se réfèrent au découpage des données appliqué pour apprendre un réseau de neurones.
- `Base_Test` contient toutes les données utilisées pour tester un modèle.

◆ *Fonctions*

- `Mesures(Base_Données)` renvoie les mesures de la Base fournie en argument.
- `Classes(Base_Données)` renvoie les classes de la Base fournie en argument.
- `Type(Modèle)` Renvoie le type du Modèle : Régression, Réseau de neurones, etc.
- `Diviser_LOSOTO(Base_Données, Sujet, Tâche)` divise la Base fournie en argument en « laissant dehors » le Sujet et la Tâche indiqués.
- `Diviser_LOSO(Base_Données, Sujet)` divise la Base en « laissant dehors » le Sujet indiqué.
- `Apprendre(Modèle, Base_Apprentissage)` apprend le Modèle sur la Base_Apprentissage en utilisant l'algorithme approprié.
- `Entraîner(Modèle, Base_Entraînement, Base_Validation)` ajuste les paramètres du Modèle sur la Base_Entraînement en utilisant la Base_Validation pour arrêter l'apprentissage au bon moment.
- `Prédire(Modèle, Données)` prédit la classe des Données à partir du Modèle fourni.
- `Ajouter(A, B)` renvoie la concaténation de A et B.
- `Nb_Paramètres(Modèle)` renvoie le nombre de paramètres du Modèle fourni.
- `Moyenne(Données)` renvoie la moyenne du vecteur de Données fourni.
- `Répliquer(Base_Données, Nb_Réplifications)` réplique la Base de données fournie, le nombre de fois indiqué, en identifiant chaque réplikat obtenu.
- `Fonction_Bruit(Base_Données)` applique à la Base de données fournie la fonction de bruit vue au paragraphe III.D du Chapitre Six - . Cette transformation peut concerner une seule ou toutes les caractéristiques de la base de données.
- `Indice_Minimum(Données)` renvoie l'indice du minimum dans le vecteur de Données fourni.

Annexe 12. Disponibilité du Signal IBI pour le Calcul d'une Caractéristique Cardiaque

Dans le Chapitre Sept - , l'indice de carence \mathcal{X} est apparu comme un bon prédicteur de la précision avec laquelle on pouvait calculer une caractéristique issue d'Empatica E4. Partant de ce constat, on peut poser un seuil de tolérance x sur l'indice de carence \mathcal{X} ; c'est-à-dire la proportion d'échantillons manquants que l'on s'autorise pour calculer une caractéristique cardiaque.

a) Méthode

En se basant sur l'équation (7-8), on peut définir un nouveau critère qui déterminera si une fenêtre de temps est « disponible » d'après l'indice \mathcal{X} , c'est-à-dire qu'elle contient suffisamment d'information pour le calcul envisagé. Ce critère peut être appliqué sans distinction aux signaux IBI_{bh3} et IBI_{e4} , découpés en segments de longueurs égales. Pour disposer d'une référence, on peut construire un critère similaire à partir de l'indicateur de confiance C_{bh3} basé sur l'équation (7-11). Si l'on appelle \mathcal{D} l'ensemble des fenêtres de longueur L qui sont « disponibles » d'après l'indice \mathcal{X} ou C_{bh3} , alors nos trois critères de disponibilité s'écrivent de la manière suivante.

$$\begin{aligned} k \in \{bh3, e4\}, \quad \tau \in \mathcal{D}(\mathcal{X}_k, x) &\Leftrightarrow \left(\mathcal{X}_k(\tau) - \frac{1}{L} * \min_{IBI_k^n \in \tau} (IBI_k^n) \right) < x \\ \tau \in \mathcal{D}(C_{bh3}, x) &\Leftrightarrow \min_{C_{bh3} \in \tau} (C_{bh3}) \geq 100 * (1 - x) \end{aligned} \quad (0-37)$$

Lorsque $x=0\%$, ces critères sont équivalents aux critères d'intégrité proposés dans le Chapitre Sept - section III.A. Etant donné un seuil quelconque $x \in [0 ; 1]$, on peut enfin calculer la proportion des segments IBI « disponibles » pour le calcul d'une caractéristique sur une journée d'enregistrement J . On appelle *taux de disponibilité* $\mathbf{D}_J(x)$, cette proportion de segments disponibles au seuil x sur cette journée d'enregistrement J . Tout comme nos critères, ce taux peut être décliné en quatre versions : trois basées sur l'indice de carence \mathcal{X} et une basée sur l'indicateur de référence C_{bh3} .

$$\begin{aligned} k \in \{bh3, e4\}, \quad d_J(\mathcal{X}_k, x) &= \text{Proba}_{\tau \in J} [\tau \in \mathcal{D}(\mathcal{X}_k, x)] \\ d_J(C_{bh3}, x) &= \text{Proba}_{\tau \in J} [\tau \in \mathcal{D}(C_{bh3}, x)] \end{aligned} \quad (0-38)$$

Le calcul des taux de disponibilité $d_J(x)$ sur l'ensemble de notre base de données nous permet de répondre à la question suivante : en vie réelle, quelle est la probabilité de pouvoir calculer la caractéristique f sur une journée d'enregistrement quelconque ? La réponse à cette question peut dépendre d'au moins trois facteurs (en plus de la tolérance x) :

- Les niveaux d'activité physique sur la période d'enregistrement (car le mouvement compromet la qualité des signaux cardiaques).
- La qualité d'installation des capteurs (par les participants eux-mêmes) en début de journée (car des tetrodes ou un bracelet mal positionnés peuvent altérer les signaux cardiaques).
- La durée de la fenêtre de temps (car il est plus exigeant de chercher des signaux IBI hautement fidèles sur des intervalles de temps plus longs).

La qualité d'installation des capteurs justifie le calcul des d_J sur chaque journée d'acquisition, plutôt que sur toute la base de données. Pour observer l'influence de la fenêtre de temps, ces journées seront découpées en fenêtres de tailles variables.

b) Résultats

On a ainsi calculé les taux de disponibilité correspondant aux indices C_{bh3} , \mathcal{X}_{bh3} et \mathcal{X}_{e4} , sur des fenêtres de largeur $L = 30s, 60s$ et $300s$, pour chaque journée d'enregistrement (voir équation (0-38)). Les résultats sont affichés pour différents seuils de tolérance (ici $x = 0\%, 5\%, 10\%, 50\%$ et 90%) sur la Figure 0.15. Les barres d'erreur correspondent à l'écart interquartile des taux de disponibilité journaliers (d) sur les 11 journées d'enregistrement ; et les points reliés correspondent à la valeur médiane, pour chacune des longueurs L utilisées pour les fenêtres de temps.

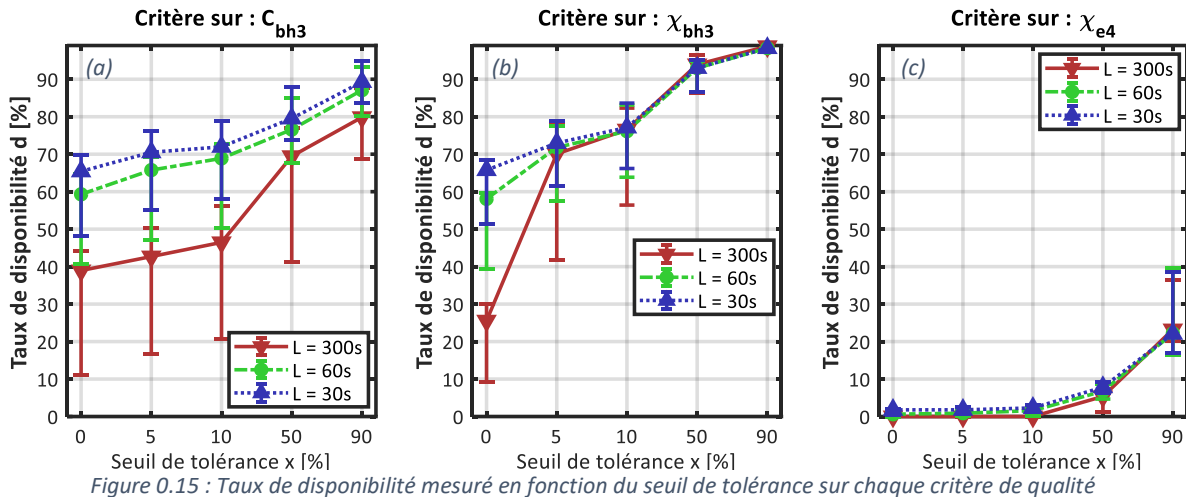


Figure 0.15 : Taux de disponibilité mesuré en fonction du seuil de tolérance sur chaque critère de qualité

Résultat pour l'indice C_{bh3} (a), l'indice \mathcal{X}_{bh3} (b) et l'indice \mathcal{X}_{e4} (c), pour des fenêtres de longueur 30s (en bleu), 1min (en vert) et 5min (en rouge). Les points représentent la valeur médiane entre les journées d'enregistrement ; les barres d'erreur représentent l'écart interquartile.

La part de segments sélectionnés (d) augmente en fonction de la tolérance (x), quels que soient le critère (C_{bh3} , \mathcal{X}_{bh3} ou \mathcal{X}_{e4}) ou la longueur de fenêtre (30s à 300s). L'utilisation de fenêtres plus longues diminue la part des segments de signal qui vérifient nos critères de disponibilité. Cette propriété est surtout vérifiée pour le critère sur C_{bh3} , le taux de disponibilité étant similaire pour les trois courbes dès 5% de tolérance avec le critère sur \mathcal{X}_{bh3} ; et dès 0% avec le critère sur \mathcal{X}_{e4} .

Avec des fenêtres de temps courtes ($L \in \{30, 60\}s$), les résultats obtenus sont similaires pour \mathcal{X}_{bh3} et C_{bh3} . Dans une journée d'enregistrement classique, on peut s'attendre à obtenir deux tiers d'échantillons complets pour une fenêtre de temps de 30s. C'est une bonne nouvelle pour le calcul de la variabilité cardiaque haute fréquence, qu'on peut effectuer sur des intervalles courts. Comme on pouvait s'y attendre, cependant, très peu de fenêtres sont disponibles à partir du signal PPG, et ce même pour des seuils de tolérance élevés. Pour une tolérance $x = 10\%$, environ 8 fenêtres de 30s à 300s sont disponibles pour la mesure IBI_{h3} mais seulement une sur cinquante pour la mesure IBI_{e4} .

Annexe 13. Valorisation des Travaux

La liste ci-dessous rassemble les communications scientifiques (avec comité de lecture) issues des travaux de cette thèse.

ARTICLES DE REVUE

- G. Vila**, C. Godin, S. Charbonnier, E. Labyt, O. Sakri, and A. Campagne (2019). "Real-Time Monitoring of Passenger's Psychological Stress", *Future Internet*, vol. 11, no. 5, p. 102
- G. Vila**, C. Godin, S. Charbonnier, E. Labyt, O. Sakri, and A. Campagne (2019). Physiological response to acute stress against confounding factors : A white-box research method. *Biomedical Physics & Engineering Express*, 5(5), 055007.

ARTICLES DE CONFERENCES INTERNATIONALES

- G. Vila**, C. Godin, S. Charbonnier, E. Labyt, O. Sakri, and A. Campagne (2018), "Pressure-Specific Feature Selection for Acute Stress Detection From Physiological Recordings," in 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp. 2341–2346.
- O. Sakri, C. Godin, **G. Vila**, E. Labyt, S. Charbonnier, and A. Campagne (2018), "A Multi-User Multi-Task Model For Stress Monitoring From Wearable Sensors," in 2018 21st International Conference on Information Fusion (FUSION), 2018, pp. 761–766.
- S. Charbonnier, **G. Vila**, C. Godin, E. Labyt, O. Sakri, et A. Campagne (2018), "A Multi-feature Fuzzy Index to Assess Stress Level from Bio-signals", in 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Honolulu, HI, p. 1086-1089.

INTERVENTIONS EN SEMINAIRES

- G. Vila** et al. (2017), "Real-Time Monitoring of Traveller's Psychological Stress",
- Présentation orale + article de colloque, International Congress on Transportation Research (ICTR), Thessalonique, Grèce (Sept. 2017).
 - Présentation poster, Journée Visage Geste Action et Comportement (GdR Isis), Paris, France (Dec. 2017).

ARTICLES EN COURS DE REDACTION

- G. Vila**, C. Godin, S. Charbonnier, and A. Campagne (2021), "Real-Time Quality Index to Control Data Loss in Real-Life Cardiac Monitoring Applications"
- G. Vila**, C. Godin, S. Charbonnier and A. Campagne (2021), "Acute Stress Detection from Biosignals with Quality Management Strategy "

BREVETS DEPOSES

- G. Vila**, C. Godin, S. Charbonnier, A. Campagne (2019) : Procédé d'estimation de la qualité d'un signal de rythme cardiaque (N° FR3098390 - EP3763283 - US2021007674)
- G. Vila** & S. Charbonnier, C. Godin, S. Charbonnier, A. Campagne (2018) : Procédé de Détermination de l'Etat de Stress d'un Individu (N° FR3083690 - EP3593709 - US2020015729)

Bibliographie

- Acharya, U. R., Joseph, K. P., Kannathal, N., Lim, C. M., & Suri, J. S. (2006). Heart rate variability : A review. *Medical and Biological Engineering and Computing*, 44(12), 1031-1051. <https://doi.org/10.1007/s11517-006-0119-0>
- Agresti, A., & Coull, B. A. (1998). Approximate is Better than “Exact” for Interval Estimation of Binomial Proportions. *The American Statistician*, 52(2), 119-126. <https://doi.org/10.1080/00031305.1998.10480550>
- Aldecoa Sánchez del Río, B., Lopetegi, T., & Romero, I. (2011). Assessment of different methods to estimate electrocardiogram signal quality. *2011 Computing in Cardiology*, 609-612.
- Allen, J. (2007). Photoplethysmography and its application in clinical physiological measurement. *Physiological Measurement*, 28(3), R1-R39. <https://doi.org/10.1088/0967-3334/28/3/R01>
- Apple. (2018). *Apple Watch Series 4—Health & ECG*. Apple. <https://www.apple.com/apple-watch-series-4/health/>
- Apple. (2019). *Use the Breathe app*. Apple Support. <https://support.apple.com/en-us/HT206999>
- Averill, J. R., & Rosenn, M. (1972). Vigilant and nonvigilant coping strategies and psychophysiological stress reactions during the anticipation of electric shock. *Journal of Personality and Social Psychology*, 23(1), 128-141. <https://doi.org/10.1037/h0032758>
- Backs, R. W., & Seljos, K. A. (1994). Metabolic and cardiorespiratory measures of mental effort : The effects of level of difficulty in a working memory task. *International Journal of Psychophysiology*, 16(1), 57-68. [https://doi.org/10.1016/0167-8760\(94\)90042-6](https://doi.org/10.1016/0167-8760(94)90042-6)
- Behar, J., Oster, J., Li, Q., & Clifford, G. D. (2012). A single channel ECG quality metric. *2012 Computing in Cardiology*, 381-384.
- Benedek, M., & Kaernbach, C. (2010). A continuous measure of phasic electrodermal activity. *Journal of Neuroscience Methods*, 190(1), 80-91. <https://doi.org/10.1016/j.jneumeth.2010.04.028>
- Bernardo, J. M., Bayarri, M. J., Berger, J. O., Dawid, A. P., Heckerman, D., Smith, A. F. M., & West, M. (2007). Generative or discriminative ? Getting the best of both worlds. *Bayesian statistics*, 8(3), 3-24.
- Boccaletti, C., Castrica, F., Fabbri, G., & Santello, M. (2008). A Non-invasive Biopotential Electrode for the Correct Detection of Bioelectrical Currents. *Proceedings of the Sixth IASTED International Conference on Biomedical Engineering*, 353-358. <http://dl.acm.org/citation.cfm?id=1713360.1713434>
- Boucsein, W. (2012). *Electrodermal Activity*. Springer Science & Business Media.
- Bouten, C. V. C., Koekkoek, K. T. M., Verduin, M., Kodde, R., & Janssen, J. D. (1997). A triaxial accelerometer and portable data processing unit for the assessment of daily physical activity. *IEEE Transactions on Biomedical Engineering*, 44(3), 136-147. <https://doi.org/10.1109/10.554760>
- Brickenkamp, R., & Zillmer, E. (1998). *The d2 test of attention : Manual* (1. US ed., [Nachdr.]). Seattle : Hogrefe & Huber. <https://us.hogrefe.com/shop/d2-test-of-attention.html>
- Broek, E. L. van den, Janssen, J. H., Westerink, J. H. D. M., & Healey, J. A. (2009). *Prerequisites for Affective Signal Processing (ASP)* (P. Encarnacao & A. Veloso, Éd.; p. 426-433). INSTICC Press. <http://doc.utwente.nl/73169/>
- Brosschot, J. F. (2010). Markers of chronic stress : Prolonged physiological activation and (un)conscious perseverative cognition. *Neuroscience & Biobehavioral Reviews*, 35(1), 46-50. <https://doi.org/10.1016/j.neubiorev.2010.01.004>

- Buijs, R. M., van Eden, C. G., Goncharuk, V. D., & Kalsbeek, A. (2003). The biological clock tunes the organs of the body : Timing by hormones and the autonomic nervous system. *The Journal of Endocrinology*, *177*(1), 17-26.
- Bullinger, M., Naber, D., Pickar, D., Cohen, R. M., Kalin, N. H., Pert, A., & Bunney, W. E. (1984). Endocrine effects of the cold pressor test : Relationships to subjective pain appraisal and coping. *Psychiatry Research*, *12*(3), 227-233. [https://doi.org/10.1016/0165-1781\(84\)90028-3](https://doi.org/10.1016/0165-1781(84)90028-3)
- Butler, P. J., Green, J. A., Boyd, I. L., & Speakman, J. R. (2004). Measuring metabolic rate in the field : The pros and cons of the doubly labelled water and heart rate methods. *Functional Ecology*, *18*(2), 168-183. <https://doi.org/10.1111/j.0269-8463.2004.00821.x>
- C. Setz, B. Arnrich, J. Schumm, Marca, R. L., Tröster, G., & Ehlert, U. (2010). Discriminating Stress From Cognitive Load Using a Wearable EDA Device. *IEEE Transactions on Information Technology in Biomedicine*, *14*(2), 410-417. <https://doi.org/10.1109/TITB.2009.2036164>
- Calder, B. J., Phillips, L. W., & Tybout, A. M. (1982). The Concept of External Validity. *Journal of Consumer Research*, *9*(3), 240-244. <https://doi.org/10.1086/208920>
- Can, Y. S., Arnrich, B., & Ersoy, C. (2019). Stress Detection in Daily Life Scenarios Using Smart Phones and Wearable Sensors : A Survey. *Journal of biomedical informatics*, 103139.
- Cannon, W. B. (1929). Organization for physiological homeostasis. *Physiological reviews*, *9*(3), 399-431.
- Carre Technologies inc. (2012). *Hexoskin Smart Shirts—Cardiac, Respiratory, Sleep & Activity Metrics*. Carre Technologies Inc (Hexoskin). <https://www.hexoskin.com/>
- Carroll, D., Phillips, A. C., & Balanos, G. M. (2009). Metabolically exaggerated cardiac reactions to acute psychological stress revisited. *Psychophysiology*, *46*(2), 270-275. <https://doi.org/10.1111/j.1469-8986.2008.00762.x>
- Castaldo, R., Melillo, P., & Pecchia, L. (2015). Acute Mental Stress Detection via Ultra-short term HRV Analysis. In D. A. Jaffray (Éd.), *World Congress on Medical Physics and Biomedical Engineering, June 7-12, 2015, Toronto, Canada* (p. 1068-1071). Springer International Publishing. https://doi.org/10.1007/978-3-319-19387-8_260
- Chang, M., Yih, W., & Meek, C. (2008). Partitioned logistic regression for spam filtering. *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, 97-105. <https://doi.org/10.1145/1401890.1401907>
- Choi, J., Ahmed, B., & Gutierrez-Osuna, R. (2012). Development and Evaluation of an Ambulatory Stress Monitor Based on Wearable Sensors. *IEEE Transactions on Information Technology in Biomedicine*, *16*(2), 279-286. <https://doi.org/10.1109/TITB.2011.2169804>
- Chow, S.-C., Shao, J., & Wang, H. (2002). A NOTE ON SAMPLE SIZE CALCULATION FOR MEAN COMPARISONS BASED ON NONCENTRAL t-STATISTICS. *Journal of Biopharmaceutical Statistics*, *12*(4), 441-456. <https://doi.org/10.1081/BIP-120016229>
- Cinaz, B., Arnrich, B., La Marca, R., & Tröster, G. (2013). Monitoring of mental workload levels during an everyday life office-work scenario. *Personal and Ubiquitous Computing*, *17*(2), 229-239. <https://doi.org/10.1007/s00779-011-0466-1>
- Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A Global Measure of Perceived Stress. *Journal of Health and Social Behavior*, *24*(4), 385-396. <https://doi.org/10.2307/2136404>
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*(4), 281-302. <https://doi.org/10.1037/h0040957>
- Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems*, *2*(4), 303-314. <https://doi.org/10.1007/BF02551274>
- Dedovic, K., Duchesne, A., Andrews, J., Engert, V., & Pruessner, J. C. (2009). The brain and the stress axis : The neural correlates of cortisol regulation in response to stress. *NeuroImage*, *47*(3), 864-871. <https://doi.org/10.1016/j.neuroimage.2009.05.074>

- Dedovic, K., Renwick, R., Mahani, N. K., Engert, V., Lupien, S. J., & Pruessner, J. C. (2005). The Montreal Imaging Stress Task : Using functional imaging to investigate the effects of perceiving and processing psychosocial stress in the human brain. *Journal of Psychiatry and Neuroscience*, 30(5), 319-325.
- DeLongis, A., Coyne, J. C., Dakof, G., Folkman, S., & Lazarus, R. S. (1982). Relationship of daily hassles, uplifts, and major life events to health status. *Health psychology*, 1(2), 119.
- Dickerson, S. S., & Kemeny, M. E. (2004). Acute Stressors and Cortisol Responses : A Theoretical Integration and Synthesis of Laboratory Research. *Psychological Bulletin*, 130(3), 355-391. <https://doi.org/10.1037/0033-2909.130.3.355>
- Dohrenwend, B. P. (2006). Inventorying stressful life events as risk factors for psychopathology : Toward resolution of the problem of intracategory variability. *Psychological Bulletin*, 132(3), 477-495. <https://doi.org/10.1037/0033-2909.132.3.477>
- Ekman, P., Levenson, R. W., & Friesen, W. V. (1983). Autonomic Nervous System Activity Distinguishes Among Emotions. *Science*, 221, 1208-1210.
- Empatica. (2013). *E4 sensor*. Empatica. <https://www.empatica.com/research/e4>
- Empatica. (2014). *Have you done comparative studies or validation of the heart rate obtained from the E3/E4?* Empatica Support. <http://support.empatica.com/hc/en-us/articles/200293658-Have-you-done-comparative-studies-or-validation-of-the-heart-rate-obtained-from-the-E3-E4->
- Empatica. (2015a). *Embrace2 Seizure Monitoring | Embrace Watch*. Empatica. <https://www.empatica.com/embrace2>
- Empatica. (2015b). *Have you done comparative studies or validation on Electrodermal activity sensor?* Empatica Support. <http://support.empatica.com/hc/en-us/articles/203005295-Have-you-done-comparative-studies-or-validation-on-Electrodermal-activity-sensor->
- Empatica. (2019). *How is IBI.csv obtained?* Empatica Support. <http://support.empatica.com/hc/en-us/articles/201912319-How-is-IBI-csv-obtained->
- European Agency for Safety and Health at Work., TC OSH., Birkbeck College University of London (BBK)., Robert Gründler and Danny Flemming, DGUV., & Prevent. (2014). *Calculating the costs of work-related stress and psychosocial risks : Literature review*. Publications Office. <https://data.europa.eu/doi/10.2802/20493>
- European Foundation for the Improvement of Living and Working Conditions (Éd.). (2007). *Fourth European working conditions survey*. Office for Official Publ. of the European Communities [u.a.].
- Fairclough, S. H., & Houston, K. (2004). A metabolic measure of mental effort. *Biological Psychology*, 66(2), 177-190. <https://doi.org/10.1016/j.biopsycho.2003.10.001>
- Fairclough, S. H., & Mulder, L. J. M. (2011). Psychophysiological processes of mental effort investment. *How Motivation Affects Cardiovascular Response: Mechanisms and Applications*, 61-76.
- FitBit. (2019). *Inspire HR*. <https://www.fitbit.com/fr/shop/inspire>
- Folkman, S., & Lazarus, R. S. (1985). If it changes it must be a process : Study of emotion and coping during three stages of a college examination. *Journal of personality and social psychology*, 48(1), 150.
- Folkman, S., Lazarus, R. S., Dunkel-Schetter, C., DeLongis, A., & Gruen, R. J. (1986). Dynamics of a stressful encounter : Cognitive appraisal, coping, and encounter outcomes. *Journal of Personality and Social Psychology*, 50(5), 992-1003. <https://doi.org/10.1037/0022-3514.50.5.992>
- Fowles, D. C. (1988). Psychophysiology and Psychopathology : A Motivational Approach. *Psychophysiology*, 25(4), 373-391. <https://doi.org/10.1111/j.1469-8986.1988.tb01873.x>

- Frankenhaeuser, M. (1982). Challenge-control interaction as reflected in sympathetic-adrenal and pituitary-adrenal activity: Comparison between the sexes. *Scandinavian Journal of Psychology*, 23(S1), 158-164. <https://doi.org/10.1111/j.1467-9450.1982.tb00466.x>
- Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). « Effect size estimates : Current use, calculations, and interpretation »: Correction to Fritz et al. (2011). *Journal of Experimental Psychology: General*, 141(1), 30-30. <https://doi.org/10.1037/a0026092>
- Gaillard, A. W. K. (2001). Stress, Workload, and Fatigue as Three Biobehavioral States : A General Overview. *Stress, Workload, and Fatigue*, 623-639.
- Gaillard, A. W. K., & Wientjes, C. J. E. (1994). Mental load and work stress as two types of energy mobilization. *Work & Stress*, 8(2), 141-152. <https://doi.org/10.1080/02678379408259986>
- Gandhi, S., Baghini, M. S., & Mukherji, S. (2015). Mental stress assessment—A comparison between HRV based and respiration based techniques. *2015 Computing in Cardiology Conference (CinC)*, 1029-1032. <https://doi.org/10.1109/CIC.2015.7411089>
- Gibbons, J. D., & Chakraborti, S. (2003). *Nonparametric statistical inference* (4th ed., rev. expanded). Marcel Dekker.
- Gil, E., Orini, M., Bailón, R., Vergara, J. M., Mainardi, L., & Laguna, P. (2010). Photoplethysmography pulse rate variability as a surrogate measurement of heart rate variability during non-stationary conditions. *Physiological Measurement*, 31(9), 1271-1290. <https://doi.org/10.1088/0967-3334/31/9/015>
- Gjoreski, M., Gjoreski, H., Luštrek, M., & Gams, M. (2016). Continuous Stress Detection Using a Wrist Device : In Laboratory and Real Life. *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*, 1185-1193. <https://doi.org/10.1145/2968219.2968306>
- Gjoreski, M., Luštrek, M., Gams, M., & Gjoreski, H. (2017). Monitoring stress with a wrist device using context. *Journal of Biomedical Informatics*, 73, 159-170. <https://doi.org/10.1016/j.jbi.2017.08.006>
- Green, S. R., Kragel, P. A., Fecteau, M. E., & LaBar, K. S. (2014). Development and validation of an unsupervised scoring system (Autonamate) for skin conductance response analysis. *International journal of psychophysiology : official journal of the International Organization of Psychophysiology*, 91(3), 186-193. <https://doi.org/10.1016/j.ijpsycho.2013.10.015>
- Gruetzmann, A., Hansen, S., & Müller, J. (2007). Novel dry electrodes for ECG monitoring. *Physiological Measurement*, 28(11), 1375-1390. <https://doi.org/10.1088/0967-3334/28/11/005>
- Grundlehner, B., Brown, L., Penders, J., & Gyselinckx, B. (2009). The Design and Analysis of a Real-Time, Continuous Arousal Monitor. *2009 Sixth International Workshop on Wearable and Implantable Body Sensor Networks*, 156-161. <https://doi.org/10.1109/BSN.2009.21>
- Gurel, N. Z., Jung, H., Hersek, S., & Inan, O. T. (2019). Fusing Near-Infrared Spectroscopy With Wearable Hemodynamic Measurements Improves Classification of Mental Stress. *IEEE Sensors Journal*, 19(19), 8522-8531. <https://doi.org/10.1109/JSEN.2018.2872651>
- Guyon, I., & Elisseeff, A. (2003). An Introduction to Variable and Feature Selection. *Journal of Machine Learning Research*, 3(Mar), 1157-1182.
- H2CARE. (2019). *H2—Blood Pressure Monitor*. <https://h2care.com/18>
- Hackney, A. C. (2006). Stress and the neuroendocrine system : The role of exercise as a stressor and modifier of stress. *Expert review of endocrinology & metabolism*, 1(6), 783-792. <https://doi.org/10.1586/17446651.1.6.783>
- Hakimi, N., & Setarehdan, S. K. (2018). Stress assessment by means of heart rate derived from functional near-infrared spectroscopy. *Journal of Biomedical Optics*, 23(11), 115001. <https://doi.org/10.1117/1.JBO.23.11.115001>

- Han, H. J., Labbaf, S., Borelli, J. L., Dutt, N., & Rahmani, A. M. (2020). Objective stress monitoring based on wearable sensors in everyday settings. *Journal of Medical Engineering & Technology*, 44(4), 177-189. <https://doi.org/10.1080/03091902.2020.1759707>
- Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index) : Results of Empirical and Theoretical Research. In P. A. Hancock & N. Meshkati (Éds.), *Advances in Psychology* (Vol. 52, p. 139-183). North-Holland. [https://doi.org/10.1016/S0166-4115\(08\)62386-9](https://doi.org/10.1016/S0166-4115(08)62386-9)
- Harvey, A., Nathens, A. B., Bandiera, G., & LeBlanc, V. R. (2010). Threat and challenge : Cognitive appraisal and stress responses in simulated trauma resuscitations. *Medical Education*, 44(6), 587-594. <https://doi.org/10.1111/j.1365-2923.2010.03634.x>
- Hastie, T., Tibshirani, R., & Friedman, J. H. (2009). *The elements of statistical learning : Data mining, inference, and prediction* (2nd ed). Springer.
- Hazanov, O., Kaiser, S., & Stephane, W. (2009). Stress et coping : Un état des lieux. In *Traité de psychologie des émotions*. Paris, France : Dunod.
- Healey, J. A., & Picard, R. W. (2005). Detecting stress during real-world driving tasks using physiological sensors. *IEEE Transactions on Intelligent Transportation Systems*, 6(2), 156-166. <https://doi.org/10.1109/TITS.2005.848368>
- Hellhammer, D. H., Wüst, S., & Kudielka, B. M. (2009). Salivary cortisol as a biomarker in stress research. *Psychoneuroendocrinology*, 34(2), 163-171. <https://doi.org/10.1016/j.psyneuen.2008.10.026>
- Hernandez, J., Morris, R. R., & Picard, R. W. (2011). Call Center Stress Recognition with Person-Specific Models. *Affective Computing and Intelligent Interaction*, 125-134. https://doi.org/10.1007/978-3-642-24600-5_16
- Hjortskov, N., Garde, A. H., Ørbæk, P., & Hansen, Å. M. (2004). Evaluation of salivary cortisol as a biomarker of self-reported mental stress in field studies. *Stress and Health*, 20(2), 91-98. <https://doi.org/10.1002/smi.1000>
- Hjortskov, N., Rissén, D., Blangsted, A. K., Fallentin, N., Lundberg, U., & Sjøgaard, K. (2004). The effect of mental stress on heart rate variability and blood pressure during computer work. *European Journal of Applied Physiology*, 92(1), 84-89. <https://doi.org/10.1007/s00421-004-1055-z>
- Hockey. (1997). Compensatory control in the regulation of human performance under stress and high workload : A cognitive-energetical framework. *Biological Psychology*, 45(1), 73-93. [https://doi.org/10.1016/S0301-0511\(96\)05223-4](https://doi.org/10.1016/S0301-0511(96)05223-4)
- Hou, X., Liu, Y., Sourina, O., Tan, Y. R. E., Wang, L., & Mueller-Wittig, W. (2015). EEG Based Stress Monitoring. *2015 IEEE International Conference on Systems, Man, and Cybernetics*, 3110-3115. <https://doi.org/10.1109/SMC.2015.540>
- Hovsepian, K., al'Absi, M., Ertin, E., Kamarck, T., Nakajima, M., & Kumar, S. (2015). *Stress : Towards a Gold Standard for Continuous Stress Assessment in the Mobile Environment*. 493-504. <https://doi.org/10.1145/2750858.2807526>
- Jandorf, L., Deblinger, E., Neale, J. M., & Stone, A. A. (1986). Daily Versus Major Life Events as Predictors of Symptom Frequency : A Replication Study. *The Journal of General Psychology*, 113(3), 205-218. <https://doi.org/10.1080/00221309.1986.9711031>
- Johnstone, J. A., Ford, P. A., Hughes, G., Watson, T., Mitchell, A. C. S., & Garrett, A. T. (2012). Field Based Reliability and Validity of the Bioharness™ Multivariable Monitoring Device. *Journal of Sports Science & Medicine*, 11(4), 643-652.
- Julious, S. A. (2004). Sample sizes for clinical trials with Normal data. *Statistics in Medicine*, 23(12), 1921-1986. <https://doi.org/10.1002/sim.1783>
- Kaelbling, L. P., Littman, M. L., & Moore, A. W. (1996). Reinforcement Learning : A Survey. *J. Artif. Intell. Res.* <https://cds.cern.ch/record/570063>

- Kalimeri, K., & Saitis, C. (2016). Exploring Multimodal Biosignal Features for Stress Detection during Indoor Mobility. *ResearchGate*. 18th ACM International Conference on Multimodal Interaction. <https://doi.org/10.1145/2993148.2993159>
- Kanner, A. D., Coyne, J. C., Schaefer, C., & Lazarus, R. S. (1981). Comparison of two modes of stress measurement : Daily hassles and uplifts versus major life events. *Journal of Behavioral Medicine*, 4(1), 1-39. <https://doi.org/10.1007/BF00844845>
- Karantonis, D. M., Narayanan, M. R., Mathie, M., Lovell, N. H., & Celler, B. G. (2006). Implementation of a real-time human movement classifier using a triaxial accelerometer for ambulatory monitoring. *IEEE transactions on information technology in biomedicine*, 10(1), 156-167.
- Karasek, R., & Theorell, T. (1990). *Healthy work : Stress, productivity, and the reconstruction of working life*. Basic Books.
- Karlsson, M., Hörnsten, R., Rydberg, A., & Wiklund, U. (2012). Automatic filtering of outliers in RR intervals before analysis of heart rate variability in Holter recordings : A comparison with carefully edited data. *Biomedical engineering online*, 11(1), 2.
- Kataoka, H., Kano, H., Yoshida, H., Saijo, A., Yasuda, M., & Osumi, M. (1998). Development of a skin temperature measuring system for non-contact stress evaluation. *Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Vol.20 Biomedical Engineering Towards the Year 2000 and Beyond (Cat. No.98CH36286)*, 2, 940-943 vol.2. <https://doi.org/10.1109/IEMBS.1998.745598>
- Kemper, K. J., Hamilton, C., & Atkinson, M. (2007). Heart rate variability : Impact of differences in outlier identification and management strategies on common measures in three clinical populations. *Pediatric research*, 62(3), 337.
- Kennedy, H. L. (2006). The History, Science, and Innovation of Holter Technology. *Annals of Noninvasive Electrocardiology*, 11(1), 85-94. <https://doi.org/10.1111/j.1542-474X.2006.00067.x>
- Kerassidis, S. (1994). Is palmar and plantar sweating thermoregulatory? *Acta Physiologica Scandinavica*, 152(3), 259-263. <https://doi.org/10.1111/j.1748-1716.1994.tb09805.x>
- Kilpatrick, D. G. (1972). Differential Responsiveness of Two Electrodermal Indices to Psychological Stress and Performance of a Complex Cognitive Task. *Psychophysiology*, 9(2), 218-226. <https://doi.org/10.1111/j.1469-8986.1972.tb00756.x>
- Kim, H.-G., Cheon, E.-J., Bai, D.-S., Lee, Y. H., & Koo, B.-H. (2018). Stress and Heart Rate Variability : A Meta-Analysis and Review of the Literature. *Psychiatry Investigation*, 15(3), 235-245. <https://doi.org/10.30773/pi.2017.08.17>
- Kirschbaum, C., Pirke, K.-M., & Hellhammer, D. H. (1993). The 'Trier Social Stress Test' – A Tool for Investigating Psychobiological Stress Responses in a Laboratory Setting. *Neuropsychobiology*, 28(1-2), 76-81. <https://doi.org/10.1159/000119004>
- Knight, J. F., & Baber, C. (2005). A Tool to Assess the Comfort of Wearable Computers. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 47(1), 77-91. <https://doi.org/10.1518/0018720053653875>
- Kocielnik, R., Sidorova, N., Maggi, F. M., Ouwerkerk, M., & Westerink, J. H. D. M. (2013). Smart technologies for long-term stress monitoring at work. *Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems*, 53-58. <https://doi.org/10.1109/CBMS.2013.6627764>
- Kohavi, R., & Sommerfield, D. (1995). Feature Subset Selection Using the Wrapper Method : Overfitting and Dynamic Search Space Topology. *KDD*, 192-197.
- Kreibig, S. D. (2010). Autonomic nervous system activity in emotion : A review. *Biological Psychology*, 84(3), 394-421. <https://doi.org/10.1016/j.biopsycho.2010.03.010>

- Lakens, D. (2017). Equivalence Tests : A Practical Primer for t Tests, Correlations, and Meta-Analyses. *Social Psychological and Personality Science*, 8(4), 355-362. <https://doi.org/10.1177/1948550617697177>
- Lakens, D., Scheel, A. M., & Isager, P. M. (2018). Equivalence Testing for Psychological Research : A Tutorial. *Advances in Methods and Practices in Psychological Science*, 1(2), 259-269. <https://doi.org/10.1177/2515245918770963>
- Larousse, É. (s. d.). *Définitions : Stress - Dictionnaire de français Larousse*. Consulté 28 mai 2019, à l'adresse <https://www.larousse.fr/dictionnaires/francais/stress/74848>
- Larsson, S.-E., Larsson, R., Zhang, Q., Cai, H., & Ake Oberg, P. (1995). Effects of psychophysiological stress on trapezius muscles blood flow and electromyography during static load. *European Journal of Applied Physiology and Occupational Physiology*, 71(6), 493-498. <https://doi.org/10.1007/BF00238550>
- Lazarus, R. S., DeLongis, A., Folkman, S., & Gruen, R. (1985). Stress and adaptational outcomes : The problem of confounded measures. *American Psychologist*, 770-779.
- Lazarus, R. S., & Folkman, S. (1987). Transactional theory and research on emotions and coping. *European Journal of Personality*, 1(3), 141-169. <https://doi.org/10.1002/per.2410010304>
- Lesage, F.-X., Berjot, S., & Deschamps, F. (2012). Clinical stress assessment using a visual analogue scale. *Occupational Medicine*, 62(8), 600-605. <https://doi.org/10.1093/occmed/kqs140>
- Leutheuser, H., Heyde, C., Roecker, K., Gollhofer, A., & Eskofier, B. M. (2017). Reference-Free Adjustment of Respiratory Inductance Plethysmography for Measurements during Physical Exercise. *IEEE Transactions on Biomedical Engineering*, 64(12), 2836-2846. <https://doi.org/10.1109/TBME.2017.2675941>
- Li, Q., & Clifford, G. D. (2012). Dynamic time warping and machine learning for signal quality assessment of pulsatile signals. *Physiological Measurement*, 33(9), 1491-1501. <https://doi.org/10.1088/0967-3334/33/9/1491>
- Li, Y., Wang, Z., Zhang, L., Yang, X., & Song, J. (2014). Characters available in photoplethysmogram for blood pressure estimation : Beyond the pulse transit time. *Australasian Physical & Engineering Sciences in Medicine*, 37(2), 367-376. <https://doi.org/10.1007/s13246-014-0269-6>
- Linden, W. von der. (1987). A microanalysis of autonomic activity during human speech. *Psychosomatic medicine*, 49(6), 562-578. <https://doi.org/10.1097/00006842-198711000-00002>
- Lomb, N. R. (1976). Least-squares frequency analysis of unequally spaced data. *Astrophysics and Space Science*, 39(2), 447-462. <https://doi.org/10.1007/BF00648343>
- Mandrick, K., Peysakhovich, V., Rémy, F., Lepron, E., & Causse, M. (2016). Neural and psychophysiological correlates of human performance under stress and high mental workload. *ResearchGate*, 121. <https://doi.org/10.1016/j.biopsycho.2016.10.002>
- Mara, C. A., & Cribbie, R. A. (2012). Paired-Samples Tests of Equivalence. *Communications in Statistics - Simulation and Computation*, 41(10), 1928-1943. <https://doi.org/10.1080/03610918.2011.626545>
- Marteau, T. M., & Bekker, H. (1992). The development of a six-item short-form of the state scale of the Spielberger State—Trait Anxiety Inventory (STAI). *British Journal of Clinical Psychology*, 31(3), 301-306. <https://doi.org/10.1111/j.2044-8260.1992.tb00997.x>
- Martins-Pinge, M. C. (2011). Cardiovascular and autonomic modulation by the central nervous system after aerobic exercise training. *Brazilian Journal of Medical and Biological Research*, 44(9), 848-854. <https://doi.org/10.1590/S0100-879X2011007500102>
- McEwen, B. S. (2000). The neurobiology of stress : From serendipity to clinical relevance. *Brain research*, 886(1), 172-189.
- McEwen, B. S. (2005). Stressed or stressed out : What is the difference? *Journal of Psychiatry and Neuroscience*, 30(5), 315.

- Mehler, B., Reimer, B., Coughlin, J. F., & Dusek, J. A. (2009). Impact of Incremental Increases in Cognitive Workload on Physiological Arousal and Performance in Young Adult Drivers. *Transportation Research Record*, 2138(1), 6-12. <https://doi.org/10.3141/2138-02>
- Melillo, P., Formisano, C., Bracale, U., & Pecchia, L. (2013). Classification tree for real-life stress detection using linear Heart Rate Variability analysis. Case study : Students under stress due to university examination. *World Congress on Medical Physics and Biomedical Engineering May 26-31, 2012, Beijing, China*, 477-480. https://doi.org/10.1007/978-3-642-29305-4_126
- Melomind. (2019). *Melomind*. Melomind. <https://www.melomind.com>
- Mohino-Herranz, I., Gil-Pita, R., Ferreira, J., Rosa-Zurera, M., & Seoane, F. (2015). Assessment of Mental, Emotional and Physical Stress through Analysis of Physiological Signals Using Smartphones. *Sensors (Basel, Switzerland)*, 15(10), 25607-25627. <https://doi.org/10.3390/s151025607>
- Mozos, O. M., Sandulescu, V., Andrews, S., Ellis, D., Bellotto, N., Dobrescu, R., & Ferrandez, J. M. (2016). Stress Detection Using Wearable Physiological and Sociometric Sensors. *International Journal of Neural Systems*, 27(02), 1650041. <https://doi.org/10.1142/S0129065716500416>
- Muaremi, A., Arnrich, B., & Tröster, G. (2013). Towards Measuring Stress with Smartphones and Wearable Devices During Workday and Sleep. *BioNanoScience*, 3(2), 172-183. <https://doi.org/10.1007/s12668-013-0089-2>
- Mulder, L. J. M. (1992). Measurement and analysis methods of heart rate and respiration for use in applied environments. *Biological Psychology*, 34(2-3), 205-236. [https://doi.org/10.1016/0301-0511\(92\)90016-N](https://doi.org/10.1016/0301-0511(92)90016-N)
- Neurosky. (2014). *EEG Sensors—EEG Headsets | NeuroSky*. <http://neurosky.com/biosensors/eeg-sensor/biosensors/>
- Ng, A. Y., & Jordan, M. I. (2002). *On Discriminative vs. Generative Classifiers : A comparison of logistic regression and naive Bayes*. 841-848.
- Ollander, S., Godin, C., Campagne, A., & Charbonnier, S. (2016). A comparison of wearable and stationary sensors for stress detection. *Systems, Man, and Cybernetics (SMC), 2016 IEEE International Conference on*, 004362-004366. <http://ieeexplore.ieee.org/abstract/document/7844917/>
- Orphanidou, C., Bonnici, T., Charlton, P., Clifton, D., Vallance, D., & Tarassenko, L. (2015). Signal-Quality Indices for the Electrocardiogram and Photoplethysmogram : Derivation and Applications to Wireless Monitoring. *IEEE Journal of Biomedical and Health Informatics*, 19(3), 832-838. <https://doi.org/10.1109/JBHI.2014.2338351>
- Orphanidou, C., & Drobnjak, I. (2017). Quality Assessment of Ambulatory ECG Using Wavelet Entropy of the HRV Signal. *IEEE Journal of Biomedical and Health Informatics*, 21(5), 1216-1223. <https://doi.org/10.1109/JBHI.2016.2615316>
- Paas, F. G. W. C., & Van Merriënboer, J. J. G. (1994). Instructional control of cognitive load in the training of complex cognitive tasks. *Educational Psychology Review*, 6(4), 351-371. <https://doi.org/10.1007/BF02213420>
- Paas, F. G. W. C., van Merriënboer, J. J. G., & Adam, J. J. (1994). Measurement of Cognitive Load in Instructional Research. *Perceptual and Motor Skills*, 79(1), 419-430. <https://doi.org/10.2466/pms.1994.79.1.419>
- Pan, J., & Tompkins, W. J. (1985). A Real-Time QRS Detection Algorithm. *IEEE Transactions on Biomedical Engineering, BME-32(3)*, 230-236. <https://doi.org/10.1109/TBME.1985.325532>
- Papini, G. B., Fonseca, P., Aubert, X. L., Overeem, S., Bergmans, J. W. M., & Vullings, R. (2017). Photoplethysmography beat detection and pulse morphology quality assessment for signal reliability estimation. *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 117-120. <https://doi.org/10.1109/EMBC.2017.8036776>

- Periyasamy, V., Pramanik, M., & Ghosh, P. K. (2017). Review on Heart-Rate Estimation from Photoplethysmography and Accelerometer Signals During Physical Exercise. *Journal of the Indian Institute of Science*, 97(3), 313-324. <https://doi.org/10.1007/s41745-017-0037-1>
- Peters, M. L., Godaert, G. L. R., Ballieux, R. E., van Vliet, M., Willemsen, J. J., Sweep, F. C. G. J., & Heijnen, C. J. (1998). Cardiovascular and endocrine responses to experimental stress : Effects of mental effort and controllability. *Psychoneuroendocrinology*, 23(1), 1-17. [https://doi.org/10.1016/S0306-4530\(97\)00082-6](https://doi.org/10.1016/S0306-4530(97)00082-6)
- Picard, R. W. (1997). *Affective Computing*. Citeseer. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.47.9195&rep=rep1&type=pdf>
- Picard, R. W., Fedor, S., & Ayzenberg, Y. (2016). Multiple arousal theory and daily-life electrodermal activity asymmetry. *Emotion Review*, 8(1), 62-75.
- Pickering Thomas G., Hall John E., Appel Lawrence J., Falkner Bonita E., Graves John, Hill Martha N., Jones Daniel W., Kurtz Theodore, Sheps Sheldon G., & Roccella Edward J. (2005). Recommendations for Blood Pressure Measurement in Humans and Experimental Animals. *Circulation*, 111(5), 697-716. <https://doi.org/10.1161/01.CIR.0000154900.76284.F6>
- Plarre, K., Raij, A., Guha, S., al'Absi, M., Ertin, E., & Kumar, S. (2010). Automated detection of sensor detachments for physiological sensing in the wild. *Wireless Health 2010 on - WH '10*, 216. <https://doi.org/10.1145/1921081.1921119>
- Porges, S. W. (1995). Cardiac vagal tone : A physiological index of stress. *Neuroscience & Biobehavioral Reviews*, 19(2), 225-233. [https://doi.org/10.1016/0149-7634\(94\)00066-A](https://doi.org/10.1016/0149-7634(94)00066-A)
- Reisman, S. (1997). Measurement of physiological stress. *Proceedings of the IEEE 23rd Northeast Bioengineering Conference*, 21-23. <https://doi.org/10.1109/NEBC.1997.594939>
- Rimmele, U., Seiler, R., Marti, B., Wirtz, P. H., Ehlert, U., & Heinrichs, M. (2009). The level of physical activity affects adrenal and cardiovascular reactivity to psychosocial stress. *Psychoneuroendocrinology*, 34(2), 190-198. <https://doi.org/10.1016/j.psyneuen.2008.08.023>
- Robert J. Hockey, G. (1997). Compensatory control in the regulation of human performance under stress and high workload : A cognitive-energetical framework. *Biological Psychology*, 45(1), 73-93. [https://doi.org/10.1016/S0301-0511\(96\)05223-4](https://doi.org/10.1016/S0301-0511(96)05223-4)
- Rowell, L. B. (1993). *Human Cardiovascular Control*. Oxford University Press.
- Roy, R. N., Bonnet, S., Charbonnier, S., & Campagne, A. (2016). Efficient Workload Classification based on Ignored Auditory Probes : A Proof of Concept. *Frontiers in Human Neuroscience*, 10. <https://doi.org/10.3389/fnhum.2016.00519>
- Roy, R. N., Charbonnier, S., Campagne, A., & Bonnet, S. (2016). Efficient mental workload estimation using task-independent EEG features. *Journal of Neural Engineering*, 13(2), 026019. <https://doi.org/10.1088/1741-2560/13/2/026019>
- Rubinstein, Y. D., Hastie, T., & others. (1997). Discriminative vs Informative Learning. *KDD*, 5, 49-53.
- Russell, J. A., & Mehrabian, A. (1977). Evidence for a three-factor theory of emotions. *Journal of Research in Personality*, 11(3), 273-294. [https://doi.org/10.1016/0092-6566\(77\)90037-X](https://doi.org/10.1016/0092-6566(77)90037-X)
- Sakri, O., Godin, C., Vila, G., Labyt, E., Charbonnier, S., & Campagne, A. (2018). A Multi-User Multi-Task Model For Stress Monitoring From Wearable Sensors. *2018 21st International Conference on Information Fusion (FUSION)*, 761-766. <https://doi.org/10.23919/ICIF.2018.8455378>
- Salazar-Ramirez, A., Irigoyen, E., Martinez, R., & Zalabarria, U. (2018). An enhanced fuzzy algorithm based on advanced signal processing for identification of stress. *Neurocomputing*, 271, 48-57. <https://doi.org/10.1016/j.neucom.2016.08.153>
- Samsung. (2017). *Samsung Health*. The Official Samsung Galaxy Site. <https://www.samsung.com/global/galaxy/apps/samsung-health/>
- Samsung. (2018). *Galaxy Watch*. Samsung fr. <http://www.samsung.com/fr/wearables/galaxy-watch/>

- Sander, D., & Scherer, K. R. (2009). *Traité de psychologie des émotions*. Dunod.
- Sandulescu, V., Andrews, S., Ellis, D., Bellotto, N., & Mozos, O. M. (2015). Stress Detection Using Wearable Physiological Sensors. In J. M. F. Vicente, J. R. Álvarez-Sánchez, F. de la P. López, F. J. Toledo-Moreo, & H. Adeli (Éds.), *Artificial Computation in Biology and Medicine* (p. 526-532). Springer International Publishing. https://doi.org/10.1007/978-3-319-18914-7_55
- Scargle, J. D. (1982). Studies in astronomical time series analysis. II - Statistical aspects of spectral analysis of unevenly spaced data. *The Astrophysical Journal*, 263, 835. <https://doi.org/10.1086/160554>
- Schachter, S., & Singer, J. (1962). Cognitive, social, and physiological determinants of emotional state. *Psychological Review*, 69(5), 379-399. <https://doi.org/10.1037/h0046234>
- Schäfer, A., & Vagedes, J. (2013). How accurate is pulse rate variability as an estimate of heart rate variability? : A review on studies comparing photoplethysmographic technology with an electrocardiogram. *International Journal of Cardiology*, 166(1), 15-29. <https://doi.org/10.1016/j.ijcard.2012.03.119>
- Schmeck, A., Opfermann, M., van Gog, T., Paas, F., & Leutner, D. (2015). Measuring cognitive load with subjective rating scales during problem solving : Differences between immediate and delayed ratings. *Instructional Science*, 43(1), 93-114. <https://doi.org/10.1007/s11251-014-9328-3>
- Schwabe, L., Haddad, L., & Schachinger, H. (2008). HPA axis activation by a socially evaluated cold-pressor test. *Psychoneuroendocrinology*, 33(6), 890-895. <https://doi.org/10.1016/j.psyneuen.2008.03.001>
- Sedghamiz, H. (2014). *Matlab Implementation of Pan Tompkins ECG QRS detector*. <https://doi.org/10.13140/RG.2.2.14202.59841>
- Selye, H. (1936). A Syndrome Produced by Diverse Nocuous Agents. *The Journal of Neuropsychiatry and Clinical Neurosciences*, 10(2), 230a-2231. <https://doi.org/10.1176/jnp.10.2.230a>
- Setz, Arnrich, Schumm, Marca, R. L., Tröster, G., & Ehlert, U. (2010). Discriminating Stress From Cognitive Load Using a Wearable EDA Device. *IEEE Transactions on Information Technology in Biomedicine*, 14(2), 410-417. <https://doi.org/10.1109/TITB.2009.2036164>
- Sharma, N., & Gedeon, T. (2012). Objective measures, sensors and computational techniques for stress recognition and classification : A survey. *Computer Methods and Programs in Biomedicine*, 108(3), 1287-1301. <https://doi.org/10.1016/j.cmpb.2012.07.003>
- Sherwood, A., Allen, M. T., Obrist, P. A., & Langer, A. W. (1986). Evaluation of Beta-Adrenergic Influences on Cardiovascular and Metabolic Adjustments to Physical and Psychological Stress. *Psychophysiology*, 23(1), 89-104. <https://doi.org/10.1111/j.1469-8986.1986.tb00602.x>
- Sherwood(Chair), A., Allen, M. T., Fahrenberg, J., Kelsey, R. M., Lovallo, W. R., & van Doornen, L. J. P. (1990). Methodological Guidelines for Impedance Cardiography. *Psychophysiology*, 27(1), 1-23. <https://doi.org/10.1111/j.1469-8986.1990.tb02171.x>
- Shi, Y., Nguyen, M. H., Blitz, P., French, B., Fisk, S., De la Torre, F., Smailagic, A., Siewiorek, D. P., al'Absi, M., & Ertin, E. (2010). Personalized stress detection from physiological measurements. *International symposium on quality of life technology*, 28-29.
- Shieh, G., Jan, S.-L., & Randles, R. H. (2007). Power and sample size determinations for the Wilcoxon signed-rank test. *Journal of Statistical Computation and Simulation*, 77(8), 717-724. <https://doi.org/10.1080/10629360600635245>
- Shouldice, R. B., Heneghan, C., Petres, G., Zaffaroni, A., Boyle, P., McNicholas, W., & Chazal, P. de. (2010). Real time breathing rate estimation from a non contact biosensor. *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, 630-633. <https://doi.org/10.1109/IEMBS.2010.5627275>
- Siegrist, J., & Peter, R. (1994). Job stressors and coping characteristics in work-related disease : Issues of validity. *Work & Stress*, 8(2), 130-140. <https://doi.org/10.1080/02678379408259985>

- Sierra, A., Avila, C. S., Casanova, J. G., & Pozo, G. B. del. (2011). A Stress-Detection System Based on Physiological Signals and Fuzzy Logic. *IEEE Transactions on Industrial Electronics*, *58*(10), 4857-4865. <https://doi.org/10.1109/TIE.2010.2103538>
- Sin, J.-U., Yun, Y.-R., & Park, S.-J. (1999). Estimation of Stress Status Using Bio-signals and Fuzzy Theory. *Journal of the Ergonomics Society of Korea*, *18*(1), 121-131.
- Sinex, J. E. (1999). Pulse oximetry : Principles and limitations. *The American Journal of Emergency Medicine*, *17*(1), 59-66. [https://doi.org/10.1016/S0735-6757\(99\)90019-0](https://doi.org/10.1016/S0735-6757(99)90019-0)
- Skoluda, N., Strahler, J., Schlotz, W., Niederberger, L., Marques, S., Fischer, S., Thoma, M. V., Spoerri, C., Ehlert, U., & Nater, U. M. (2015). Intra-individual psychological and physiological responses to acute laboratory stressors of different intensity. *Psychoneuroendocrinology*, *51*, 227-236. <https://doi.org/10.1016/j.psyneuen.2014.10.002>
- Sloan, R. P., Korten, J. B., & Myers, M. M. (1991). Components of heart rate reactivity during mental arithmetic with and without speaking. *Physiology & Behavior*, *50*(5), 1039-1045. [https://doi.org/10.1016/0031-9384\(91\)90434-P](https://doi.org/10.1016/0031-9384(91)90434-P)
- Smith, A. D. H., Crabtree, D. R., Bilzon, J. L. J., & Walsh, N. P. (2010). The validity of wireless iButtons® and thermistors for human skin temperature measurement. *Physiological Measurement*, *31*(1), 95-114. <https://doi.org/10.1088/0967-3334/31/1/007>
- Smith, G. T. (2005). On Construct Validity : Issues of Method and Measurement. *Psychological Assessment*, *17*(4), 396-408. <https://doi.org/10.1037/1040-3590.17.4.396>
- Smyth, J., Ockenfels, M. C., Porter, L., Kirschbaum, C., Hellhammer, D. H., & Stone, A. A. (1998). Stressors And Mood Measured on a Momentary Basis are Associated With Salivary Cortisol Secretion. *Psychoneuroendocrinology*, *23*(4), 353-370. [https://doi.org/10.1016/S0306-4530\(98\)00008-0](https://doi.org/10.1016/S0306-4530(98)00008-0)
- Spagnolli, A., Guardigli, E., Orso, V., Varotto, A., & Gamberini, L. (2014). Measuring User Acceptance of Wearable Symbiotic Devices : Validation Study Across Application Scenarios. In G. Jacucci, L. Gamberini, J. Freeman, & A. Spagnolli (Éds.), *Symbiotic Interaction* (Vol. 8820, p. 87-98). Springer International Publishing. https://doi.org/10.1007/978-3-319-13500-7_7
- Spielberger, C. D. (1983). *Manual for the State-Trait Anxiety Inventory STAI (Form Y)* (« Self-Evaluation Questionnaire »). <http://ubir.buffalo.edu/xmlui/handle/10477/1873>
- SpireHealth. (2018). *Spire Health : Clinical-Grade Health Monitoring and Insights*. <https://spirehealth.com/>
- Staal, M. A. (2004). *Stress, Cognition, and Human Performance : A Literature Review and Conceptual Framework*. <https://ntrs.nasa.gov/search.jsp?R=20060017835>
- Stop Sleep France. (2013). *Bague StopSleep, solution anti-somnolence*. <https://www.stopsleep.fr/>
- Stroop, J.R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643-662. <https://doi.org/10.1037/h0054651>
- Sul, A., Shin, J., Lee, C., Yoon, Y., & Principe, J. (2002). Evaluation of stress reactivity and recovery using biosignals and fuzzy theory. *Proceedings of the Second Joint 24th Annual Conference and the Annual Fall Meeting of the Biomedical Engineering Society [Engineering in Medicine and Biology, 1, 32-33 vol.1]*. <https://doi.org/10.1109/IEMBS.2002.1134373>
- Sun, F.-T., Kuo, C., Cheng, H.-T., Buthpitiya, S., Collins, P., & Griss, M. (2012). Activity-Aware Mental Stress Detection Using Physiological Sensors. In M. Gris & G. Yang (Éds.), *Mobile Computing, Applications, and Services* (p. 282-301). Springer. https://doi.org/10.1007/978-3-642-29336-8_16
- Sutton, C., Sindelar, M., & McCallum, A. (2006). Reducing weight undertraining in structured discriminative learning. *Proceedings of the main conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics*, 89-95. <https://doi.org/10.3115/1220835.1220847>

- Sweeney, K. T., Kearney, D., Ward, T. E., Coyle, S., & Diamond, D. (2013). Employing ensemble empirical mode decomposition for artifact removal : Extracting accurate respiration rates from ECG data during ambulatory activity. *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 977-980. <https://doi.org/10.1109/EMBC.2013.6609666>
- Sweller, J., van Merriënboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive Architecture and Instructional Design. *Educational Psychology Review*, 10(3), 251-296. <https://doi.org/10.1023/A:1022193728205>
- Sztajzel, J. (2004). Heart rate variability : A noninvasive electrocardiographic method to measure the autonomic nervous system. *Swiss medical weekly*, 134(35-36), 514-522.
- Taelman, J., Vandeput, S., Gligorijević, I., Spaepen, A., & Huffel, S. V. (2011). Time-frequency heart rate variability characteristics of young adults during physical, mental and combined stress in laboratory environment. *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 1973-1976. <https://doi.org/10.1109/IEMBS.2011.6090556>
- Taylor, E. W., Jordan, D., & Coote, J. H. (1999). Central Control of the Cardiovascular and Respiratory Systems and Their Interactions in Vertebrates. *Physiological Reviews*, 79(3), 855-916. <https://doi.org/10.1152/physrev.1999.79.3.855>
- Thomas, S. S., Nathan, V., Zong, C., Soundarapandian, K., Shi, X., & Jafari, R. (2016). BioWatch : A Noninvasive Wrist-Based Blood Pressure Monitor That Incorporates Training Techniques for Posture and Subject Variability. *IEEE Journal of Biomedical and Health Informatics*, 20(5), 1291-1300. <https://doi.org/10.1109/JBHI.2015.2458779>
- Tobii. (2015, juin 25). *Tobii Pro Glasses 2 wearable eye tracker* [Information]. <https://www.tobii.com/product-listing/tobii-pro-glasses-2/>
- Tomaka, J., & Blascovich, J. (1994). Effects of justice beliefs on cognitive appraisal of and subjective physiological, and behavioral responses to potential stress. *Journal of personality and social psychology*, 67(4), 732.
- Tomaka, J., Blascovich, J., Kelsey, R. M., & Leitten, C. L. (1993). Subjective, physiological, and behavioral effects of threat and challenge appraisal. *Journal of Personality and Social Psychology*, 65(2), 248-260. <https://doi.org/10.1037/0022-3514.65.2.248>
- Tomaka, J., Blascovich, J., Kibler, J., & Ernst, J. M. (1997). Cognitive and physiological antecedents of threat and challenge appraisal. *Journal of Personality and Social Psychology*, 73(1), 63-72. <https://doi.org/10.1037/0022-3514.73.1.63>
- Tomaka, J., Blascovich, J., & Swart, L. (1994). Effects of vocalization on cardiovascular and electrodermal responses during mental arithmetic. *International Journal of Psychophysiology*, 18(1), 23-33.
- Tomaka, J., Palacios, R., Schneider, K. T., Colotla, M., Concha, J. B., & Herral, M. M. (1999). Assertiveness predicts threat and challenge reactions to potential stress among women. *Journal of Personality and Social Psychology*, 76(6), 1008.
- Tomaka, J., & Palacios-Esquivel, R. L. (1997). Motivational systems and stress-related cardiovascular reactivity. *Motivation and Emotion*, 21(4), 275-296.
- Tulen, J. H. M., Moleman, P., van Steenis, H. G., & Boomsma, F. (1989). Characterization of stress reactions to the Stroop Color Word Test. *Pharmacology Biochemistry and Behavior*, 32(1), 9-15. [https://doi.org/10.1016/0091-3057\(89\)90204-9](https://doi.org/10.1016/0091-3057(89)90204-9)
- Turner, J. R., & Carroll, D. (1985). Heart Rate and Oxygen Consumption during Mental Arithmetic, a Video Game, and Graded Exercise : Further Evidence of Metabolically-Exaggerated Cardiac Adjustments? *Psychophysiology*, 22(3), 261-267. <https://doi.org/10.1111/j.1469-8986.1985.tb01597.x>

- Ulrich-Lai, Y. M., & Herman, J. P. (2009). Neural regulation of endocrine and autonomic stress responses. *Nature Reviews. Neuroscience*, *10*(6), 397-409. <https://doi.org/10.1038/nrn2647>
- van Dooren, M., de Vries, J. J. G. (Gert-J., & Janssen, J. H. (2012). Emotional sweating across the body : Comparing 16 different skin conductance measurement locations. *Physiology & Behavior*, *106*(2), 298-304. <https://doi.org/10.1016/j.physbeh.2012.01.020>
- van Lien, R., Neijts, M., Willemsen, G., & de Geus, E. J. C. (2015). Ambulatory measurement of the ECG T-wave amplitude. *Psychophysiology*, *52*(2), 225-237. <https://doi.org/10.1111/psyp.12300>
- van Stegeren, A. H., Wolf, O. T., & Kindt, M. (2008). Salivary alpha amylase and cortisol responses to different stress tasks : Impact of sex. *International Journal of Psychophysiology*, *69*(1), 33-40. <https://doi.org/10.1016/j.ijpsycho.2008.02.008>
- Veltman, J. A., & Gaillard, A. W. K. (1998). Physiological workload reactions to increasing levels of task difficulty. *Ergonomics*, *41*(5), 656-669. <https://doi.org/10.1080/001401398186829>
- Verwey, W. B., & Veltman, H. A. (1996). Detecting short periods of elevated workload : A comparison of nine workload assessment techniques. *Journal of Experimental Psychology: Applied*, *2*(3), 270-285. <https://doi.org/10.1037/1076-898X.2.3.270>
- Vila, G., Godin, C., Charbonnier, S., Labyt, E., Sakri, O., & Campagne, A. (2018). Pressure-Specific Feature Selection for Acute Stress Detection From Physiological Recordings. *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2341-2346. <https://doi.org/10.1109/SMC.2018.00402>
- Vila, G., Godin, C., Charbonnier, S., Sakri, O., Labyt, E., & Campagne, A. (2019). Physiological response to acute stress against confounding factors : A white-box research method. *Biomedical Physics & Engineering Express*, *5*(5), 055007. <https://doi.org/10.1088/2057-1976/ab360e>
- Villarejo, M. V., Zapirain, B. G., & Zorrilla, A. M. (2012). A Stress Sensor Based on Galvanic Skin Response (GSR) Controlled by ZigBee. *Sensors*, *12*(5), 6075-6101. <https://doi.org/10.3390/s120506075>
- Vrijkotte, T. G., Van Doornen, L. J., & De Geus, E. J. (2000). Effects of work stress on ambulatory blood pressure, heart rate, and heart rate variability. *Hypertension*, *35*(4), 880-886.
- 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), 1063-1070. <https://doi.org/10.1037/0022-3514.54.6.1063>
- Wellek, S. (1996). A New Approach to Equivalence Assessment in Standard Comparative Bioavailability Trials by Means of the Mann-Whitney Statistic. *Biometrical Journal*, *38*(6), 695-710. <https://doi.org/10.1002/bimj.4710380608>
- Wetherell, M. A., Crown, A. L., Lightman, S. L., Miles, J. N. V., Kaye, J., & Vedhara, K. (2006). The four-dimensional stress test : Psychological, sympathetic–adrenal–medullary, parasympathetic and hypothalamic–pituitary–adrenal responses following inhalation of 35% CO₂. *Psychoneuroendocrinology*, *31*(6), 736-747. <https://doi.org/10.1016/j.psyneuen.2006.02.005>
- Wijsman, J. (2013). Trapezius Muscle EMG as Predictor of Mental Stress. *ACM Transactions on Computational Logic*, 21.
- Wijsman, J., Grundlehner, B., Liu, H., Penders, J., & Hermens, H. (2013). *Wearable Physiological Sensors Reflect Mental Stress State in Office-Like Situations*. 600-605. <https://doi.org/10.1109/ACII.2013.105>
- Wijsman, J. L. P. (2014). *Sensing stress : Stress detection from physiological variables in controlled and uncontrolled conditions*. University of Twente.
- World Health Organization. (2003). *The solid facts : Social determinants of health* (2nd ed). WHO Regional Office for Europe.
- Yazdani, H., Mahnam, A., Edrisi, M., & Esfahani, M. A. (2016). Design and Implementation of a Portable Impedance Cardiography System for Noninvasive Stroke Volume Monitoring. *Journal of Medical Signals and Sensors*, *6*(1), 47-56.

- Yin, B., & Chen, F. (2007). Towards Automatic Cognitive Load Measurement from Speech Analysis. In J. A. Jacko (Éd.), *Human-Computer Interaction. Interaction Design and Usability* (p. 1011-1020). Springer Berlin Heidelberg.
- Zadeh, L. A. (1988). Fuzzy logic. *Computer*, 21(4), 83-93. <https://doi.org/10.1109/2.53>
- Zephyr™. (2012). *Zephyr™*. <https://www.zephyranywhere.com/>
- Zhai, J., & Barreto, A. (2006a). Stress detection in computer users based on digital signal processing of noninvasive physiological variables. *Engineering in Medicine and Biology Society, 2006. EMBS'06. 28th Annual International Conference of the IEEE*, 1355-1358. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4462012
- Zhai, J., & Barreto, A. (2006b). Stress Recognition Using Non-invasive Technology. *FLAIRS Conference*, 395-401. <http://www.aaai.org/Papers/FLAIRS/2006/Flairs06-077.pdf>

Résumé en Français

Le stress aigu est une réponse adaptative du corps aux épreuves de la vie quotidienne. Les corrélats physiologiques du stress sont connus et mesurables à travers le rythme cardiaque, la conductance de la peau ou l'activité respiratoire. Récemment, l'essor des capteurs physiologiques portables a donné une perspective concrète à la détection d'épisodes de stress dans la vie quotidienne. Mais les corrélats physiologiques du stress ne lui sont pas spécifiques : ils sont également sensibles à de nombreux facteurs confondants. En outre, la mesure physiologique en vie réelle est sujette à de nombreux artefacts et les modèles actuellement proposés manquent d'une stratégie de gestion efficace de la qualité des mesures. Enfin, la littérature a tendance à promouvoir des modèles de type boîte-noire, dont la décision ne peut être interprétée dans un but thérapeutique.

Pour répondre à ces problématiques, nous proposons des méthodes de type boîte-blanche qui aboutissent à un modèle de détection du stress susceptible de fonctionner dans la vie quotidienne. Nous proposons d'abord une sélection des caractéristiques sensibles au stress et peu sensibles aux facteurs confondants. Sur la base des caractéristiques sélectionnées, nous proposons un détecteur de stress robuste aux artefacts et qui adopte une structure modulaire, afin de s'adapter en temps réel à la qualité des mesures physiologiques. L'idée derrière ce modèle est de pouvoir se séparer sans nouvel apprentissage des caractéristiques momentanément indisponibles ou trop bruitées. Pour valoriser cette capacité adaptative de notre détecteur de stress, nous proposons finalement un algorithme permettant d'estimer en temps réel la qualité des données de rythme cardiaque. Couplé à cet indice de qualité, notre modèle de détection du stress constitue un pas de plus vers une mesure objective de ce phénomène et de sa prévalence dans notre quotidien.

Mots-Clés : Stress, Mesures physiologiques, Fusion de Données, Classification, Méthodes Expérimentales, Qualité des Signaux

English Abstract

Acute stress is our body's adaptive response to the challenges of daily life. Stress has well-known physiological correlates, which can be monitored through heart rate, skin conductance or respiratory activity. Over the past decade, the rise of wearable physiological sensors has provided tangible insights into the detection of stressful episodes in daily life. Yet, physiological stress correlates are also sensitive to many confounding factors. Moreover, real-life physiological measurements involve frequent artifacts ; and current literature models lack an effective strategy for managing sensor data quality. Finally, current accuracy-based approaches tend to promote black-box models whose decision cannot be interpreted for therapeutic purposes.

To tackle these issues, we provide white-box methods to design a stress detection model able to operate in real-life settings. We first propose a feature selection method that identifies features that are sensitive to stress and relatively insensitive to confounding factors. We then develop a stress detection model designed to withstand artefactual data ; and which incorporates a modular structure to timely adapt to the quality of physiological measurements. The idea behind this model is to isolate, without need for training, features that are temporarily unavailable or too noisy for stress detection. To capitalize on such an adaptive ability, we finally provide an algorithm to timely estimate the quality of heart rate data. Combined with this quality index, our stress detection model is one more step towards an objective measurement of this phenomenon and its occurrence in our daily lives.

KeyWords : Stress, Physiological signals, Data Fusion, Classification, Experimentation, Data Quality