

Table des matières

Liste des Tableaux.....	9
Liste des Figures	12
Glossaire.....	16
Introduction Générale	17
Section I État de l'Art	22
Chapitre 1 Le Stress.....	23
1.1 Qu'est ce que le stress ?.....	23
1.1.1 « Fight or Flight » - Walter Cannon	23
1.1.2 Syndrome d'Adaptation Général - Hans Selye	24
1.2 Peur, Anxiété, Phobie	26
1.3 Thérapies Cognitives Comportementales, <i>in vivo</i> et <i>in virtuo</i>	27
1.4 Réponse au stress : coping et appraisal.....	29
1.5 Facteurs d'influence dans l'expérience et l'expression du stress	31
Chapitre 2 L'expression du stress.....	35
2.1 Mesures de facteurs individuels.....	35
2.2 Analyse de signaux pour la caractérisation du stress	38
2.2.1 Dans la physiologie.....	38
2.2.2 Dans la voix	39
2.2.3 Dans les expressions faciales	40
2.2.4 Dans la posture et les gestes	41
2.2.5 Dans les signaux sociaux	42
Chapitre 3 Apprentissage automatique pour la reconnaissance des états affectifs	43
3.1 Affective Computing	43
3.2 Corpus de stress existants et expérimentations stressantes	44
3.3 Algorithmes de classification pour les données affectives	47
3.4 Sélection de paramètres	48
3.5 Fusion d'information pour une classification multimodale.....	49

3.6	Evaluations de performances	52
	Synthèse de l'État de l'Art.....	55
Section II	Collecte, Annotation et Analyse de Signaux de Stress.....	56
Chapitre 4	Constitution de Corpus de Stress.....	57
4.1	Capture de données inédites	57
4.2	E-thérapies : public sociophobe en situation anxiogène simulée	58
4.2.1	Collecte ETCC-HM : interactions homme-machine	58
4.2.2	Collecte ETCC-HH : interaction entres humains.....	63
4.3	Comparses : public non-pathologique en situation stressante ponctuelle.....	65
4.4	Nao Fait de l'Humour: public non-pathologique en interaction sociale humoristique.....	68
4.5	Récapitulatif des corpus.....	71
Chapitre 5	Annotations	72
5.1	Considérations générales	72
5.2	Auto annotation du stress	76
5.2.1	Sur E-Thérapies.....	76
5.2.2	Sur Comparses.....	78
5.2.3	Sur Nao Fait de l'Humour.....	79
5.3	Annotation du stress par un observateur externe.....	80
5.3.1	Protocole d'annotation du stress	80
5.3.2	Accord inter-annotation	82
5.3.3	Sur E-Thérapies.....	86
5.3.4	Sur Comparses.....	88
5.3.5	Sur Nao Fait de l'Humour.....	90
5.4	Comparaison entre auto-annotation et annotation externe.....	94
5.4.1	Sur E-Thérapies.....	94
5.4.2	Sur Comparses.....	97
5.4.3	Sur Nao Fait de l'Humour.....	98
Chapitre 6	Analyse du signal pour l'extraction d'indices du stress.....	99
6.1	Extraction d'indices	99

6.1.1	Observation du signal.....	99
6.1.2	Audio	100
6.1.3	Visage	101
6.1.4	Posture.....	104
6.1.5	Récapitulatif des indices	107
6.2	Etude de cas.....	108
6.2.1	Corrélations entre mesures audio, posturales, et évaluation de performance	108
6.2.2	Corrélations entre affect bursts et facteurs individuels	109
6.2.3	Variations des paramètres considérés pour des cas particulier de stress ..	110
	Synthèse de la Collecte, Annotation et Analyse de Signaux de Stress.....	113
	Section III Détection Automatique du Stress.....	114
	Chapitre 7 Modèles monomodaux de reconnaissance du stress.....	115
7.1	Pré-traitements.....	115
7.2	Audio	120
7.2.1	Voix de sujets phobiques et non-phobiques en interaction avec des humains	120
7.2.2	Voix de sujets phobiques et non-phobiques en interaction avec une machine	123
7.2.3	Voix de sujets non pathologiques en prise de parole en public	127
7.3	Vidéo	135
7.3.1	Expressions faciales de sujets phobiques et non-phobiques en interaction avec une machine	135
7.3.2	Posture de sujets non pathologiques en prise de parole en public	137
7.4	Performances des modèles monomodaux	140
	Chapitre 8 Fusion d'informations multimodales pour la reconnaissance du stress	143
8.1	Fusion de la voix et des expressions faciales de sujets phobiques et non-phobiques en interaction avec une machine	145
8.2	Fusion de la voix et de la posture de sujets non pathologiques en prise de parole en public	149
8.3	Apports de la fusion.....	153
	Chapitre 9 Facteurs individuels d'influence sur la reconnaissance du stress.....	158
9.1	Phobie Sociale	158
9.2	Traits de personnalité.....	163

9.3	Modèles de reconnaissance du stress tenant compte des informations de profil	168
9.3.1	Sujets phobiques et non-phobiques	168
9.3.2	Sujets non-pathologiques	170
9.4	Implications pour la conception de logiciels de remédiation	172
	Synthèse de la Détection Automatique du Stress.....	173
	Conclusion.....	174
	Bibliographie	179
	Publications.....	199
	Annexes	200

Liste des Tableaux

Tableau 1 : Matrice de Confusion pour l'évaluation des modèles de reconnaissance automatique, distinguant les types d'erreurs (VP = vrais positifs, VN = vrais négatifs, FP = faux positifs, FN = faux négatifs).....	52
Tableau 2 : description des sujets de la collecte ETCC-HM	60
Tableau 3 : Illustration et présentation des 3 scénarios du protocole expérimental ETCC-HM pour la collecte de données en interaction homme-machine	61
Tableau 4 : Récapitulatif des données du corpus ETCC-HM.....	62
Tableau 5 : Description des sujets de la collecte ETCC-HH.....	63
Tableau 6 : Récapitulatif des données du corpus ETCC-HH.....	64
Tableau 7 : Description des sujets des collecte Compare.....	66
Tableau 8 : Récapitulatif des données des corpus CP1-CP2-CP3-CP4.....	67
Tableau 9 : Description des sujets de la collecte IS	69
Tableau 10 : Récapitulatif des données des corpus IS-Comp et IS-Nao.....	70
Tableau 11 : Récapitulatif des corpus collectés au cours de l'étude. Les corpus encadrés en gras sont utilisés dans la suite de l'étude.	71
Tableau 12 : Meta informations disponibles par corpus (✓ disponible ; × non renseigné ; - non applicable)	73
Tableau 13 : Stress/Difficulté auto-annoté par tâche par population sur ETCC-HM (stress moyen auto-annoté entres parenthèses).....	77
Tableau 14 : Stress/Difficulté auto-annoté par tâche par population sur ETCC-HH (stress moyen auto-annoté entres parenthèses).....	78
Tableau 15 : Signes indicatifs du stress, traduits d'après [65]	80
Tableau 16 : Critères d'annotation du stress	81
Tableau 17 : Régression d'échelle.....	83
Tableau 18 : Kappa de Cohen et taux d'accord brut pour les corpus Compare	84
Tableau 19 : Taux d'accord entre auto-annotations et annotations externes sur ETCC-HM .	97
Tableau 20 : Taux d'accord entre auto-annotations et annotations externes sur ETCC-HH..	97
Tableau 21 : Taux d'accord entre auto-annotations et annotations externes sur CP2.....	98
Tableau 22 : Taux d'accord entre auto-annotations et annotations externes sur IS-Comp...	98

Tableau 23 : Les 384 descripteurs audio du set emo_IS09 d'OpenSMILE.....	100
Tableau 24 : Les 98 paramètres décrivant les expressions faciales ; un 99 ^{ème} paramètre indique la probabilité de la présence d'un visage dans la séquence d'images observée.....	103
Tableau 25 : Les 15 descripteurs posturaux calculés à partir de la silhouette des sujets	106
Tableau 26 : Récapitulatif des indices extraits des enregistrements de cette étude.	107
Tableau 27 : Indices audio et visuels de l'étude [II]	108
Tableau 28 : Nombre de segments de 1s,2s et 5s utilisés pour construire les modèles, par corpus.....	117
Tableau 29 : Sélection des 25 meilleurs, 50 meilleurs et 100 meilleurs paramètres robustes aux variations interindividuelles et inter-contextuelles pour la reconnaissance du stress....	119
Tableau 30 : Comparaison des scores de f-mesure pour les modèles entraînés à partir des données audio du corpus ETCC-HH (22 phobiques / 18 témoins en interaction avec une machine). DL = modèle dépendant du locuteur ; IL = modèle indépendant du locuteur ; Auto = stress auto-annoté par les sujets ; Externe = stress annoté par des observateurs. Scores moyens sur 3 sous-ensembles de données (déviations standard en italique pour les modèles IL). Les meilleurs scores sont en gras	120
Tableau 31 : Comparaison des scores de f-mesure pour les modèles entraînés à partir des données audio du corpus ETCC-HM (15 phobiques / 9 témoins en interaction avec une machine). DL = modèle dépendant du locuteur ; IL = modèle indépendant du locuteur ; Auto = stress auto-annoté par les sujets ; Externe = stress annoté par des observateurs. Scores moyens sur 3 sous-ensembles de données (déviations standard en italique pour les modèles IL). Les meilleurs scores sont en gras.	124
Tableau 32 : Paramètres audio utilisés pour la reconnaissance du stress sur le corpus de sujets en prise de parole en public dans l'étude [III]	128
Tableau 33 : Comparaison des scores de f-mesure pour les modèles entraînés à partir des données audio du corpus CP2 (19 personnes en prise de parole en public). DL = modèle dépendant du locuteur ; IL = modèle indépendant du locuteur ; Auto = stress auto-annoté par les sujets ; Externe = stress annoté par des observateurs. Scores moyens sur 3 sous-ensembles de données (déviations standard en italique pour les modèles IL). Les meilleurs scores sont en gras.....	130
Tableau 34 : Comparaison des scores de f-mesure pour les modèles entraînés à partir des données faciales du corpus ETCC-HM (15 phobiques / 9 témoins en interaction avec une machine). DL = modèle dépendant du locuteur ; IL = modèle indépendant du locuteur ; Auto = stress auto-annoté par les sujets ; Externe = stress annoté par des observateurs. Scores moyens sur 3 sous-ensembles de données (déviations standard en italique pour les modèles IL). Les meilleurs scores sont en gras.	135

Tableau 35 : Comparaison des scores de f-mesure pour les modèles entraînés à partir des données posturales du corpus CP2 (19 personnes en prise de parole en public). DL = modèle dépendant du locuteur ; IL = modèle indépendant du locuteur ; Auto = stress auto-annoté par les sujets ; Externe = stress annoté par des observateurs. Scores moyens sur 3 sous-ensembles de données (déviati <u>o</u> n standard en italique pour les modèles IL). Les meilleurs scores sont en gras.....	138
Tableau 36 : Paramètres audio moins robustes aux variations interindividuelles et inter-contextuelles pour la reconnaissance du stress.....	142
Tableau 37 : Nombre de segments de 1s, 2s et 5s utilisés pour la fusion, par corpus.....	143
Tableau 38 : Paramètres audio et faciaux utilisés pour la reconnaissance bimodale du stress sur le corpus ETCC-HM (13 sujets phobiques et 8 sujets témoins en interaction avec une machine).....	145
Tableau 39 : Comparaison des scores de f-mesures des modèles entraînés à partir des données fusionnées audio et faciales du corpus ETCC-HM (13 phobiques / 8 témoins en interaction avec une machine). P = fusion précoce ; V = fusion par votes simples ; W = fusion par votes pondérés ; C = fusion en cascade; Auto = stress auto-annoté par les sujets ; Externe = stress annoté par des observateurs. Scores moyens sur 3 sous-ensembles de données (déviati <u>o</u> n standard en italique). Les meilleurs scores sont en gras.	146
Tableau 40 : Paramètres audio et posturaux utilisés pour la reconnaissance bimodale du stress sur le corpus CP2 (18 sujets en prise de parole en public).	149
Tableau 41 : Comparaison des scores de f-mesures des modèles entraînés à partir des données fusionnées audio et posturales du corpus CP2 (18 personnes en prise de parole en public). P = fusion précoce ; V = fusion par votes simples ; W = fusion par votes pondérés ; C = fusion en cascade; Auto = stress auto-annoté par les sujets ; Externe = stress annoté par des observateurs. Scores moyens sur 3 sous-ensembles de données (déviati <u>o</u> n standard en italique). Les meilleurs scores sont en gras.....	150
Tableau 42 : Paramètres robustes aux variations de configurations pour la reconnaissance du stress par les modèles conçus en fusion pondérée sur le corpus ETCC-HM (13 phobiques / 8 témoins en interaction avec une machine). Les paramètres en gras sont communs aux deux annotations.	155
Tableau 43 : Paramètres robustes aux variations de configurations pour la reconnaissance du stress par les modèles conçus en fusion pondérée sur le corpus CP2 (18 sujets en prise de parole en public). Les paramètres en gras sont communs aux deux annotations.....	156
Tableau 44 : Paramètres audio extraits des données du corpus ETCC-HH (22 phobiques / 18 témoins en interaction avec des humains) pour l'étude des comportements vocaux des populations.....	159

Liste des Figures

Figure 1: Le stress, réaction physiologique du corps face à une menace	25
Figure 2 : Modèle de communication « en lentille », adapté de Scherer	33
Figure 3 : Processus d'apprentissage et de classification de données.....	44
Figure 4 : Entraînement et utilisation d'un système de reconnaissance automatique avec fusion tardive	51
Figure 5 : Courbe sensibilité/spécificité, ou courbe ROC (Receiver Operating Characteristic)	53
Figure 6 : Interface de contrôle MOz des animations ETCC-HM.....	59
Figure 7 : Disposition de l'expérience de collecte ETCC-HM.....	62
Figure 8 : Disposition de l'expérience de collecte ETCC-HH.....	64
Figure 9 : Collecte de données Comparse, deuxième campagne (2012)	66
Figure 10 : Interface de contrôle MOz de l'application IS	69
Figure 11 : Collecte de données lors d'Interspeech 2013, à Lyon (photo A. Crenn)	70
Figure 12 : Granularité de l'annotation.....	74
Figure 13 : Stress auto-annoté par tâche par population sur ETCC-HM	76
Figure 14 : Stress auto-annoté par tâche par population sur ETCC-HH.....	78
Figure 15 : Répartition des auto-annotations du stress (de 1 à 5) sur CP2.....	79
Figure 16 : Répartition des auto-annotations de gêne (de 1 à 5) sur IS	79
Figure 17 : Accord inter-annotateurs suivant le Kappa observé proposé par Landis et Koch [168].....	82
Figure 18 : Stress annoté par 2 experts par tâche par population sur ETCC-HM	86
Figure 19 : Stress annoté par 2 experts par tâche par population sur ETCC-HH.....	87
Figure 20 : Stress annoté par deux observateurs par tâche sur CP2	88
Figure 21 : Stress annoté par deux observateurs par tâche sur CP3	88
Figure 22 : Stress annoté par deux observateurs par tâche sur CP4	89
Figure 23 : Proportion de segments stress/non-stress pour chaque tâche sur CP2	90
Figure 24 : Durée relative des segments stress/non-stress pour chaque tâche sur CP2.....	90
Figure 25 : Répartition des annotations de gêne par tâche sur IS-Comp.....	91

Figure 26 : Gêne moyenne par population par tâche sur IS-Comp.....	92
Figure 27 : Nombre de rires par tâche sur IS-Comp et IS-Nao	92
Figure 28 : Nombre de sourires par tâche sur IS-Comp.....	93
Figure 29 : Fréquence des annotations de stress par type d'annotation sur ETCC-HM	94
Figure 30 : Fréquence des annotations de stress par type d'annotation sur ETCC-HH.....	95
Figure 31 : Fréquence des annotations de stress de la population phobique sur ETCC-HM	95
Figure 32 : Fréquence des annotations de stress de la population phobique sur ETCC-HH	96
Figure 33 : Processus d'analyse par fenêtres glissantes avec 50% de recouvrement.....	100
Figure 34 : Le modèle à 66 points de FaceTracker	102
Figure 35 : Silhouette capturée par la Kinect, et boîte englobante.....	105
Figure 36 : Quantité de Mouvement à partir de la silhouette capturée par la Kinect	106
Figure 37 : Densité de probabilité du premier composant de l'Analyse en Composantes Principales (PCA) des paramètres audio sur le corpus ETCC-HH	110
Figure 38 : Densité de probabilité du premier composant de l'Analyse en Composantes Principales (PCA) des paramètres faciaux sur le corpus ETCC-HM	111
Figure 39 : Densité de probabilité du premier composant de l'Analyse en Composantes Principales (PCA) des paramètres posturaux sur le corpus CP2.....	112
Figure 40 : Découpage des corpus en 3 sous-ensembles pour des modèles indépendants du locuteur.....	116
Figure 41 : Découpage des corpus en 3 sous-ensembles de sujets assurant une répartition équilibré des annotations suivant le type de données produit par sujet.....	116
Figure 42 : Meilleures configurations pour les modèles entraînés sur les données audio du corpus ETCC-HH (22 phobiques / 18 témoins en interaction avec des humains). DL = modèle dépendant du locuteur ; IL = modèle indépendant du locuteur ; Auto = stress auto-annoté par les sujets ; Externe = stress annoté par des observateurs.....	122
Figure 43 : Meilleures configurations pour les modèles entraînés sur les données audio du corpus ETCC-HM (15 phobiques / 9 témoins en interaction avec une machine) . DL = modèle dépendant du locuteur ; IL = modèle indépendant du locuteur ; Auto = stress auto-annoté par les sujets ; Externe = stress annoté par des observateurs.....	126
Figure 44 : Scores de f-mesure pour les modèles SVM audio dépendants du locuteur sur les annotations par tour de parole du stress chez les sujets en prise de parole en public, Entraînement en cross-validation 10-folds.....	129

Figure 45 : Meilleures configurations pour les modèles entraînés sur les données audio du corpus CP2 (19 personnes en prise de parole en public) . DL = modèle dépendant du locuteur ; IL = modèle indépendant du locuteur ; Auto = stress auto-annoté par les sujets ; Externe = stress annoté par des observateurs.....	133
Figure 46 : Scores de f-mesure par type de modèle et par population à partir des auto-annotations du stress sur ETCC-HM (13 sujets phobiques et 8 sujets témoins en interaction avec une machine).....	160
Figure 47 : Scores de f-mesure par type de modèle et par population à partir des annotations externes du stress sur ETCC-HM (13 sujets phobiques et 8 sujets témoins en interaction avec une machine).....	161
Figure 48 : Influence de l'extraversion sur les performances des modèles auto-annotés sur les données du corpus CP2 (18 sujets en prise de parole en public).	164
Figure 49 : Influence du névrotisme sur les performances des modèles auto-annotés sur les données du corpus CP2 (18 sujets en prise de parole en public).....	164
Figure 50 : Influence de l'alexithymie sur les performances des modèles auto-annotés sur les données du corpus CP2 (18 sujets en prise de parole en public).....	165
Figure 51 : Influence du trait anxieux sur les performances des modèles auto-annotés sur les données du corpus CP2 (18 sujets en prise de parole en public).....	165
Figure 52 : Reconnaissance du névrotisme par des modèles audio, posturaux, et en fusion pondérée sur le corpus CP2 (18 sujets en prise de parole en public). Score moyen et déviation standard sur 3 sous-ensembles de sujets.....	166
Figure 53 : Reconnaissance de l'anxiété par des modèles audio, posturaux, et en fusion pondérée sur le corpus CP2 (18 sujets en prise de parole en public). Score moyen et déviation standard sur 3 sous-ensembles de sujets.....	167
Figure 54 : Reconnaissance du stress en fusion pondéré d'informations audio et faciales avec et sans informations de profil à partir des auto-annotations sur le corpus ETCC-HM (13 phobiques et 8 témoins en interaction avec une machine). Score de f-mesure moyen et déviation standard sur 3 sous-ensembles de sujets.....	169
Figure 55 : Reconnaissance du stress en fusion pondéré d'informations audio et faciales avec et sans informations de profil à partir des annotations externes sur le corpus ETCC-HM (13 phobiques et 8 témoins en interaction avec une machine). Score de f-mesure moyen et déviation standard sur 3 sous-ensembles de sujets.....	169
Figure 56 : Reconnaissance du stress en fusion pondéré d'informations audio et posturales avec et sans informations de profil à partir des auto-annotations sur le corpus CP2 (18 sujets en prise de parole en public). Score de f-mesure moyen et déviation standard sur 3 sous-ensembles de sujets.....	170

Figure 57 : Reconnaissance du stress en fusion pondéré d'informations audio et posturales avec et sans informations de profil à partir des auto-annotations sur le corpus CP2 (18 sujets en prise de parole en public). Score de f-mesure moyen et déviation standard sur 3 sous-ensembles de sujets..... 171

Glossaire

- AAM Active Appearance Model
- Affect Bursts courtes expressions affectives non verbales, e.g les rires
- Affective Computing l'utilisation des états affectifs dans les interactions Homme-Machine
- Appraisal mécanisme d'appréhension des évènements et de leur impact sur les buts, valeurs et désirs d'un individu
- BFI Big Five Inventory
- CI Contraction Index
- COPE Cope Inventory
- Coping mécanisme d'adaptation pour gérer le stress
- DL Dépendant du Locuteur
- FACS Facial Action Coding System
- FPS Frames Per Seconds
- HAD Hospital Anxiety and Depression scale
- Kappa mesure d'accord
- IL Indépendant du locuteur
- LSAS Liebowitz Social Anxiety Scale
- MFCC Mel-Frequency Cepstral Coefficients
- Modèle Dans cette étude, modèle SVM conçu pour la distinction automatique du stress et du non-stress
- MOz Magicien d'Oz
- OCEAN Openness, Conscientiousness, Extraversion, Agreeableness, Neuroticism
- PCA Principal Component Analysis
- PST Public Speaking Task
- Remédiation rééducation de fonctions cognitives altérées
- RV Réalité Virtuelle
- SHS Sens of Humour Scale
- Sociophobe Personne souffrant de Phobie Sociale
- STAI State Trait Anxiety Index
- Stress réaction physiologique et cognitive à une menace réelle ou imaginée
- SVM Support Vector Machines
- TAS Toronto Alexithymia Scale
- TCC Thérapie Cognitive Comportementale
- TSST Trier Social Stress Test
- QES Questionnaire d'Éreutophobie de la Salpêtrière
- QoM Quantity of Motion

Introduction Générale

“Everybody is always under some degree of stress. Even while quietly asleep our heart must continue to beat, our lungs to breathe, and even our brain works in the form of dreams. Stress can be avoided only by dying.”

– Hans Selye *'The Nature of Stress'*, 1985 [283]

Motivation et Contexte

Réaction à toute forme de pression physique ou psychologique sur le corps et l'esprit, le stress est une constante dans nos vies. Si cette réponse permet d'échapper à une menace ponctuelle, elle devient un handicap lorsque les pressions qui la déclenchent sont constantes. L'influence négative du stress sur le comportement des individus limite leur capacité à réaliser leurs aspirations personnelles et professionnelles.

L'anxiété ressentie de façon ponctuelle, par exemple dans une situation où notre performance est évaluée, est naturelle ; mais elle peut devenir un handicap et une souffrance dans le cas de troubles anxieux affectant le fonctionnement des individus au point de les empêcher de travailler ou d'établir et maintenir des relations sociales.

Si le stress est inévitable comme le suggère Hans Selye dans la citation introduisant ce document, il est possible d'apprendre à contrôler nos réactions dans une situation stressante, de façon à ne pas laisser notre anxiété nous dominer.

Les travaux présentés dans cette étude ont été menés dans le cadre du projet FEDER E-Thérapies. Ce projet est une première étude visant à valider l'utilisation de logiciels comme outils de remédiation pour le dépistage et le traitement de troubles anxieux et de troubles du langage, but ambitieux. Le développement de tels systèmes peut permettre de créer des solutions de remédiation pour des troubles largement prévalent dans la société. Notre étude porte sur l'hypothèse que la détection des troubles anxieux des sujets par une machine est possible.

L'objectif du projet E-Thérapie est l'élaboration d'un système plaçant les sujets en interaction avec un Agent Conversationnel Animé (ACA) virtuel, dans des scénarios anxiogènes élaborés par des psychiatres. La détection du stress des patients doit servir à orienter le dialogue mené par l'ACA, mais également être utilisée pour fournir aux utilisateurs un *biofeedback*, un retour en direct de leur niveau de stress.

Ce logiciel est conçu comme un système paramétrable à destination des psychiatres traitants les troubles phobiques, comme un outil de Thérapie Cognitive Comportementale (TCC) complémentaire des exercices de mise en situations simulées et *in vivo* existants. L'utilisation d'un outil informatique permet la répétition contrôlée d'exercices adaptés par le psychiatre aux besoins des patients et augmentation progressive de leur difficulté, tout en mesurant l'évolution du stress au cours de l'exercice et durant la thérapie.

Dans cette étude, nous abordons l'utilisation d'animations interactives simplifiées sans agents conversationnels pour la mise en situation anxiogène, afin de collecter des données auprès de sujets phobiques et de sujets témoins pour l'étude de la perception humaine et de la détection automatique du stress.

Nos travaux portent également sur des données collectées dans le cadre du projet ANR Comparse dont l'objectif est l'étude de l'impact des états affectifs sur le comportement, avec et sans ACA. Dans ce contexte également, nous n'avons travaillé que sur les interactions n'utilisant pas d'agent. Enfin l'Action Incitative « Nao Fait de l'Humour » nous a permis de mener une étude sur les rires dans les interactions homme-machine, dans des situations plus ou moins stressantes. Ces deux projets nous ont également donné l'opportunité d'étudier l'influence de différentes stratégies de coping*, mécanisme d'adaptation pour répondre à un évènement stressant.

Les travaux présentés dans cette thèse mettent en avant plusieurs axes de recherche pour la conception d'outils de détection du stress, dans des applications médicales ou d'entraînement à des situations stressantes sous la forme de *serious games*. Nous abordons l'évaluation de nouvelles solutions – particulièrement l'utilisation d'outils informatiques de détection des états affectifs – pour l'implémentation d'exercices préexistants de TCC*, à destination d'un public phobique. Nous étudions également l'analyse automatique des comportements affectifs des humains par des machines pour la reconnaissance du stress, afin d'aider des individus stressés à développer des comportements positifs pour réguler leurs émotions, par exemple dans le contexte d'une prise de parole en public ou d'un entretien d'embauche.

Challenges

Cette étude fait essentiellement face à deux défis :

Parvenir à identifier le stress chez des populations variées dans différents contextes : les projets dans lesquels se place cette étude impliquent des sujets d'études différents (des phobiques avec une propension exagérée au stress, des individus non-pathologiques ayant tous différentes stratégies de coping) et dans des contextes anxiogènes variés. Dans cette étude nous cherchons à déterminer s'il existe des indices « universels » du stress, ou si au contraire les caractéristiques individuelles de chacun favorisent la recherche d'indices spécifiques suivant les populations considérées.

Obtenir des indices acoustiques et visuels permettant d'identifier le stress à l'aide de capteurs peu invasifs : le fait de se savoir observé est un facteur déclencheur ou aggravant du stress, en particulier chez des individus anxieux en situation de performance. L'utilisation de mesures invasives comme des capteurs placés sur le corps risquant de provoquer un stress supplémentaire, nous avons opté pour l'utilisation d'indices mesurables à l'aide d'un micro et d'une caméra. Par ailleurs, dans l'objectif d'une diffusion auprès d'un large public de logiciels de remédiation, il est plus pertinent d'utiliser des capteurs facilement accessibles en dehors d'un laboratoire de recherche.

Considérations éthiques

L'étude des comportements des individus (stress, états affectifs) est un sujet qui requiert de tenir compte de considérations éthiques relatives aux travaux impliquant des sujets humains et des informations à caractère hautement personnel.

La provocation d'un état anxieux, en particulier chez des individus fragiles ayant une tendance pathologique au stress, nécessite de prendre des précautions dans la création des protocoles expérimentaux. Afin de limiter l'impact néfaste de nos expérimentations sur les participants, nous avons porté une attention particulière à deux points :

- le consentement explicite des participants, par la signature d'un accord décrivant en détail l'expérience, garantissant l'anonymat des données et leur exploitation à des fins de recherche uniquement, et la possibilité pour le sujet de se retirer de l'expérimentation à tout moment y compris *a posteriori*
- un dialogue ouvert avec le chercheur – soit entre les tâches soit *a posteriori* selon l'expérimentation – sur les objectifs de recherches des différentes tâches

La conception d'outils de reconnaissance automatique des états de stress souligne également la question de la fiabilité du système, et de l'impact des erreurs sur les utilisateurs. Dans l'objectif de concevoir un outil pour aider un individu à maîtriser son stress, une perception erronée de l'état de l'utilisateur pourrait le déstabiliser, voir renforcer son anxiété. Dans ce contexte, l'expertise des thérapeutes est essentielle et les systèmes automatisés doivent être conçus comme des outils pour les assister et non des solutions autonomes.

Apports de l'étude

Cette étude a contribué à plusieurs aspects de la reconnaissance automatique du stress dans le but de constituer des outils de remédiation :

- Afin d'étudier le stress auprès de différentes populations et dans différents contextes nous avons collaboré à la création de protocoles expérimentaux, et à la collecte de données à l'aide de ces protocoles, résultant en des corpus multimodaux originaux de comportements stressés réalistes.
- Pour exploiter ces données, nous avons proposé un protocole d'annotation du stress, à partir des indices utilisés en psychologie. Ce protocole a été appliqué à certains corpus de cette étude et les résultats ont été comparés aux auto-évaluations des participants des différentes expérimentations.
- Nous avons utilisé les données collectés et annotés pour extraire des paramètres, habituellement utilisés dans la reconnaissance des états affectifs, afin d'entraîner et d'évaluer des modèles de reconnaissance automatique du stress.
- Nous avons évalué les performances des modèles constitués suivant différents facteurs impactant l'expérience et l'expression du stress, en particulier auprès de populations souffrant de trouble phobique.

Organisation du Manuscrit

Ce manuscrit s'organise en trois sections :

La première section présente un état de l'art axé autour de la notion de stress et de la reconnaissance automatique du stress. Cette section est divisée en trois chapitres : le premier chapitre s'intéresse à la définition du stress d'un point de vue physiologique, psychologique et sociologique, et aborde également les traitements par Thérapie Cognitive Comportementale *in vivo* et un panorama des approches en Réalité Virtuelle (RV) ; le deuxième chapitre s'intéresse aux signes du stress et à leurs mesures ; le troisième chapitre s'intéresse à la méthodologie de mise en œuvre d'un système de détection automatique du stress, il présente un panorama des corpus de stress existants et les techniques d'apprentissage automatique en usage pour la reconnaissance du stress .

La deuxième section présente nos travaux de collecte et d'annotation de données, et d'analyse de signaux acoustiques et visuels pour l'étude du stress. Cette section se divise en trois chapitres : le premier chapitre décrit la constitution de corpus de données réalistes dans trois contextes susceptibles d'éliciter du stress ; le deuxième chapitre détaille le protocole d'annotation élaboré pour l'évaluation du stress dans nos données, et les résultats de ces annotations ; le troisième chapitre présente les paramètres vocaux, faciaux et posturaux utilisés pour la perception du stress dans nos données.

La troisième et dernière section présente les expériences menées sur les données collectées pour valider la faisabilité d'un système de reconnaissance automatique du stress à partir d'indices vocaux, faciaux et posturaux. Cette section se divise en trois chapitres : le premier chapitre présente une expérience d'apprentissage du stress à partir d'indices extraits d'une seule modalité ; le deuxième chapitre présente une expérience d'apprentissage du stress à partir de la fusion d'indices multimodaux ; le troisième chapitre présente une expérience d'apprentissage du stress sur différentes populations distinguées par leur attitude face à une situation anxiogène.

Section I État de l'Art

Dans cette partie, nous présentons un aperçu des travaux existants sur la compréhension et la reconnaissance du stress, de façon pluridisciplinaire et orientée vers la conception d'outils de remédiation. Cette partie se divise en trois chapitres :

Le Chapitre 1 « Le Stress » explicite les notions de stress, de phobie, et de coping (i.e. la gestion du stress) d'un point de vue physiologique et cognitif, et montre l'aspect fortement individuel du stress suivant des facteurs contextuels, socioculturels et personnels. Ce chapitre aborde également les traitements de la phobie par la Thérapie Cognitive Comportementale, et les approches en Réalité Virtuelle de ce type de traitement.

Le Chapitre 2 « L'expression du stress » présente les indices mesurables du stress, à la fois par des questionnaires d'évaluation psychologiques, et par des mesures de comportements physiologiques, vocaux, faciaux et posturaux.

Le Chapitre 3 « Apprentissage automatique pour la reconnaissance des états affectifs » présente les corpus de stress existants et les protocoles expérimentaux utilisés pour éliciter du stress, ainsi que les méthodologies, corpus et techniques d'apprentissage automatique utilisées dans la reconnaissance du stress.

Chapitre 1 Le Stress

1.1 Qu'est ce que le stress ?

1.1.1 « *Fight or Flight* » - Walter Cannon

Le stress est une réaction physiologique à une situation perçue comme menaçante pour l'individu stressé. C'est la réponse « *Fight or Flight* »¹ originellement décrite par le physiologiste Walter B. Cannon [43], une réaction du système nerveux orthosympathique à une situation compromettant la survie de l'individu afin de préparer le corps pour une réponse offensive ou défensive pour échapper au danger. Les glandes surrénales libèrent dans le sang de l'adrénaline et du cortisol, dont l'action sur le corps favorise une réaction musculaire violente imminente : les rythmes cardiaque et respiratoire s'accélèrent et les réserves métaboliques sont libérées pour fournir au corps l'énergie dont il va avoir besoin, la circulation sanguine est modifiée pour favoriser les muscles au détriment des systèmes digestif et immunitaires, la capacité de coagulation est augmentée en prévision d'hémorragie, les glandes lacrymales et salivaires sont inhibées, et la vision périphérique et l'audition sont altérées pour concentrer l'attention sur la menace.

Cette réaction, mécanisme de survie constaté chez tous les vertébrés et illustré dans la Figure 1, est un atout évolutionnaire : les individus capables de réagir rapidement à un danger et d'y échapper pourront survivre et se reproduire. Mais si une réponse de stress aiguë était un avantage il y a quelques millénaires dans un environnement où les périls physiques étaient réels et ponctuels, elle devient un handicap lorsque les menaces sont figuratives et constantes.

Bien que les menaces physiques restent une réalité dans certaines circonstances (conflits armés, criminalité, catastrophe naturelle, pauvreté...), les dangers auxquels nous faisons face au quotidien sont souvent le fruit de notre interprétation d'une situation : peur de l'échec professionnel lorsque nos contraintes de travail dépassent nos ressources, peur de décevoir ses proches lorsque nos décisions ne s'accordent pas aux leurs, peur du jugement de l'autre lorsque nos choix de vie ne correspondent pas aux valeurs portées par la communauté dans laquelle nous vivons... Ces peurs sont des réactions émotionnelles à des dissonances entre notre perception de la réalité et les attentes imposées par des éléments sur lesquels nous n'avons pas de contrôle : milieu socioculturel, économique ou politique. Cependant les mécanismes évolutionnaires en place y réagissent de la même façon qu'à la peur plus primaire d'un prédateur ou d'une chute mortelle.

¹ Littéralement « Combat ou Fuite »

État de l'Art

1.1.2 *Syndrome d'Adaptation Général – Hans Selye*

La réponse physiologique qu'est le stress prépare le corps à échapper au danger, et retourne vers un état normal lorsque celui-ci est passé. Face à une menace plus diffuse dont le terme n'est jamais atteint, le corps reste en état d'alerte permanent : c'est le *Syndrome d'Adaptation Général* décrit par Hans Selye [282]. Selye décrit le stress comme une réaction du corps à toute demande qui lui est faite ; si cette demande n'excède pas les capacités du corps celui-ci répond à la demande puis récupère, mais si le corps épuise ses ressources les conséquences peuvent être sévères.

Les répercussions physiques du stress prolongé sont la source d'une variété de problèmes de santé [45] : l'élévation du rythme cardiaque et la coagulation augmentée provoque des maladies cardio-vasculaires, la sous-irrigation du système digestif et l'altération des réserves de glucose deviennent un dérèglement du système digestif, l'hyperactivité musculaire devient un dérèglement des cycles de sommeil, l'altération de l'audition et de la vision devient une dégradation des capacités de concentration... Le corps entre dans un cercle vicieux où ces effets néfastes sur son intégrité sont perçus comme des menaces et maintiennent la réponse de stress active alors même qu'elle est la source du problème.

Au delà des conséquences physiques directes, la réponse primaire « Fight or Flight » qui se traduisait originellement par une stratégie offensive (attaquer le prédateur) ou défensive (courir dans la direction opposée) se traduit par des comportements contemporains équivalents : attitude agressive verbale ou physique envers les autres ou envers soi-même, ou au contraire fuite par l'isolation sociale ou la toxicomanie [161].

A plus grande échelle, le stress comme tout facteur dégradant la santé physique et morale des individus a un impact économique. Une enquête de 2005 de la fondation européenne pour l'amélioration des conditions de vie et de travail [93] révèle que le stress au travail touche 20% des salariés européens tout secteur d'activité confondu ; une enquête de 2009 de l'Institut National de Recherche et de Sécurité [138] estime que le stress est responsable de 50% des journées de travail perdues pour un coût de 20 milliards d'euros.

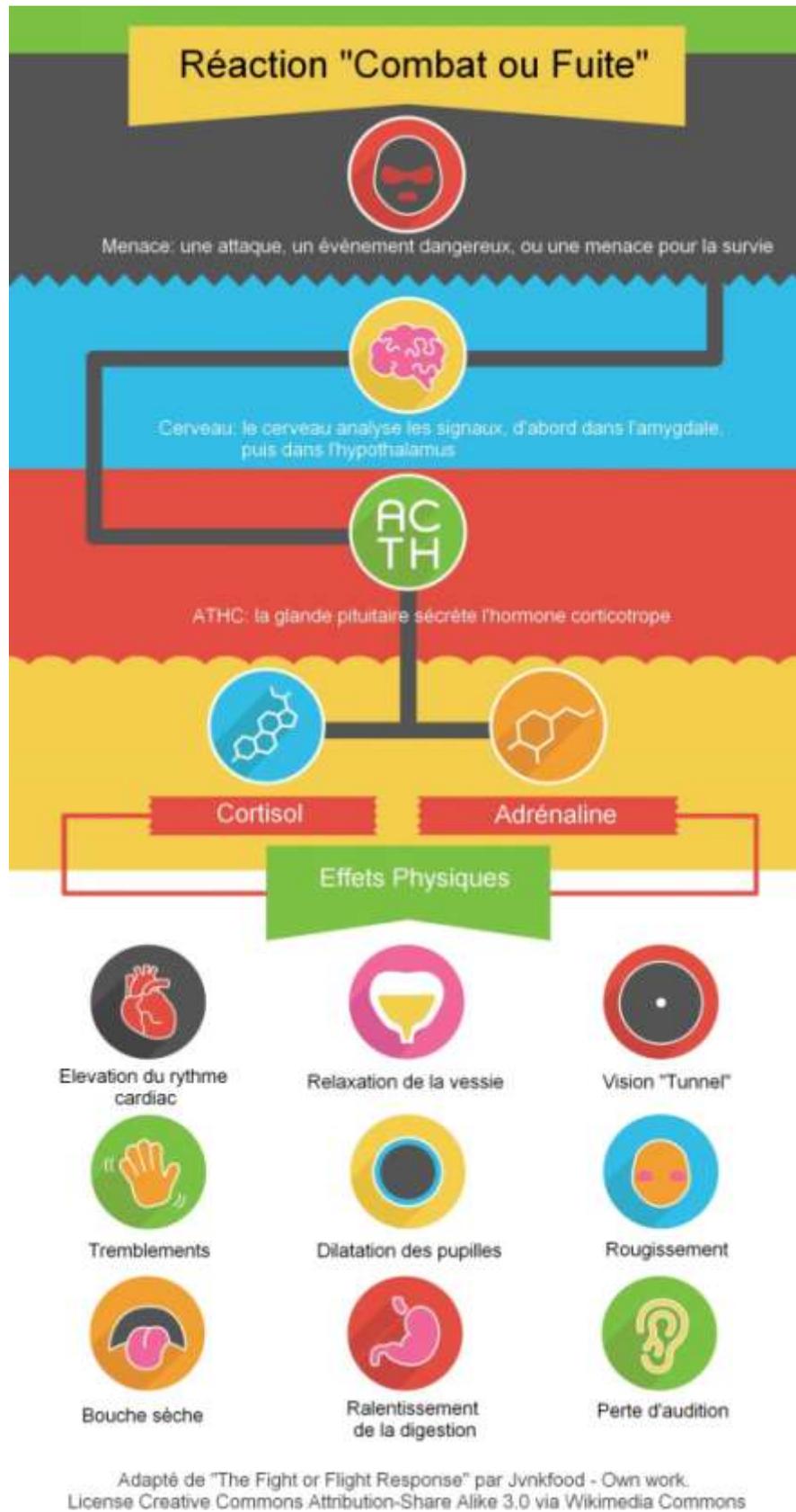


Figure 1: Le stress, réaction physiologique du corps face à une menace

1.2 Peur, Anxiété, Phobie

Nous avons présenté le stress comme une réaction physiologique à une menace objective ou subjective, un mécanisme de survie à court terme mais un cycle destructeur à long terme si la menace est perçue comme inaltérable de façon prolongée. Deux états peuvent déclencher la réaction physiologique de stress : la peur, et l'anxiété. Öhman [219] suggère que la distinction entre ces deux états se fait non pas par les symptômes ressentis (et comparables), mais par des éléments subjectifs comme la certitude de la menace perçue, la forme de réponse à cette menace, et la durée prolongée de l'expérience.

Sylvers et. al. [297] résument la peur comme une émotion de courte durée, vécue dans l'instant présent, déclenchée par une menace spécifique, et qui permet de fuir ou d'affronter la menace. A l'inverse, ils définissent l'anxiété comme un sentiment de longue durée, vécu dans un futur projeté, déclenché par une menace diffuse ou mal identifiée, et qui résulte en un comportement de prudence excessive vis-à-vis d'une menace potentielle, interférant avec les mécanismes de coping. Cette distinction est importante car si la peur et l'anxiété déclenchent une réponse stressée, dans le cas de la peur la menace est identifiée, tandis que dans le cas de l'anxiété les symptômes ressentis ne semblent rattachés à rien, et il est nécessaire de remonter à la source du problème avant de pouvoir envisager d'agir.

Gorman [111] définit l'anxiété comme l'appréhension désagréable d'une situation qui potentiellement n'arrivera jamais. C'est par exemple la peur de ne pas parvenir à prendre la parole devant un public alors même que l'on connaît son discours par cœur et que l'on a déjà effectué l'exercice avec succès des dizaines de fois dans le passé. Généralement l'état anxieux se dissipe lorsqu'il devient apparent que la situation redoutée ne se produira pas, par exemple une fois les premières phrases prononcées sur scène. Cependant il arrive que l'état anxieux ne disparaisse pas et au contraire s'intensifie, jusqu'au point où les individus sont incapables d'affronter la situation anxiogène. On parle alors de troubles anxieux associés à une situation spécifique.

Les troubles anxieux sont reconnus comme trouble mental selon le Diagnostic and Statistical Manual of Mental Disorders (DSM) de l'American Psychiatric Association (APA) [3]. Il existe une grande variété de troubles anxieux, qui peuvent être distingués selon la manifestation de l'anxiété : symptômes épisodiques, ou symptômes continus. Les phobies sont des troubles épisodiques déclenchés par un objet ou une situation particulière, comme la coulrophobie (la peur des clowns), la geliophobie (la peur du rire), la chiraptophobie (la peur du contact physique), l'aviophobie (la peur de prendre l'avion), l'aéroacrophobie (la peur des hauteurs) ou encore l'éreutophobie (la peur de rougir).

État de l'Art

Dans cette étude, nous nous intéressons particulièrement à la phobie sociale, trouble anxieux caractérisé par la peur du regard d'autrui et la crainte d'un jugement négatif dans des situations d'interactions sociales. Elle s'accompagne d'épisodes aigus d'anxiété lors de la confrontation aux situations redoutées, et se traduit généralement par l'adoption de stratégie d'évitement. Heimberg et. al. [125] proposent trois sous-catégories de phobie sociale : phobie généralisée (crainte de toutes les situations sociales), non-généralisées (dans quelques situations sociales, le sujet n'est pas anxieux), et spécifique (crainte d'une ou deux situations sociales spécifiques). Liebowitz [183] évalue d'ailleurs la sévérité de la phobie sociale par un questionnaire évaluant la tendance à l'évitement d'une vingtaine de situations, le Liebowitz Social Anxiety Scale.

La peur du jugement est souvent décorrélée de la performance effective des sujets : Hofmann et Barlow donnent ainsi l'exemple de chanteurs populaires qui révèlent être pétrifiés sur scène, et ce bien après les premières minutes « normales » de trac [131]. Il s'agit d'ailleurs d'un trouble anxieux particulièrement prévalent : Lépine et. al. [181] estiment que 2 à 4% de la population française souffre d'une forme de phobie sociale ; la prévalence au sein de la population américaine est estimée à 6,8% [152]. Il s'agit d'un trouble anxieux particulièrement handicapant pour les malades, affectant aussi bien leur vie professionnelle que personnelle de par leur tendance à éviter les situations d'interaction [264].

1.3 Thérapies Cognitives Comportementales, *in vivo* et *in virtuo*

Un traitement particulièrement efficace en réponse à la phobie sociale en particulier, et aux troubles anxieux en général, est la Thérapie Cognitive Comportementale (TCC*) [130]. La TCC* postule que les processus cognitifs internes influent nos comportements externes [83] ; elle vise à corriger certain comportement, comme l'évitement pathologique dans le cas des phobies. Le thérapeute identifie les processus cognitifs associés à une situation anxiogène et les revisite avec le patient afin de les réaligner avec ses désirs et ses buts [142]. Il s'agit d'une thérapie « par l'action », où le thérapeute fait rejouer au patient des situations anxiogènes qui ont déclenché un comportement non désirable, et le guide pour identifier des comportements alternatifs souhaitables par l'analyse de ses processus cognitifs.

Les TCC* ont montré leur efficacité dans le traitement de phobies spécifiques : une étude de Öst [223] a par exemple montré qu'une exposition graduelle de 3 à 5 heures à un stimulus anxiogène suffisait à réduire les symptômes anxieux, y compris chez des phobiques sévères.

État de l'Art

Les TCC* sont conçues autour du principe d'exposition [5]; s'il existe plusieurs façon d'exposer un patient à un stimulus (e.g. imaginer une situation spécifique, jouer une scène avec le thérapeute dans le rôle du stresser), l'exposition *in vivo* (i.e. affronter en réel la situation anxiogène, comme prendre l'avion pour un aviophobe) s'avère plus efficace que des techniques dépendantes de la capacité des patients à imaginer une scène. Elle est cependant difficilement contrôlable par le thérapeute. Bouchard [29] présente l'exposition *in virtuo* (i.e. en réalité virtuelle, RV*) comme une alternative intéressante à l'exposition *in vivo*, citant notamment comme avantages la standardisation et sécurisation de la mise en situation, le contrôle du niveau d'intensité, la protection de la confidentialité patient-thérapeute (par rapport à des mises en situation *in vivo* impliquant des personnes tiers), et la réduction des coûts pour le patient (e.g. dans le cas de l'aviophobie, la peur de l'avion, il n'est pas nécessaire d'acheter des billets). Si les coûts d'acquisition de matériel pour le thérapeute restent élevés, le développement et la généralisation de ces techniques tendent à les réduire. Côte [68] ajoute que la RV* permet de dépasser la réalité dans l'exposition : un patient souffrant d'aéroacrophobie, la peur des hauteurs, pourra aller jusqu'à sauter dans le vide. Garcia-Palacios et. al. [107] ont par ailleurs montré que les patients sont plus enclins à accepter une thérapie en RV* qu'une thérapie *in vivo*, car elle leur paraît moins menaçante.

Les études comparatives [222] [226] [167] [242] [191] sur l'efficacité de thérapies d'expositions virtuelles par rapport à des expositions *in vivo* tendent à montrer, pour des phobies spécifiques, des résultats comparables entre les deux techniques et une réduction manifeste des symptômes anxieux par rapport à un groupe de contrôle.

L'utilisation de la RV* pour le traitement des phobies sociales a également été validée pour la prise de parole en public [4] et notamment face à différents types d'audiences virtuelles ayant une attitude positive, négative ou neutre vis-à-vis de la performance des sujets [235]. Une étude de Klinger et. al. [159] comparant exposition *in vivo* et en RV* a également montré la validité de l'exposition virtuelle sur 4 scénarios anxiogènes pour les sociophobes: affirmation de soit (e.g. défendre son point de vu, ses intérêts), performance (e.g. prise de parole en public), intimité (e.g. discussion avec des proches) et observation (e.g. agir tout en étant observé).

Les études sur l'usage de TCC* par exposition *in virtuo* mesurent le stress des participants pre- et post-traitement par des questionnaires d'évaluation du stress [203][81], et dans certains cas par des relevés physiologiques au cours de l'expérience : Wiederhold utilise ces relevés physiologiques pour fournir un « biofeedback » au sujet, un retour visuel de son état physiologique [326][327]. A notre connaissance, il n'existe pas de système utilisant d'autres mesures pour la détection du stress, par exemple dans la voix ou dans l'attitude corporelle.

1.4 Réponse au stress : coping et appraisal

Nous avons établi l'impact potentiellement néfaste du stress sur le corps lorsque celui-ci est ressenti de façon prolongée, et sur la vie quotidienne lorsque l'anxiété ponctuelle devient une phobie. S'il n'est pas envisageable de supprimer la réponse physiologique qu'est le stress, nous ne sommes pas pour autant sans défenses face à la menace qu'elle peut représenter.

Le « coping » est un mécanisme de régulation, défini par Lazarus comme « un ensemble d'efforts cognitifs et comportementaux visant à maîtriser, réduire ou tolérer des demandes spécifiques internes et/ou externes, vécues par le sujet comme menaçantes, épuisantes ou dépassant ses ressources » (traduit de [178]). Il s'agit des stratégies que nous mettons en place pour supporter le stress auquel nous nous trouvons confrontés. La notion de coping intègre aussi bien des stratégies réactives (i.e. en réponse à un événement stressant) que proactives (i.e. en anticipation d'un événement stressant), et inclut aussi bien des réactions comportementales que cognitives [280][279].

De nombreuses études se sont attachées à déterminer des dimensions afin de catégoriser les multitudes de stratégies de coping existantes [162][287]. On peut les présenter selon deux méta-catégories : les stratégies de coping positives ou fonctionnelles, et les stratégies de coping négatives ou dysfonctionnelles.

Les stratégies négatives incluent les attitudes de déni et d'évitement des situations stressantes, la fuite – dont le recours à l'automédication – ou encore les troubles anxieux. Ces stratégies sont qualifiées de dysfonctionnelles car si elles éliminent temporairement les symptômes du stress, elles ne font que renforcer la réaction négative lorsque la situation stressante se présentera à nouveau.

A l'inverse, les stratégies fonctionnelles sont des mécanismes permettant d'appréhender une situation stressante de façon positive ; elles incluent l'anticipation d'un événement afin de s'y préparer plus efficacement, la recherche de support social (en partageant un événement stressant avec des proches) ou moral (en cherchant un sens à l'évènement stressant), l'humour (en dédramatisant l'évènement stressant), ou encore la pratique d'une bonne hygiène de vie pour aider le corps à répondre aux demandes physiologiques du stress [32]. Lazarus et Folkman [177] distinguent deux procédés de coping : l'un centré sur les actions que peut entreprendre un individu pour agir sur le problème ayant déclenché son stress afin de le maîtriser, on parlera de coping situationnel ; l'autre centré sur la régulation de la réponse émotionnelle au problème, on parlera de coping émotionnel.

État de l'Art

L'un des éléments qui rend la réaction de stress si individuelle est la *perception* de l'évènement stressant comme une menace. Comme nous l'avons expliqué précédemment pour définir les phobies, cette perception peut être largement subjective. Dès lors, notre capacité à appréhender les situations auxquelles nous sommes confrontés -- et l'impact qu'elles auront sur nous -- va forcément influencer notre capacité à gérer le stress. Le concept d'« appraisal » a été introduit dans la recherche sur les émotions par Arnold [6], puis repris par Lazarus [174] dans le cas spécifique du stress, et par Scherer [258]. Ils définissent l'appraisal comme la réponse émotionnelle d'un individu à un évènement, régie par la perception de l'individu de l'impact que cet évènement a sur lui.

Cette définition explique les différences de degré d'intensité et de durée du stress que peuvent ressentir deux individus confrontés à une même situation : l'un peut percevoir la situation comme un challenge et en tirer une grande motivation, tandis que l'autre peut la percevoir comme un obstacle insurmontable et en tirer un fort découragement [229]. Cette réponse est elle-même influencée par un certain nombre de facteurs à la fois liés à la personnalité et au contexte. Lazarus [176] liste les facteurs personnels les plus influents : la motivation, les buts, les valeurs et les attentes ; et les facteurs contextuels les plus influents : la prédictibilité, la contrôlabilité et l'imminence de l'évènement stressant.

Les concepts de coping et d'appraisal se chevauchent, Gratch et Marsella [112] expliquent que les stratégies de coping déterminent la façon dont nous répondons à une situation que nous avons appréhendée comme ayant un impact significatif sur nous. Lazarus [175] décrit cette relation comme un phénomène en trois temps : l'appraisal primaire où l'individu évalue en quoi un évènement impacte ses valeurs et ses buts, l'appraisal secondaire où l'individu évalue l'influence qu'il peut exercer sur l'évènement afin de le réaligner avec ses buts et valeurs (et suivant ses stratégies de coping), et finalement la réévaluation de la situation suite à la perception du contrôle exercé sur l'évènement stressant. Scherer [258] considère également l'appraisal comme un processus récursif où une situation stressante est réévaluée au vu des solutions envisageables pour y remédier.

1.5 Facteurs d'influence dans l'expérience et l'expression du stress

Nous avons vu dans les sous-chapitre 1.2 et 1.4 que la réponse au stress est différente pour chaque personne ; dans cette section, nous nous intéressons aux éléments qui peuvent expliquer cette variabilité. Nous tiendrons compte à la fois de la variabilité du stress ressenti et du stress exprimé par un individu, et du fait que ces deux mesures ne sont pas forcément égales.

Les stratégies de coping mises en place par chaque individu dépendent de facteurs à la fois socio-culturels, contextuels et personnels. Dans *'Etude Sociologique du Stress* [230], Pearlin présente le processus de réponse au stress comme étant contrôlé par trois concepts: les stressseurs, les modérateurs, et les impacts. Les *stressseurs* sont tous les facteurs provoquant un changement et un besoin d'adaptation chez l'individu, qu'ils soient externes (i.e. les pressions environnementales ou sociales comme un déménagement ou un nouveau poste à responsabilités) ou internes (i.e. les réactions biologiques ou psychologiques comme la ménopause ou une dépression). Il peut tous aussi bien s'agir d'évènements ponctuels (e.g. le cambriolage du domicile) que d'évènements chroniques (e.g. un parent dépendant). Les *impacts* sont les effets résultant de l'exposition de l'individu aux stressseurs après médiation des modérateurs. Ils englobent aussi bien les résultats physiques (e.g. des troubles musculo-squelettiques) que psychologiques (e.g. un trouble anxieux déclenché par l'exposition à des stressseurs similaires). Les *modérateurs* sont définis comme l'ensemble des ressources d'un individu qui vont atténuer l'impact des stressseurs. Il peut s'agir de ressources sociales (e.g. le soutien émotionnel de proches) ou individuelles (e.g. la confiance en soi).

La variabilité des stressseurs peut expliquer la diversité de perception du stress. Dohrenwend définit les évènements stressants comme « toute situation objective suffisamment importante pour modifier les activités normales de la plupart des personnes » (traduit de [84]). Holmes et Rahe [135] ont développé une liste de 43 évènements bouleversant de façon plus ou moins importante la place d'un individu dans sa structure sociale (du décès d'un conjoint à un départ en vacances) afin de comparer la façon dont les individus évaluent l'impact de ces évènements sur eux-mêmes. Cette échelle a été utilisée pour comparer l'évaluation du stress généré par de mêmes évènements sociaux par des populations de milieux culturels différents, révélant l'aspect non-universel de l'importance perçue d'un évènement [195][160][120][298].

État de l'Art

Si un réseau social solide est un élément essentiel pour le coping fonctionnel [210], les pressions sociales peuvent également se révéler source de stress. Aneshensel [10] définit la notion de stress social, déclenché par le non respect des normes sociales en vigueur dans la communauté d'un individu, résultat inévitable de toute structure sociale. Mirowsky et Ross [210] se sont intéressés aux stressseurs produits par le milieu social d'un individu dans une culture occidentale. Leurs résultats suggèrent quatre types de stressseurs sociaux : la position dans l'échelle sociale (les individus ayant un statut plus important sont moins stressés), le genre et le rôle social associé (les femmes sont plus stressées que les hommes), le statut marital (les personnes mariées sont moins stressées que les personnes divorcées, veuves, ou célibataires), et la fréquence de confrontation à des événements indésirables (plus ces événements sont fréquents plus les individus sont stressés).

Les normes socioculturelles qui influencent notre perception en « créant » à la fois du stress et des ressources pour y répondre, modifient également notre expression. La théorie sociocognitive de Bandura [14] présente la régulation des réponses émotionnelles comme des comportements appris par l'assimilation et la réplication des modèles comportementaux auxquels nous sommes exposés de façon répétée. En particulier, la phobie sociale se caractérise comme un trouble cognitif déclenché par les conventions sociales, et la peur de ne pas les respecter [61][244].

Scherer [256][260] propose de modéliser ces influences par un « Modèle en Lentille » adapté de Brunswick [35]. Il présente un processus contrôlé par deux concepts : les mécanismes « push » et les mécanismes « pull ». Les push sont l'externalisation réflexe de processus biologiques internes – comme un éclat de rire ou un cri de peur – et par nature fortement variables d'un individu à l'autre. Les pulls sont des modèles comportementaux définis par des normes socioculturelles qui imposent un cadre strict à l'expression par des codes de communications préétablis, allant des règles grammaticales et syntaxiques de langage aux attitudes considérées comme « correctes » dans un contexte donné (e.g. ne pas parler fort dans un lieu public). Ces comportements sont symboliques et fortement conventionnels, et partagés par l'ensemble des individus d'un même groupe.

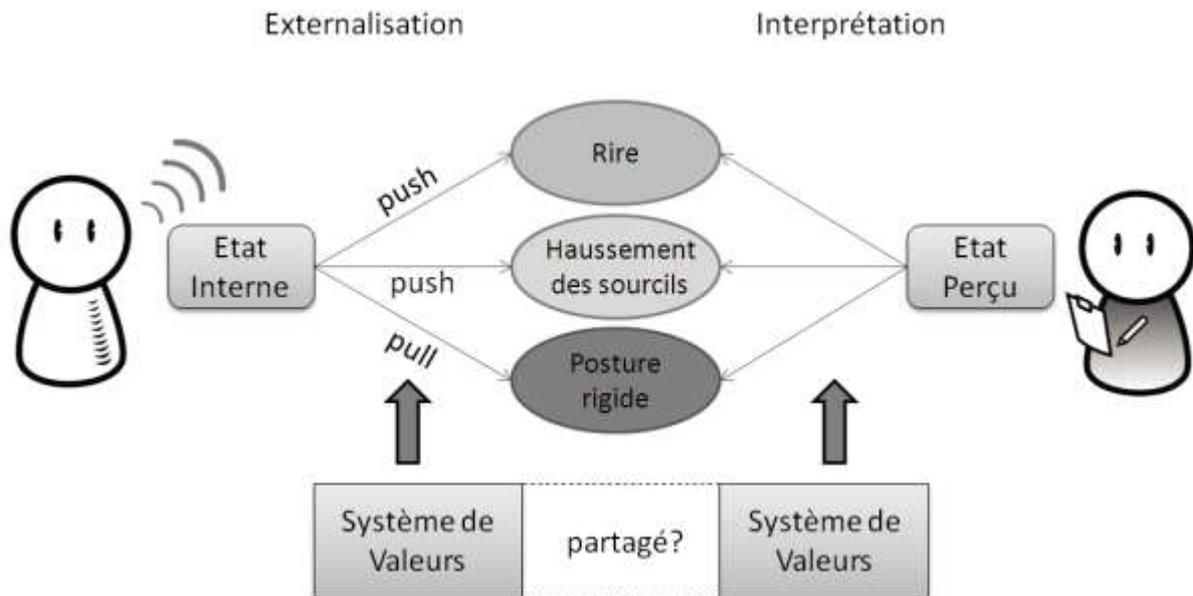


Figure 2 : Modèle de communication « en lentille », adapté de Scherer

La Figure 2 illustre ce processus. La partie gauche représente l'encodage de l'état émotionnel d'un individu par un certain nombre de caractéristiques qui vont nuancer son message ; la partie droite représente le décodage du message et l'interprétation des caractéristiques exprimées (ou non exprimées) pour inférer l'état émotionnel sous-jacent au message. L'influence d'un modèle de valeur – et les implications d'un modèle non partagé entre les deux extrémités du processus – sur l'expression et sur l'interprétation des « traces émotionnelles » du message est explicite. Par exemple, dans un contexte où le système de valeur d'un individu dicte une attitude sérieuse mais où celui-ci est amusé, celui-ci va contenir son rire par une posture rigide. Un observateur ne partageant pas ses valeurs (i.e. ne percevant pas la situation comme « sérieuse ») pourrait interpréter son comportement comme de la gêne.

L'influence du milieu socioculturel sur la perception des stresseurs n'est pas suffisante pour expliquer la variabilité des réactions au stress ; deux individus issus d'un même milieu – et ayant accès aux mêmes ressources de coping – confrontés à un même évènement peuvent réagir de façons très différentes. Ces différences intra-culturelles peuvent être expliquées par les différences interindividuelles liées aux traits de personnalité propres à chacun. La modélisation de la personnalité est généralement basée sur cinq dimensions regroupées sous le terme de Modèle Cinq Facteurs (Five-Factor Model, FFM): il s'agit de l'Ouverture (appréciation des nouvelles expériences et idées), de la Conscienciosité (autodiscipline et poursuite de buts), de l'Extraversion (spontanéité et recherche du contact social), de l'Agréabilité (désir de coopération) et du Névrotisme (stabilité émotionnelle et tendance aux émotions négatives). Ces cinq dimensions sont notamment exploitées dans les modèles des Big Five [133] et OCEAN* [197].

État de l'Art

Plusieurs études ont montré un lien entre traits de personnalité et propension aux émotions : l'extraversion est généralement associée aux émotions positives [322], tandis que le névrotisme est lié aux émotions négatives [170][141]. De façon plus indirecte, l'extraversion est associée à un désir de contact et de statut social, et le névrotisme à un désir de calme et de tranquillité [220]. Ces désirs vont agir comme des pull dans le modèle Lentille de Scherer et encourager ou réprimer l'expression de certaines émotions.

Les études des liens entre personnalité et choix des stratégies de coping ont montré des corrélations positives entre névrotisme et stratégies dysfonctionnelles d'évitement [317], entre optimisme et stratégies fonctionnelles par l'action [255], et l'influence des traits névrotisme, ouverture et extraversion sur la façon d'appréhender les événements stressants et le choix de stratégies fonctionnelles par l'action [265].

Chapitre 2 L'expression du stress

2.1 Mesures de facteurs individuels

Nous avons présenté le stress au Chapitre 1 comme une réaction fortement spécifique à un individu, dépendant à la fois de facteurs personnels et socioculturels qui vont influencer à la fois sa perception et son expression. Il convient alors de caractériser ces éléments d'influence pour observer le stress chez des individus de profils comparables (ou au contraire, entre différents profils).

Schwarzer et Schwarzer [280] se sont intéressés aux outils de mesure des stratégies de coping et comparent 13 questionnaires, montrant une grande variabilité dans les échelles proposées. À l'évidence il n'existe pas une unique « bonne » échelle de mesure des stratégies de coping et le choix de l'instrument de mesure à utiliser est dépendant de la population ciblée et du type d'utilisation souhaité de cette mesure.

Certains questionnaires [196][163][24][234] sont composés autour du type de situation stressante, les auteurs considérant la stratégie de coping comme dépendante du stresser. D'autres questionnaires [208][104][8][100][92] cherchent à regrouper différentes stratégies de coping en méta-catégories suivant l'approche de Lazarus (coping par l'action pour la résolution de problème, et coping émotionnel pour altérer la perception de l'évènement) [174]. Ces méta-catégories sont variables, mais incluent majoritairement les notions de « confrontation du problème » et d'« évitement du problème ». À l'inverse d'autres questionnaires préfèrent un grand nombre de catégories de stratégies pour distinguer spécifiquement les comportements adoptés en termes de confrontation ou d'évitement [46][293].

Deux questionnaires [226][82] ciblent en particulier les adolescents et les enfants, dont l'expérience du stress est différente de celle des adultes de par leur développement cognitif encore en cours. Ces deux échelles distinguent 12 et 5 stratégies, respectivement.

État de l'Art

Les stratégies de coping et d'appraisal ont également été étudiées pour générer des comportements réalistes dans des personnages virtuels animés : Marsella et Gratch [194] proposent un modèle à six stratégies de coping (planification, acceptation, recadrage positif, désengagement, déni, rejet ou acceptation de la responsabilité) pour modéliser le comportement d'un agent ; Ochs et al. [218] proposent un modèle d'appraisal d'évènements en quatre points (conséquences de l'évènement sur les buts de l'agent, causes de l'évènement, la cohérence des conséquences de l'évènement et des attentes de l'agent, la capacité à gérer les conséquences de l'évènement) pour implémenter un comportement empathique chez un agent d'un système de messagerie ; Faur et al. [99] proposent un modèle de comportement pour un compagnon virtuel basé sur les théories d'autorégulation : l'agent réagit à un évènement en interprétant son impact sur ses désirs et ses croyances, suivant une stratégie de promotion (i.e. favoriser la réalisation de ses aspirations) ou de prévention (i.e. limiter les impacts négatifs sur ses objectifs).

L'influence des traits de personnalité sur l'expérience du stress a été présentée dans le sous-chapitre 1.5, et a montré l'intérêt de mesurer ces facteurs. De même que pour la mesure des stratégies de coping, il existe de nombreuses échelles d'évaluation des traits de personnalité.

Un certain nombre des échelles existantes sont conçues pour mesurer les 5 traits de personnalité proposés par le FFM psychologue Lewis Goldberg [110] et communément appelés « modèle OCEAN* » : Elles varient par le nombre de questions (de 10 à 240), la forme des questions (un questionnaire [67][190][145][243] ou un entretien semi-ouvert [305]) et la population ciblée (une échelle spécifique a été développée pour les enfants [159]).

D'autres échelles mesurent des traits semblables à ceux des théories du FFM : le Minnesota Multiphasic Personality Inventory [121] est utilisé pour le diagnostic de la psychopathologie et mesure l'agressivité, le psychotisme, le refus des contraintes, le névrotisme et l'introversion ; l'Eysenck Personality Questionnaire [97] mesure le psychotisme, le névrotisme, l'extraversion, et la propension au mensonge ; l'HEXACO Personality Inventory [11] mesure les cinq traits FFM et y ajoute le trait d'honnêteté/humilité. Le 16PF Questionnaire [53] mesure 16 traits de personnalité, et dans ses versions les plus récentes [54] propose une correspondance avec des traits similaires à ceux des théories du FFM (Extraversion, Anxiété, Ouverture d'esprit, Réceptivité, Contrôle de soi) mais tenant compte de la corrélation de ces 5 traits les uns avec les autres.

Le Myers-Briggs Type Indicator [216] est un questionnaire couramment utilisé dans le milieu professionnel pour évaluer la prise de décision, et distingue 16 types de personnalités suivant la combinaison de 4 facteurs : ressource d'énergie (Introversion/Extraversion), recueil d'information (Sensation/Intuition), prise de décision (Emotion/Réflexion), et appréhension de la situation (Perception/Jugement).

État de l'Art

La spécificité de la phobie – et en particulier de la phobie sociale – dans l'expérience du stress a été présentée dans le Chapitre 1, et il convient de faire l'inventaire des outils de mesure de l'anxiété et de la phobie. Ce sont des questionnaires relativement courts (10 à 30 questions), utilisés soit en auto-évaluation par le patient, soit utilisés par un médecin psychiatre pour conduire un entretien d'évaluation

Certains questionnaires évaluent l'anxiété de manière générale : le Generalized Anxiety Disorder-7 [291] est utilisé pour le diagnostic du Syndrome d'Anxiété Généralisé ; le Fear Questionnaire [193] mesure 3 phobies (l'agoraphobie, la phobie sociale et la phobie du sang ou des blessures) ; le Hamilton Anxiety and Depression Scale [117] tient compte de la forte comorbidité entre trouble anxieux et dépression et est utilisé pour évaluer la sévérité du trouble anxieux chez un patient déjà diagnostiqué.

D'autres sont conçus afin d'évaluer spécifiquement la phobie sociale, soit à des fins de diagnostic [183][126], soit afin de mesurer la sévérité du trouble [324][66]. Les différents éléments des questionnaires sont utilisés pour évaluer le type de situations sociales qui déclenchent chez le patient un trouble anxieux, et le type de situations sociales que le patient cherche au maximum à éviter. Le Questionnaire d'Ereutophobie [232] mesure en particulier la peur du rougissement en public, considérée comme une sous-catégorie de phobie sociale.

2.2 Analyse de signaux pour la caractérisation du stress

2.2.1 Dans la physiologie

S'il existe des variations individuelles, socioculturelles et contextuelles dans l'expérience et l'expression du stress, nous pouvons considérer un certain nombre de signes communs, à commencer par les modifications physiologiques déclenchées par le système nerveux orthosympathique, et les sensations physiques qui en résultent : tension musculaire, bouche sèche, l'estomac « qui se retourne », l'impression « d'entendre son cœur battre dans ses oreilles », la vision « en tunnel »... Puisque le stress est d'abord une réaction physiologique, il est naturel de chercher à mesurer le stress selon les variations physiologiques d'un individu. Plusieurs études ont montré l'efficacité de ces mesures – en particulier les variations des rythmes cardiaque et respiratoire, la conductance de la peau, et le taux de cortisol présent dans le sang ou la salive – pour la distinction des états émotionnels [109] ou la reconnaissance du stress [217][123] en laboratoire.

L'inconvénient principal de ces mesures est le caractère invasif des capteurs impliqués : s'il est envisageable de relier des sujets expérimentaux à des appareils de mesure en laboratoire et d'avoir du personnel expérimenté à disposition pour placer les capteurs correctement, ce procédé devient contraignant pour des mesures « en milieu naturel ». Par ailleurs, ces capteurs auront potentiellement un « effet Blouse Blanche » — une augmentation de la pression sanguine en présence du médecin chargé de la mesurer [192] — non négligeable sur les sujets, exacerbant leur stress et biaisant l'expérience. Des études sont menées pour obtenir ces mesures physiologiques de façon plus discrète, en utilisant un bracelet autonome [252] ou même sans contact via une image vidéo [240].

2.2.2 Dans la voix

La voix est le second canal d'information étudié – après la physiologie – pour la reconnaissance du stress, suivant le constat que les modifications physiques engendrées par le stress (tensions musculaires, modifications de la respiration) vont affecter le conduit vocal, instrument de production de la voix. De façon générale, la voix et les variations de ses attributs acoustiques ont été largement étudiées pour caractériser des états affectifs (émotions catégorisées de façon discrète [15] [316] [95] [227] [300] ou dimensionnelle [171] [94]), des états physiques (ébrioité [308], douleur [31], maladie neuro-dégénérative [134]), des conditions psychologiques (schizophrénie [42], dépression [241] [215]) ou encore des traits de personnalité [34] [256] [273]. Les paramètres acoustiques les plus utilisés portent sur la fréquence fondamentale, l'énergie, les paramètres spectraux (Linear predictive coding LPC) et cepstraux (Mel-frequency cepstral coefficients MFCC*), et le rythme de la parole. Il s'agit d'indices également utilisés dans la reconnaissance des émotions dans la voix [299][270][59], mais aussi pour la caractérisation d'autres formes de son (e.g. classification de genres musicaux [307], des bruits domestiques [281]).

Les spécificités du stress ont également été investiguées : les travaux de Scherer et. al. [257][261] sur les caractéristiques acoustiques des voix stressées ont valorisé l'utilisation du rythme, des contours de l'énergie, de la fréquence fondamentale et des paramètres spectraux pour la reconnaissance du stress induit par une charge cognitive (i.e. la réalisation d'une tâche tout en étant soumis à des stimuli susceptibles de détourner l'attention). Ruiz et. al. [250] ont utilisé la fréquence fondamentale et les formants pour distinguer différents degrés de stress sur des données induites et réelles. Les travaux de Zhou [333] et Fernandez et Picard [102] ont valorisé l'utilisation du Teager Energy Operator (TEO) – proposé par Teager [304] pour modéliser les variations du flux d'air dans le conduit respiratoire – et de mesures dérivées de cet indicateur pour la reconnaissance de stress réel et induit.

Des travaux sur l'influence de l'anxiété sur la parole ont révélé des spécificités dans le discours d'individus en état d'anxiété ponctuel ou généralisé : augmentation des disfluences [149][182], trémolos dans la voix [288], diminution de la fréquence fondamentale et de ses variations et ralentissement du débit de parole [115]. Après traitement du trouble anxieux, ces paramètres reviennent à des valeurs standard [172].

2.2.3 Dans les expressions faciales

Si la voix est un canal riche en informations affectives, les silences et la communication non-verbale peuvent être tout aussi porteurs de sens. Il convient alors de tenir compte d'autres canaux de communication pour trouver des indices susceptibles de traduire l'état affectif d'un individu. Charles Darwin [72] théorise que les expressions faciales des émotions chez les humains sont le résultat de notre évolution, et trouvent leur origine dans des comportements animaux jouant un rôle favorable dans la sélection naturelle. De ce constat découle la notion d'universalité des expressions faciales, transcendant les différences culturelles. Si cette notion est discutée, elle est à la base des travaux de Paul Ekman sur l'encodage des expressions faciales. Ekman a développé le système Facial Action Coding System (FACS*) permettant de décrire toute expression comme la combinaison de l'activation de différents muscles du visage, ou Actions Units (AU) [90].

Les systèmes de reconnaissance automatique des expressions faciales cherchent à identifier les expressions des 6 émotions primaires identifiées par Ekman : peur, colère, joie, tristesse, dégoût, surprise. On peut distinguer deux approches, l'une statique à partir d'images fixes, l'autre dynamique à partir d'une suite chronologique d'images [143]. Quelle que soit l'approche, le système commence par identifier la présence et la position d'un visage dans l'image, puis cherche à identifier la bouche, les yeux et les sourcils, éléments mobiles dans un visage humain, pour en étudier la position relative. Cette identification se fait soit par la recherche de formes de la bouche, des yeux et des sourcils (on parle de modèle d'« apparence » [144][21]), soit par la recherche de points (e.g. les coins des yeux, de la bouche) ou l'application d'un modèle « fil-de-fer » d'un visage humain (on parle alors de modèle « géométrique » [310][189]), ou éventuellement par une combinaison de ces approches. A partir des indices collectés, le système confronte l'apparence actuelle du visage à un modèle pour estimer quelle est l'expression la plus probable. Le système FACS* est également utilisé pour la *génération* d'expressions faciales réalistes chez des personnages virtuels [69][231][129].

Si plusieurs études ont montré l'influence de la dépression [207], de la schizophrénie [87], de l'anxiété [249] ou encore de la phobie sociale [146][108] sur la *perception* des émotions faciales, l'influence de ces pathologies et états anxieux sur l'expression affective dans les expressions faciales est rarement étudiée [230].

2.2.4 Dans la posture et les gestes

Chez l'homme, la danse est probablement l'exemple le plus explicite de l'expression d'émotions par la posture et le geste. Le danseur et chorégraphe Rudolf Laban [166] s'est d'ailleurs intéressé à la codification du mouvement expressif afin de pouvoir communiquer plus facilement ses attentes à ses danseurs. Il propose de décrire le mouvement suivant 4 dimensions : corps (quelles sont les parties du corps en mouvement, les connexions entre les différentes parties du corps), effort (la dynamique du mouvement pour traduire une intention), forme (les changements de forme du corps pendant le mouvement), espace (l'occupation de l'espace par le mouvement et les trajectoires suivies).

Les travaux de McNeill [205] et Cassel [49] [50] sur la codification du mouvement se fondent sur l'association d'un geste à une fonction communicative spécifique : iconique (pour décrire la forme ou le volume d'un objet), métaphorique (une représentation abstraite, comme l'utilisation d'un objet), déictique (désigne un point dans l'espace ou une personne), emphatique (accompagne la parole pour accentuer certains mots), ou emblématique (des gestes au sens sémantique convenus comme un « au revoir » de la main). A l'inverse, les travaux de Zhao [332] se détachent de la sémantique du geste et s'appuient sur les propositions de Laban d'effort et de force. Camurri et. al. [41] préfèrent également une qualification non sémantique du geste et de la posture, et proposent les critères de d'Indice de Contraction et de Quantité de Mouvement décrivant l'occupation du corps dans l'espace et dans le temps

Le lien entre mouvement et émotion a fait l'objet de plusieurs travaux cherchant à associer un type de mouvement à une classe d'émotion. Dès l'enfance les humains sont capables de distinguer la joie de la colère en s'appuyant sur des indices gestuels [28], et si l'attribution de gestes spécifiques à des émotions n'est pas universelle [320], la qualité de la gestuelle (i.e. vélocité, amplitude, fréquence...) contribue à la distinction d'états d'excitation ou d'apathie, d'ouverture ou de repli [76] [320].

L'analyse du mouvement pour l'extraction d'information affective a été transférée aux machines [156], soit par des techniques de « Motion Capture » [148], soit par des techniques moins invasives de capture vidéo et plus récemment de capture de profondeur avec le déploiement massif de capteurs de moindre coût grâce à la Kinect [237]. Ces techniques ont été explorées pour la reconnaissance des émotions chez des danseurs [40], dans des gestes des bras actés [52], sur des données actées en fusion avec des indices audio [153] et faciaux [51]. L'influence du milieu culturel sur la perception [157] et l'expression [44] du lien entre émotion et mouvement a également été étudiée. A notre connaissance il n'existe pas d'étude spécifique entre mouvements du corps et stress.

2.2.5 Dans les signaux sociaux

L'expression émotionnelle de l'homme se fait à travers plusieurs modalités, et ce de façon simultanée. La voix, le visage et le corps se synchronisent pour communiquer notre état affectif et moduler notre message dans les interactions sociales.

Scherer décrit la notion d' « Affect Burst »² comme des « expressions affectives brèves, discrètes, non verbales simultanément dans la voix et sur le visage, en réaction à un événement clairement identifiable » (traduit de [259]). Cette définition couvre les interjections, (de surprise, de dégoût, de déception, de compréhension...) et les moues et gestes qui les accompagnent, les soupirs, les bâillements, ou encore les rires. Schröder [266] a étudié le lien entre certains affects bursts et 10 catégories d'émotions, et montré que les humains parviennent à identifier très clairement une émotion à partir d'un affect burst audio. Simon et. al. [286] ont également utilisé les affects bursts audio pour discriminer 9 émotions négatives et 13 émotions positives mêlant émotions basiques (joie, colère...) et sociales (honte, culpabilité...). Les études tenant compte de l'expression faciale des affects bursts sont plus rares [164] [306] et montrent un impact significatif sur l'interprétation du signal : un froncement de sourcils combiné à un signal vocal de tension est interprété comme de l'aversion, mais ces signaux pris de façon indépendante ne sont pas perçus comme de l'aversion [128].

² Littéralement « explosion affective »

Chapitre 3 Apprentissage automatique pour la reconnaissance des états affectifs

3.1 Affective Computing

L'« Affective Computing »³ est un terme initialement proposé par Rosalind Picard [239] pour décrire un domaine de recherche axé sur l'étude et la conception de systèmes capables de reconnaître, interpréter, analyser ou simuler des états affectifs humains. C'est un champ d'étude vaste, à l'intersection de plusieurs disciplines de traitement et de synthèse automatique du signal et s'appuyant sur les sciences humaines pour appréhender les états affectifs selon un angle psychologique, sociologique ou encore cognitif.

Les données susceptibles de fournir à un système les informations nécessaires pour comprendre l'état affectif d'un utilisateur humain sont multimodales (nous avons vu dans le chapitre 2.2 des informations physiologiques, audio, visuelles) et les interprétations possibles sont variables (suivant une échelle discrète ou continue, comme une quantité numérique ou comme une information symbolique...). L'espace de représentation de ces informations est par conséquent très complexe, et il est difficilement envisageable de le parcourir à l'aide de simples règles de décisions. Les techniques d'apprentissage automatique rendent cet espace plus facilement appréhendable, et ont largement contribué au développement de l'affective computing.

De manière générale, les techniques d'apprentissage dites « supervisées » consistent à observer un certain nombre d'exemples représentatifs des catégories que l'on souhaite distinguer, décrits chacun par un ensemble de descripteurs (e.g. la couleur, la forme, la taille...) appelé « vecteur », et une étiquette indiquant à quelle catégorie l'exemple appartient. Le système « apprend » un modèle des caractéristiques de chaque classe, qui sera ensuite confronté à d'autres exemples pour estimer la validité du modèle conçu. Le processus est illustré par la Figure 3.

³ Littéralement l'« Informatique Emotionnelle »

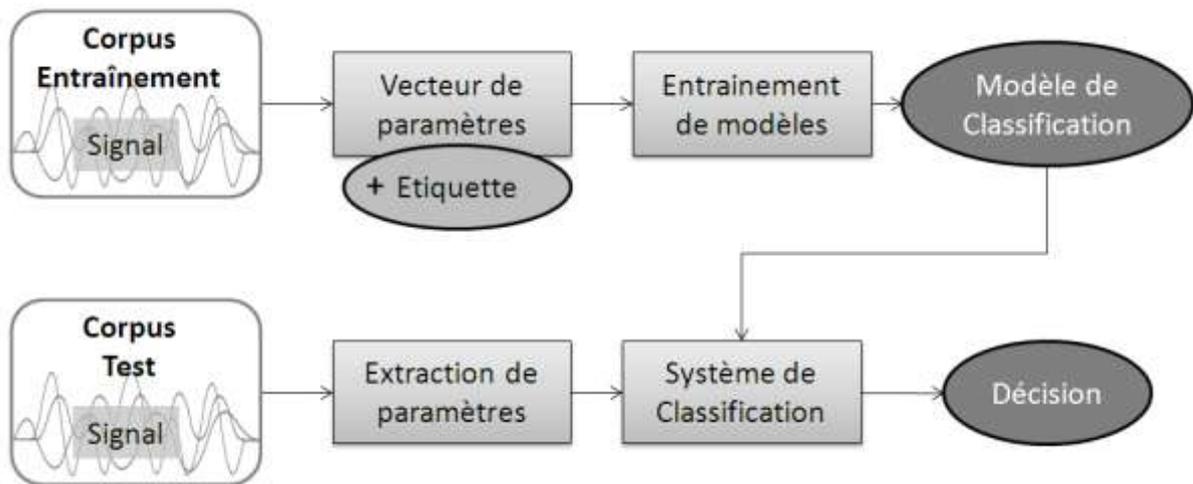


Figure 3 : Processus d'apprentissage et de classification de données

Dans le cas des techniques dites « non-supervisées », les vecteurs ne sont pas étiquetés et le système établit des catégories à partir des caractéristiques distinguant les vecteurs les uns des autres [26].

3.2 Corpus de stress existants et expérimentations stressantes

Le développement d'un système d'apprentissage nécessite l'accès à des données « représentatives » du concept à apprendre. La popularité croissante de l'Affective computing* a donné lieu à la constitution de nombreux corpus de données affectives de types variables : données actées ou en contexte, dans une ou plusieurs langues, mesures audio, visuelles, gestuelles... ou multimodales, annotation par catégories d'émotions de nombre variable, ou par dimension [85] [91] [17]. Le stress est parfois étudié comme l'une de ces catégories, mais il existe également quelques corpus constitués spécifiquement dans le but d'étudier le stress, comprenant des données « stressées » et « non-stressées », ou différents degrés d'intensité de stress.

Le corpus le plus fréquemment utilisé pour l'étude du stress dans la parole est le « Speech Under Simulated and Actual Stress » (SUSAS) [118]. Il s'agit de l'enregistrement audio de données en partie actées et en partie naturelles, collectées dans un contexte aéronautique (pilotes en conditions simulées et conversations naturelles en vol). Une partie des données est cantonnée à un vocabulaire restreint.

L'enregistrement du stress cognitif généré par une tâche complexe permet de constituer des corpus de données réalistes. C'est le cas des données utilisées par Fernandez [102], qui élicite du stress chez des conducteurs automobiles en leur faisant effectuer une tâche de calcul mental à voix haute. Le stress élicité par la conduite a également été confirmé par des mesures physiologiques [123].

État de l'Art

Le stress élicité par des pressions externes (les stressseurs présentés en 1.5) a fait l'objet d'études plus récentes. Zuo et. al. [337] utilisent un questionnaire pour interroger et enregistrer les réponses d'étudiants en période d'examen sur les événements stressants dans leur vie. Les données collectées comprennent des enregistrements audio au contenu libre, en anglais et en mandarin.

La constitution d'un corpus de données affectives soulève la question du type de données à enregistrer. La collecte de données « actées » (i.e. jouées par un acteur professionnel) présente un certain nombre d'avantages pour l'étude systématique d'un problème donné : le contenu linguistique peut être contrôlé, et chaque acteur produit l'intégralité des états affectifs que l'on souhaite observer pour étudier des différences spécifiques [18].

Cependant les études comparatives entre reconnaissance d'états affectifs actés et naturels révèlent que si les modèles sont performants pour la classification des données actées, ces performances ne se transfèrent pas sur des données naturelles [315]. Une explication est que les données actées sont prototypiques et clairement distinctes, alors que les données naturelles présentent plus souvent un mélange d'émotions et des degrés d'intensité plus faible [70]. Pour obtenir des données plus réalistes, certaines études utilisent des données actées en contexte réaliste issues de films [62] [80]. Les limites de cette approche restent la variabilité des contextes, et l'expression d'états « extrêmes » (e.g. expressions de terreur issues de films d'horreur) qui rendent la collecte de données comparables difficile. Les travaux récents tendent de plus en plus à utiliser des données « semi-naturelles » réalistes, obtenues à partir de protocoles expérimentaux conçu pour éliciter des réponses affectives chez les sujets [268] [173].

Il est éthiquement délicat, et en pratique compliqué d'étudier le stress élicité par des événements réels. Il existe des exemples de telles études sur des données issues d'enregistrements d'appels aux secours [78] [180] ou des enregistrements de pilotes en vol [117], mais la majorité se fait sur des données élicitées par des expérimentations [63].

Le cinéma est une source de stimuli émotionnels de tous types [301]. La provocation d'émotions en laboratoire à l'aide de stimuli externes a fait l'objet de plusieurs études [236] [113] [246] [254] cherchant à établir un set de films pour éliciter différentes catégories d'émotions (les 6 basiques d'Ekman et une variété d'émotions plus subtiles comme le contentement ou la tendresse) de façon fiable. Lazarus [179] a utilisé un procédé similaire pour éliciter du stress psychologique à l'aide d'un film « perturbant ».

État de l'Art

Outre l'exposition à des stimuli, le stress est provoqué expérimentalement par des tâches actives. Il peut s'agir de tâches physiques (e.g. plonger la main dans une eau glacée [202]) ou de tâches cognitives, le stress est alors en réponse à une demande de performance. Le test de Stroop [294] est l'une de ces tâches : lors de ce test, le sujet doit identifier la couleur d'un mot sans lire le mot lui-même, qui désigne une couleur non-congruente (e.g. le mot « rouge » écrit en bleu). L'impacte de la dissonance sur le processus cognitif du sujet est mesuré par son temps de réaction. Des variations de ce test ont été utilisées pour l'évaluation du stress induit par des conditions pathologiques (stress post-traumatique [204], dépression [328] [329]). Les tâches de calcul mental sont également utilisées pour induire du stress cognitif, seules [75] ou en parallèle d'une tâche motrice [102].

La Prise de Parole en Public (Public Speaking Task PST*) a été largement utilisée comme inducteur de stress « social », face à des humains ou face à des personnages virtuels [136] et sur tout types de public (adultes [155], adolescent [295] [325], phobiques sociaux [73] [132] [4]). Kirschbaum et al. [154] ont développé un protocole expérimental, le « Trier Social Stress Test » (TSST*), simulant les conditions d'un entretien d'embauche. Il s'agit d'une forme de PST* où l'évaluation de la performance, déclencheur majeur du stress dans ce type de tâche, est explicite puisqu'il ne s'agit plus simplement de parler à un public mais de convaincre un recruteur. Ce protocole a été largement utilisé pour l'induction de stress [165] [25]. Une variante du TSST* a également été développée pour l'étude du stress en groupe [318].

Si les données naturelles ou réalistes obtenues à partir d'expérimentations sont plus proches de l'expression non contrainte d'états affectifs au quotidien, leur utilisation pour l'entraînement de systèmes de reconnaissance automatique n'est pas simple, d'une part parce que ces expressions sont rares (à l'exception de situations émotionnellement fortes, les expressions affectives représentent environ 10% des échanges [23] [75]), d'autre part parce qu'à l'inverse des données actées leur contenu est souvent un mélange d'émotions et d'intensité, et doit être annoté pour être identifié.

Cette annotation réalisée par des humains est coûteuse en temps et nécessite l'intervention de plusieurs annotateurs -- dont la perception est subjective -- pour atteindre une catégorisation consensuelle [1]. L'intervention de plusieurs annotateurs ajoute le problème de la mesure d'accord entre leurs annotations pour estimer la fiabilité des étiquettes attribuées aux données. Cet accord est généralement mesuré par le Kappa de Cohen [64] ou l'une de ses variantes lorsque plus de deux annotateurs sont impliqués [292].

État de l'Art

La façon d'annoter les données va également influencer l'exploitabilité du corpus : le choix d'un schéma d'annotation (protocole décrivant à la fois les catégories possibles selon lesquelles distinguer les données et les marqueurs permettant aux annotateurs de justifier leur choix [86] [77] [266]), les conditions d'annotations (à partir d'une seule modalité ou de façon multimodale [263], avec des informations contextuelles ou « en aveugle » [38]), la façon de segmenter les données (suivant les tours de paroles d'un dialogue, suivant les pauses dans le discours, par segments de durée fixe) sont autant de paramètres qui vont impacter le corpus résultant.

3.3 Algorithmes de classification pour les données affectives

Parmi les nombreux modèles existants pour l'apprentissage automatique, les études portant sur leurs applications dans le domaine de l'affective computing utilisent majoritairement un (ou plusieurs, à des fins de comparaisons) des cinq suivants [122] :

K Plus Proches Voisins (K-Nearest Neighbors, KNN) : Le modèle KNN identifie la classe à laquelle appartient un vecteur de descripteurs donné en lui attribuant la classe majoritairement représentée parmi les k vecteurs les plus similaires vus lors de la phase d'entraînement. Deux paramètres permettent la configuration d'un algorithme KNN : k , le nombre de « voisins » auquel comparer un vecteur donné, et la mesure de distance utilisée pour comparer deux vecteurs.

Séparateurs à Vaste Marge (Support Vector Machine, SVM) : Les SVM sont un modèle de classification binaire⁴, utilisant la distance scalaire entre les vecteurs pour calculer un hyperplan divisant l'espace de représentation des données en deux sous-espaces représentant les deux catégories. Le meilleur hyperplan possible pour un problème donné se situe à la plus grande distance possible du plus proche vecteur de chacune des deux classes, c'est-à-dire l'hyperplan avec la plus grande marge entre les deux espaces. Pour simplifier le parcours d'un espace de très grandes dimensions, les SVM appliquent le « Kernel Trick » : plutôt que de calculer le produit scalaire entre deux vecteurs de très grandes dimensions, on le remplace par une fonction-noyau donnant une mesure de distance entre ces deux vecteurs. De ce fait, les SVM sont particulièrement adaptés pour des problèmes de classification dans des espaces complexes,

Modèle de Mélanges Gaussien (Gaussian Mixture Models, GMM) : Le modèle GMM est une modélisation statistique représentant chaque catégorie de données comme la somme pondérée d'un ensemble de fonctions de densité gaussiennes. Lorsqu'un échantillon inconnu est testé, le modèle évalue la distance de son mélange de densités gaussiennes à celui de chaque catégorie, et lui attribue la plus proche.

⁴ Les SVM peuvent être appliqués à des problèmes multi-classes en combinant plusieurs SVM pour chaque classe, chacun cherchant à séparer leur classe du « reste du monde »

État de l'Art

Modèles de Markov Cachés (Hidden Markov Models, HMM) : Le modèle HMM est également une modélisation statistique postulant que le système à modéliser est un processus de Markov (i.e. une suite d'états où l'état courant dépend du ou des états précédents) dont certains états sont « cachés ». Ce type de modèle est particulièrement adapté pour la représentation et l'identification de séquences temporelles et a été largement utilisé pour la modélisation de la parole.

Réseau de Neurones Artificiel (Artificial Neural Network, ANN) : Les réseaux de neurones artificiels sont un procédé de modélisation statistique non linéaire, reproduisant le fonctionnement – supposé – des neurones biologiques pour la reconnaissance de motifs (e.g. formes, séquences, récurrence...). Le réseau de neurones cherche à représenter une relation entre un vecteur de paramètres et la catégorie à laquelle il appartient, via les connexions entre les différents neurones qui le compose.

Les performances des différents types de modèles pour la classification des états affectifs ont fait l'objet de plusieurs études comparatives [224] [48] [331], montrant une variation d'amplitude des performances de 10 à 20% selon les modèles utilisés. Cependant il est difficile de déterminer un algorithme « idéal » à partir de la littérature, car les résultats présentés sont calculés sur des corpus différents (données actées ou naturelles), des problèmes de classification différents (classification de différents types d'états affectifs en un nombre de classes variables), et produisent des classements de performances des algorithmes différents.

3.4 Sélection de paramètres

L'utilisation d'un grand nombre de descripteurs pour caractériser les exemples fournis aux algorithmes d'apprentissage rend la construction d'un modèle capable de distinguer différentes catégories à partir de ces descripteurs complexes. Par ailleurs, les descripteurs ne contribuent pas forcément de manière équilibrée à la distinction des catégories. Il devient alors intéressant de faire une sélection des descripteurs les plus utiles à la classification avant d'entraîner un modèle.

État de l'Art

Dans le contexte de la classification automatique, nous pouvons distinguer deux approches pour la sélection de descripteurs : la sélection par filtre, et la sélection par wrapper [314]. La sélection par filtre a lieu en amont et indépendamment du procédé de classification, les paramètres sont choisis pour maximiser la distinction entre les classes et minimiser la variabilité au sein des classes. À l'inverse la sélection par wrapper identifie un sous-ensemble de paramètres maximisant le taux de bonnes reconnaissances dans le procédé de classification, et sera entre autres influencée par le type de modèle utilisé pour la classification. Nous pouvons également mentionner les approches dites « intégrées », où le procédé de sélection est directement intégré à la construction du modèle de classification [302].

La sélection de paramètre a été appliquée à des problèmes de reconnaissance des émotions [55] et a montré une influence positive sur les performances de ces systèmes [33]. Les résultats sont variables selon le type de classification utilisé, le type de descripteurs utilisés [23], et le type de données utilisés [315].

3.5 Fusion d'information pour une classification multimodale

Les études portant sur la compréhension des informations affectives dans la communication entre humains tendent à indiquer que nous étayons notre perception des émotions à partir d'informations provenant de multiples canaux (voix, contenu verbal, expressions du visage, regard, attitude corporelle...) [7]. Ces résultats confortent l'approche multimodale pour la reconnaissance automatique des états affectifs humains, mais la complexité du problème de classification augmente avec chaque nouvelle modalité apportant des descripteurs supplémentaires. Si les informations des différentes modalités sont regroupées dans un vecteur unique représentant l'exemple afin d'entraîner un modèle, on parle de fusion « précoce », car celle-ci a lieu avant l'apprentissage.

Il est possible de réduire la complexité du problème en traitant les informations apportées par chaque modalité séparément et en entraînant un modèle par modalité, puis en fusionnant les décisions de classifications prise par chaque modèle a posteriori. On parle alors de fusion « tardive », car celle-ci a lieu après l'apprentissage. L'un des intérêts de la fusion tardive, outre la complexité réduite du problème de classification, est qu'elle permet de considérer l'apport de chaque modalité à la décision finale de façon distincte, notamment en accordant à chacune un poids unique [184].

État de l'Art

Fusion par Vote : Il s'agit de la façon la plus intuitive de fusionner les classifications de plusieurs systèmes, en combinant les décisions de chaque système et en sélectionnant l'étiquette ayant obtenu le plus de « votes », c'est-à-dire la plus souvent sélectionnée par les différents systèmes. Si cette méthode est très simple, elle ne peut s'effectuer que sur un nombre impair de systèmes pour des problèmes de classification binaire, au risque d'obtenir une égalité parfaite entre les deux étiquettes. Pour contourner ce problème, il est possible de voter de façon pondérée, en attribuant un poids à chaque système, par exemple suivant ses performances, cependant cette pratique nécessite la connaissance a priori des performances.

Règles de Combinaison Algébriques : Plutôt que de faire la somme des votes des différents systèmes, il est possible d'utiliser les probabilités que chaque système attribue à chaque catégorie et de les comparer selon différents critères algébriques : le critère « maxima » (resp. « minima ») sélectionne pour chaque catégorie la probabilité la plus élevée (resp. la moins élevée) parmi les différents systèmes; le critère « somme » additionne les probabilités de chaque catégorie fournies par les différents systèmes; le critère « moyenne » fait la moyenne des probabilités, et le critère « moyenne pondérée » calcule la moyenne pondérée par les poids accordés à chaque système ; le critère « produit » fait le produit des probabilités pour chaque catégorie. Ces différentes comparaisons permettent de classer la probabilité « fusionnée » de chaque catégorie, et de sélectionner la plus probable.

Cascade de spécialistes : La fusion en cascade n'agrège pas les décisions de façon linéaire, mais considère que parmi les différents systèmes classifiant les données certains sont plus « experts » que d'autres pour certains problèmes. Lors de la phase d'entraînement, cette méthode consiste à identifier pour chaque catégorie le système le plus performant, puis à estimer les performances atteintes pour chaque catégorie, de la moins bien reconnue à la mieux reconnue. En phase de test, l'échantillon inconnu est confronté au meilleur système pour la classe la moins bien reconnue : si l'échantillon correspond, cette étiquette lui est attribuée, sinon il est confronté au meilleur système pour la deuxième catégorie la moins bien reconnue, et ainsi de suite jusqu'à atteindre un consensus.

Enfin il est possible de considérer les décisions prises par les modèles entraînés pour chaque modalité comme des descripteurs à fournir comme données d'apprentissage à un nouveau modèle. On parle alors de fusion « hybride ». L'espace à explorer pour le second modèle est largement réduit, puisqu'il est constitué d'un seul descripteur par modalité dont les valeurs se limitent aux étiquettes des catégories à apprendre.

État de l'Art

Les études portant sur la fusion d'informations affectives provenant de plusieurs modalités montrent généralement un gain des performances de classification par rapport à des systèmes n'utilisant qu'une seule modalité [147] [114] [51] [71]. Les résultats sont dépendants du type de modalités utilisé, de la clarté des données utilisées [319], et de la méthode de fusion appliquée [184].

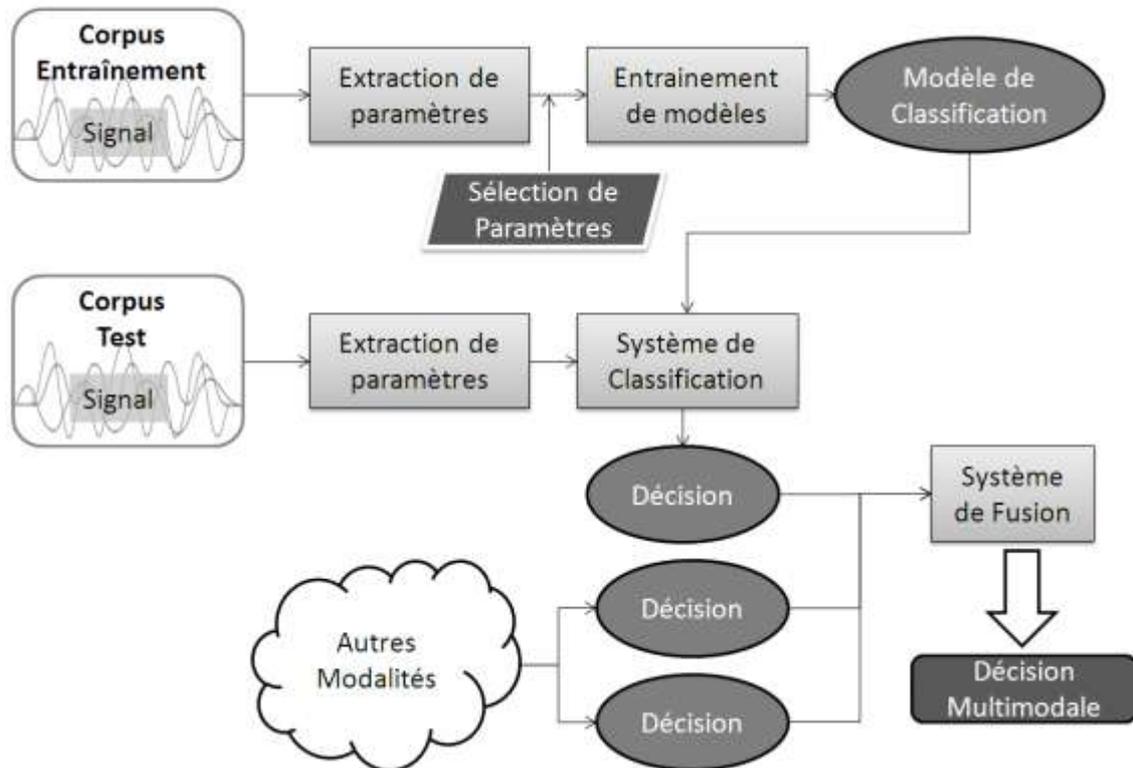


Figure 4 : Entraînement et utilisation d'un système de reconnaissance automatique avec fusion tardive

La Figure 4 schématise le procédé d'entraînement puis d'utilisation d'un système de classification pour des signaux affectifs multimodaux. La partie haute de la figure illustre l'entraînement du modèle à partir de paramètres extraits d'un signal, après éventuelle sélection des paramètres. La partie basse montre la classification d'un échantillon multimodal : les paramètres sont extraits du signal et confrontés au modèle précédemment appris pour les différentes modalités, puis les décisions de chaque modalité sont fusionnées pour aboutir à la classification finale de l'échantillon testé.

3.6 Evaluations de performances

La construction de modèles de reconnaissance automatique implique l'évaluation de l'efficacité de ces modèles. Dans le cas de modèles distinguant deux catégories d'entités⁵, cette évaluation passe par l'estimation de quatre mesures : le nombre d'entités correctement reconnues comme appartenant à une catégorie (Vrais Positifs VP), le nombre d'entités correctement reconnues comme n'appartenant pas à une catégorie (Vrais Négatifs VN), le nombre d'entités incorrectement reconnues comme appartenant à une catégorie (Faux Positifs FP), et le nombre d'entités incorrectement reconnues comme n'appartenant pas à une catégorie (Faux Négatifs FN). Ces quatre mesures constituent la Matrice de Confusion du modèle évalué, présentée dans le Tableau 1.

Entité/Classification	Positive	Négative
Positive	VP	FN
Négative	FP	VN

Tableau 1 : Matrice de Confusion pour l'évaluation des modèles de reconnaissance automatique, distinguant les types d'erreurs (VP = vrais positifs, VN = vrais négatifs, FP = faux positifs, FN = faux négatifs)

Afin de pouvoir comparer les performances de deux modèles, plusieurs scores peuvent être calculés à partir de ces quatre mesures.

Taux de Bonnes Reconnaissances

Il s'agit du taux d'entités correctement catégorisées par le modèle évalué : _____

Alternativement, on peut calculer le taux d'erreurs : _____

L'inconvénient du taux de bonnes reconnaissances est qu'il n'apporte aucune information sur les erreurs commises, c'est-à-dire s'il s'agit de faux positifs ou de faux négatifs.

Rappel (ou Sensibilité)

Le rappel évalue dans quelle mesure les entités positives ont été catégorisées comme telles : _____

Précision (ou Spécificité)

La précision évalue dans quelle mesure le modèle classe correctement la catégorie positive : _____

⁵ Ces mesures peuvent s'appliquer à des problèmes de plus de deux catégories, en utilisant une approche « un contre le reste du monde »

État de l'Art

f-mesure

La f-mesure combine les mesures de précision et de rappel d'un modèle en une mesure unique : _____

Notons que dans les équations indiquées, seules les performances du modèle sur la catégorie positive sont évaluées. Afin d'avoir une vision complète des performances du modèle, il convient de calculer ces mesures sur la catégorie négative également.

Courbe ROC

La courbe sensibilité/spécificité (Receiver Operating Characteristic ROC) est une représentation graphique du taux de vrais positifs (sensibilité) sur le taux de faux positifs (1-spécificité), illustrée en Figure 5. Afin de comparer deux modèles, l'aire sous la courbe ROC permet d'évaluer leurs performances.

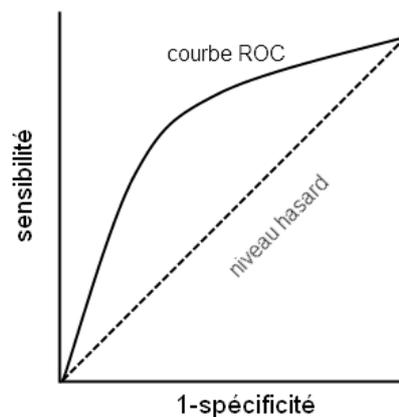


Figure 5 : Courbe sensibilité/spécificité, ou courbe ROC (Receiver Operating Characteristic)

Corrélation de Pearson

La corrélation est une mesure de la relation entre deux variables continues x et y . La corrélation de Pearson évalue la force de la relation linéaire entre x et y :

— où σ_x et σ_y représentent la déviation standard des variables x et y , respectivement, et σ_{xy} représente la covariance entre les deux variables.

Le coefficient de corrélation r varie entre -1 (forte corrélation négative) et 1 (forte corrélation positive), 0 indiquant une corrélation nulle. La corrélation r est accompagnée d'une mesure de valeur- p indiquant la pertinence statistique de r sur l'échantillon de données observé, c'est-à-dire la probabilité d'obtenir r si l'Hypothèse Nulle – aucune corrélation n'existe entre x et y – était vraie.

On pourra étudier la corrélation entre les entités catégorisées comme positives (VP et FP) et les entités réellement positives (VP et FN), et réciproquement pour les entités négatives.

État de l'Art

Challenges pour la reconnaissance d'états affectifs

La communauté de l'Affective computing* regroupe des équipes de recherche travaillant sur une même catégorie de problème (i.e. la reconnaissance automatique des émotions) mais utilisant des outils différents. Les dernières années ont par conséquent vu apparaître des « challenges », des défis proposés à la communauté afin de mesurer les performances de leurs outils sur de nouvelles données.

Chaque année depuis 2009, à l'occasion de la conférence Interspeech Björn Schuller propose à la communauté traitant la parole affective des challenges portant sur la reconnaissance des états du locuteur à partir d'enregistrements audio. Les challenges ont porté sur la reconnaissance de cinq ou deux émotions [270], l'âge, le genre et l'intérêt [271], l'ébriété et la fatigue [272], les traits de personnalité OCEAN* [273], les conflits en groupe, les affect bursts, douze émotions et différentes pathologies autistiques [274], la fatigue cognitive et la fatigue physique [275]. Les performances du modèle de référence et des modèles des participants aux challenges sont comparées en utilisant la valeur moyenne du rappel non pondéré, et certaines années l'aire sous la courbe ROC.

Depuis 2011 à l'occasion de la conférence ACM Multimedia, Michel Valstar et Bjorn Schuller proposent des challenges portant sur la reconnaissance des émotions à partir d'enregistrement audio et vidéo du corpus SEMAINE [201] et plus récemment de la dépression sur le corpus AVID La reconnaissance des émotions porte suivant les années sur tout ou partie des dimensions Excitation, Attente, Contrôle et Valence [276] [277]; depuis 2012 les dimensions sont évaluées sur une échelle continue (en 2011, l'échelle était binaire). La reconnaissance de la dépression, inclue depuis 2013 [311] [312], porte sur l'évaluation du score au Beck Depression Index (BDI) de 0 à 63. Les performances du modèle de référence et des modèles des participants sont comparées en utilisant le coefficient de Corrélacion de Pearson moyen entre prédictions et véritables catégories pour la reconnaissance des émotions, et les erreurs absolue et quadratique moyennes pour la reconnaissance de la dépression.

En 2011, au cours de la conférence IEEE Face and Gesture Recognition, Michel Valstar à proposé le challenge de reconnaissance d'expressions faciales FERA, sur une partie des données du corpus GEMEP [309]. Le challenge consistait en deux tâches : la classification de cinq expressions faciale (colère, peur, joie, soulagement, tristesse), et la reconnaissance de douze des Unités d'Action (Action Units AU) du système FACS* d'Ekman (c.f. 2.2.3). Les performances du modèle de référence et des modèles des participants sont comparées en utilisant le score de f-mesure moyen pour la reconnaissance des douze AU, et le taux de bonne reconnaissance moyen pour les cinq expressions faciales.

A notre connaissance, il n'y a pas eu à ce jour de challenges portant sur la reconnaissance du stress, ou impliquant des modalités autres qu'audio et vidéo.

Synthèse de l'État de l'Art

Nous avons présenté le stress comme une réaction physiologique du corps face à une menace, qu'elle porte sur notre intégrité physique ou sur notre capacité à réaliser nos aspirations. Cette réaction peut être déclenchée par une situation ponctuellement effrayante, ou par une anxiété chronique. Nous avons introduit en particulier le cas de la phobie sociale, trouble anxieux associé à une crainte pathologique des interactions sociales. La pratique favorisée pour la remédiation de ces troubles est la Thérapie Cognitive Comportementale, confrontant progressivement le patient à ses peurs. Nous avons présenté les résultats positifs de l'utilisation d'environnements logiciels pour l'application de ces thérapies.

La perception de la menace provoquant le stress est subjective, influencée par des facteurs à la fois contextuels, socioculturels et personnels. Nous avons décrit les notions *d'appraisal* et de *coping* comme stratégies d'appréhension et de gestion du stress ; les différences interindividuelles de stratégies face à un même évènement stressant impactent l'expérience et l'expression du stress.

Nous avons présenté des outils de mesure de facteurs individuels susceptibles de modifier les réactions face au stress. En particulier nous nous sommes intéressés aux différentes échelles psychologiques de mesure de traits de personnalités, de stratégies de coping, et d'évaluation de la phobie sociale et des troubles anxieux.

Nous avons également présenté des mesures physiologiques, vocales, faciales et posturales porteuses d'informations pour l'identification des états affectifs des humains, en particulier du stress. Nous nous sommes particulièrement intéressés aux mesures obtenues à partir de capteurs peu invasifs.

Nous avons présenté *l'affective computing*, domaine de recherche visant à fournir aux machines la capacité de reconnaître les états affectifs des humains, simuler des émotions et raisonner en intégrant ces informations émotionnelles. Cet apprentissage passe par l'étude de données affectives, actées ou réalistes, annotées perceptivement par des humains. Nous avons décrit la notion de stress dans les expérimentations affectives existantes.

Les techniques d'apprentissage automatique permettent l'exploitation de grandes quantités d'informations pour la constitution de modèles de reconnaissance des états affectifs ; nous avons abordé les techniques les plus courantes. Nous avons également mentionné l'utilisation de stratégies de sélection de paramètres pour la simplification des modèles, et de fusion d'information pour la robustesse des modèles. Nous avons enfin présenté différentes mesures et compétitions visant à évaluer les performances des modèles pour déterminer les configurations les plus efficaces pour la reconnaissance des états affectifs.

Dans la suite de cette étude, nous utilisons ces mesures et techniques pour caractériser et reconnaître le stress d'individus en conditions réalistes dans différents contextes anxiogènes.

Section II Collecte, Annotation et Analyse de Signaux de Stress

Dans cette partie, nous présentons la constitution de corpus multimodaux de stress collectés grâce à différentes expérimentations mettant les sujets dans des conditions réalistes aptes à éliciter du stress. Cette partie se divise en trois chapitres :

Le Chapitre 4 « Constitution de Corpus de Stress » présente les trois expérimentations auxquelles nous avons collaboré au cours de cette étude, conçues pour éliciter du stress auprès d'une population phobique, d'une population non-pathologique, et d'une population multiculturelle.

Le Chapitre 5 « Annotations » présente l'annotation des données collectées, à la fois par les sujets via des questionnaires d'évaluation psychologiques et de retour d'expérience, et par des annotateurs extérieurs. Ce chapitre présente notamment le protocole d'annotation sur lequel ont été formé les annotateurs.

Le Chapitre 6 « Analyse du signal pour l'extraction d'indices du stress » présente les paramètres vocaux, faciaux et posturaux extraits des données collectées et annotées. Ces paramètres sont confrontés aux informations de profil psychologique obtenues par les questionnaires d'évaluation, et observés sur des cas particuliers de stress.

Chapitre 4 Constitution de Corpus de Stress

4.1 Capture de données inédites

Dans cette étude, nous souhaitons utiliser des informations multimodales pour évaluer le stress de différentes populations dans des conditions anxigènes réalistes. N'ayant pas de corpus adéquat pour mener ce travail sur le stress, il était essentiel de collecter des corpus pour étudier des phénomènes de stress. Grâce à différents projets, nous avons pu constituer plusieurs corpus originaux avec différents types de sujets. Tous les corpus ont été enregistrés en suivant le même procédé de capture audio ; pour la vidéo nous avons utilisé plusieurs technologies (webcam, kinect).

Les données audio ont été capturées à l'aide d'un micro-cravate AKG-40, et enregistrées avec le logiciel Audacity à 16 KHz. La segmentation et l'annotation des enregistrements audio ont été réalisées à l'aide de Transcriber [19]. Cet outil permet la visualisation du signal audio, la segmentation des fichiers en sous-parties à partir des bornes indiquées par l'annotateur, l'annotation des multiples locuteurs, et la transcription du discours.

Les données vidéo ont été capturées à l'aide d'une webcam standard et enregistrées avec le logiciel WinAVI à 25 FPS. Cet outil permettant de synchroniser une entrée vidéo et une entrée audio, les données étaient synchronisées dès l'enregistrement. Les fichiers vidéo ont été segmentés à l'aide de Movie Maker, puis annotés à l'aide du logiciel de sous-titrage Subtitle Edit. Cet outil permet la visualisation du signal audio et la lecture image par image du fichier vidéo.

Les données posturales à partir de vidéo 3D ont été capturées à l'aide d'une Kinect 1^{ère} génération, et enregistrées avec le logiciel NiViewer d'OpenNI [221]. Les fichiers vidéo 3D ont été segmentés à l'aide d'une modification du code de NiViewer, développée à cet effet dans le cadre de cette étude. Les données audio et vidéo 3D ont été synchronisées à l'aide d'un clap du pied au sol, perçu dans les deux modalités. La segmentation par tâche et tour de parole a été réalisée sur les données audio.

4.2 E-thérapies : public sociophobe en situation anxiogène simulée

Cette étude a été financée par le projet FEDER E-Thérapies, dont l'un des objectifs est la conception de logiciels à but thérapeutique, à destination d'un public sociophobe. Nous avons présenté la phobie sociale dans le Chapitre 1 de ce document. Cette cible est un public intéressant pour l'étude du stress induit par une pathologie altérant la perception des interactions sociales.

Dans le cadre du projet nous avons collaboré avec Le Centre Emotion, USR 3246 hébergée dans les locaux de l'hôpital de la Pitié-Salpêtrière à Paris. Cette collaboration nous a permis d'interagir avec un groupe de thérapie, dont l'objectif est d'aider les patients sociophobes à dépasser leur phobie par des exercices de Thérapie Cognitive Comportementale (TCC*). Ces exercices consistent en des mises en situation suivant des scénarios anxiogènes avec l'accompagnement du thérapeute, afin de permettre aux participants de reconnaître et d'apprendre à contrôler leurs émotions en développant des stratégies cognitives positives.

Cette collaboration a donné lieu à deux collectes de données auprès des patients du groupe de thérapie, de patients en première consultation, et de participants non pathologiques comme groupe de contrôle. La première collecte utilise une interface graphique comme inducteur de stress ; la seconde collecte s'appuie sur l'interaction avec des humains afin d'induire du stress.

4.2.1 Collecte ETCC-HM : interactions homme-machine

La collecte ETCC-HM (pour E-Thérapie Cognitive Comportementale Humain-Machine) a été initiée afin d'évaluer la validité de l'utilisation d'un logiciel comme outil thérapeutique pour le traitement de la phobie sociale [1]. Cette collecte devait également permettre de capturer des données auprès de sociophobes et de témoins pour évaluer la possibilité de percevoir le stress, particulièrement un stress pathologique, dans des enregistrements audio et vidéo.

Afin d'éliciter des réactions anxieuses chez les participants, nous avons conçu un logiciel composé de deux parties, une interface « client » qui affiche des animations sur un écran, et une interface « serveur » qui permet à un « Magicien d'Oz » (MOz*) [22], un intervenant humain, de contrôler partiellement les animations en indiquant manuellement les émotions exprimées par les sujets. L'interface « serveur » est illustrée en Figure 6.

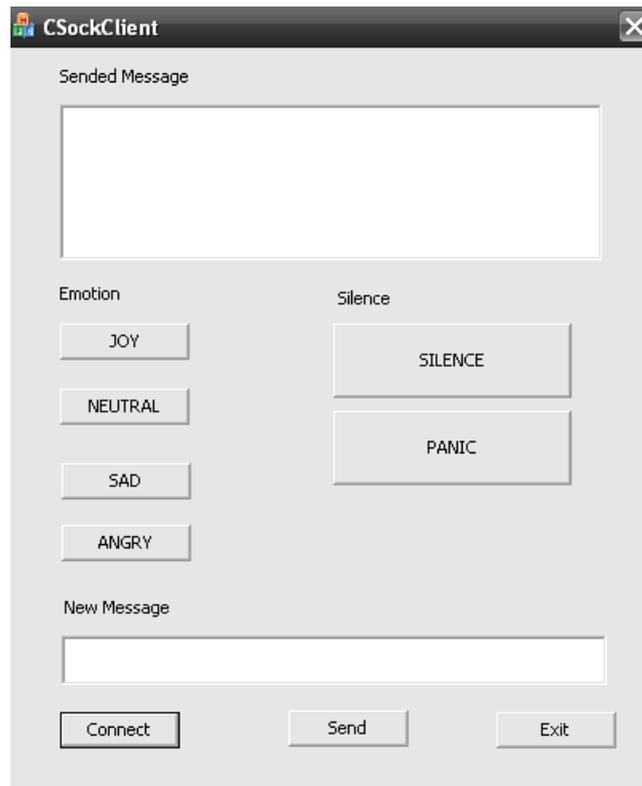


Figure 6 : Interface de contrôle MOz des animations ETCC-HM

Les animations affichées sur l'interface « client » illustrent 3 scénarios élaborés avec les médecins du Centre Emotion, comparables à ceux proposés aux patients lors des sessions de TCC*. L'une des difficultés rencontrées par les sociophobes est l'expression d'émotions, aussi nous avons conçu ces scénarios afin d'éliciter des réactions émotionnelles chez les participants.

Le premier scénario nécessite de parler à un public afin de les convaincre d'adopter un comportement respectueux de l'environnement. L'objectif est d'encourager les participants à adopter une voix posée pendant que des interlocuteurs les observent.

Le deuxième scénario nécessite d'exprimer de la joie afin d'inviter l'élue de son cœur à un rendez-vous romantique. L'objectif est d'encourager les participants à adopter une voix enjouée pour convaincre l'interlocuteur d'accepter l'invitation.

Le troisième scénario nécessite d'exprimer de la colère face à une personne insistante. L'objectif est d'encourager les participants à adopter un ton ferme et négatif pour repousser l'interlocuteur.

Collecte, Annotation et Analyse de Signaux de Stress

Plutôt que d'utiliser des personnages réalistes dans les animations, nous nous sommes inspiré des travaux de Heider et Simmel [124] et avons imaginé des scènes abstraites animées par les réactions des utilisateurs. Les interlocuteurs sont symbolisés par des triangles noirs aux yeux globuleux, le regard fixé sur l'utilisateur symbolisé à l'écran par un rond. La couleur du rond change en fonction de l'expression émotionnelle des participants, et les mouvements des triangles figurent les intentions des interlocuteurs. L'interface « client » est illustrée dans le Tableau 3 pour les 3 scénarios proposés.

Cette interface nous a permis d'enregistrer 27 personnes entre novembre 2012 et mai 2013. Un phobique a demandé à être retiré de l'étude. La population est présentée dans le Tableau 2.

Population	Nombre de sujets	Répartition hommes/femmes	Age moyen
Témoins	9	2/7	34 ans
Phobiques	15	9/8	36 ans
Total	26	11/15	35 ans

Tableau 2 : description des sujets de la collecte ETCC-HM

Les participants, après consentement explicite, ont été enregistrés à l'aide d'un micro-cravate et filmés à l'aide d'une webcam cadrée sur leur visage. Assis face à un écran, ils disposaient de quelques minutes pour tester l'application hors de tout scénario pour se familiariser avec les animations, puis réalisaient les trois scénarios dans l'ordre (le dispositif est illustré en Figure 7).

Ils avaient la possibilité de rejouer les scénarios s'ils n'étaient pas satisfaits de leur performance. La durée maximale de chaque exercice était fixée à 90 secondes, cependant les scénarios 2 et 3 pouvaient se terminer plus rapidement si les participants atteignaient l'objectif (i.e., le triangle a rejoint le rond dans le scénario 2, et le triangle a quitté l'écran dans le scénario 3) ou à l'inverse si l'objectif devenait inatteignable (i.e., le triangle a quitté l'écran dans le scénario 2, et le triangle a rejoint le rond dans le scénario 3).

Suite aux enregistrements, les participants ont rempli trois questionnaires : le Liebowitz Social Anxiety Scale (LSAS*) [183], le questionnaire de personnalité OCEAN* [145], et un questionnaire de retour d'expérience indiquant notamment la difficulté perçue de l'exercice⁶.

⁶ Les questionnaires sont disponibles en annexe

Collecte, Annotation et Analyse de Signaux de Stress

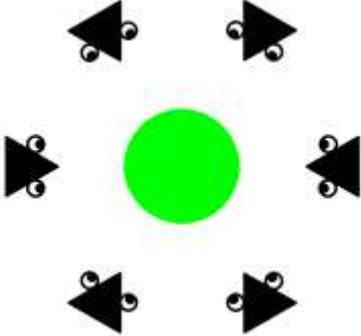
Scénario	Animation	Explications
<p>Scénario 1 : parler d'une voix neutre mais convaincante</p>		<p>Essayez de convaincre le public d'adopter un comportement respectueux de l'environnement, par exemple le tri sélectif. Ce peut être vos voisins, vos amis...Vous devez parler de façon posée (colorer le rond en vert) pendant 1min30 pour attirer autour de vous un maximum de triangles.</p>
<p>Scénario 2 : parler avec une voix positive et enjouée</p>		<p>Invitez pour la première fois une personne qui vous plaît beaucoup. Ce peut être à diner, au cinéma...vous pouvez lui donner un prénom pour vous aider. Vous devez parler de façon positive (colorer le rond en jaune) pour la convaincre de vous accompagner, vous avez 1min30 pour que le triangle se rapproche du rond et le touche.</p>
<p>Scénario 3 : parler avec une voix colérique, voire agressive</p>		<p>Repoussez les avances d'un vendeur insistant, dans un magasin ou au téléphone. Vous devez vous mettre en colère (colorer le rond en rouge) pour le repousser, car il est impossible de le raisonner, vous avez 1min30 pour faire sortir le triangle de l'écran. Attention, il peut revenir vers vous s'il n'est pas complètement sorti !</p>

Tableau 3 : Illustration et présentation des 3 scénarios du protocole expérimental ETCC-HM pour la collecte de données en interaction homme-machine

En plus du MOz*, un médecin était présent durant toute la session pour assister les participants, et discuter avec eux de l'expérimentation pour les aider de la même façon qu'en thérapie.

Collecte, Annotation et Analyse de Signaux de Stress

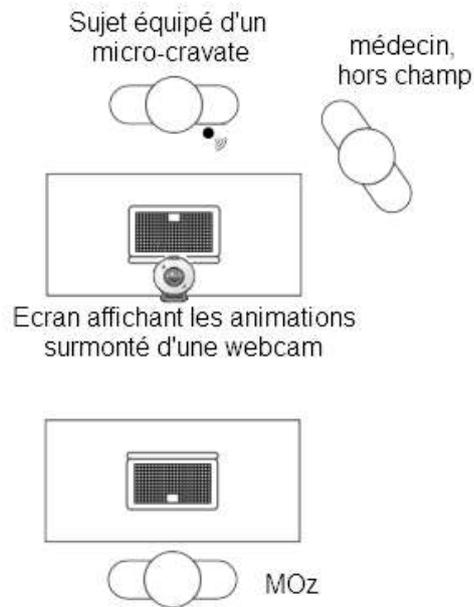


Figure 7 : Disposition de l'expérience de collecte ETCC-HM

Suite à la collecte, les données audio et vidéo ont été segmentées suivant les trois scénarios. Le Tableau 4 récapitule les données collectées à l'aide de l'expérimentation.

	Audio	Vidéo	Questionnaires
Capteur	Micro cravate, enregistrement à 16KHz (Audacity)	Webcam, enregistrement 320x240 25FPS (WinAVI)	OCEAN, LSAS, retour d'expérience
Données	2h04min	1h41min ⁷	

Tableau 4 : Récapitulatif des données du corpus ETCC-HM

⁷ Certains enregistrements vidéo ne sont pas exploitables, d'où la différence de quantité par rapport aux données audio.

4.2.2 Collecte ETCC-HH : interaction entres humains

La collecte ETCC-HH (pour E-Thérapie Cognitive Comportementale Humain-humain) a été initiée pour évaluer l'influence d'humains jouant le rôle d'un jury sur l'expression du stress chez des participants en situation d'évaluation.

L'expérimentation élaborée est une prise de parole en public, face à un jury de deux personnes sur un sujet imposé. L'exercice consiste en quatre tâches : la lecture d'un texte au sujet portant à débat extrait de Wikipedia sur le mariage homosexuel⁸, l'expression d'un avis sur le sujet du texte, la réponse à une question absurde (« Que pensez-vous des mœurs de reproduction des rhinocéros roses ? »), et une description positive de soi.

42 personnes ont participé à cette collecte ; parmi elles 10 phobiques avaient précédemment participé à la collecte ETCC-HM. Un sujet phobique (ayant également participé à la première collecte) a demandé à être retiré de l'étude, et un témoin a été exclu a posteriori car présentant un score LSAS* élevé. La population est présentée dans le Tableau 5.

Population	Nombre de sujets	Répartition hommes/femmes	Age moyen
Témoins	18	10/8	34 ans
Phobiques	22	12/10	36 ans
Total	40	22/19	35 ans

Tableau 5 : Description des sujets de la collecte ETCC-HH

Les participants, après consentement explicite, ont été enregistrés à l'aide d'un micro-cravate. Debout sur une estrade et face à un ou deux expérimentateurs, ils réalisaient les 4 tâches expliquées au fur et à mesure par les expérimentateurs (le dispositif expérimental est illustré en Figure 8).

Pour chaque tâche hormis la lecture (où le temps était libre), les participants disposaient de 90 secondes. Si leur réponse était plus courte, les expérimentateurs restaient silencieux jusqu'à ce que le temps soit écoulé.

⁸ Texte extrait avant la législation de 2013 sur l'union des couples de même sexe.

Collecte, Annotation et Analyse de Signaux de Stress

Suite aux enregistrements, les participants remplissaient trois questionnaires : le Liebowitz Social Anxiety Scale (LSAS*) [183], le Questionnaire d'Éreutophobie de la Pitié-Salpêtrière (QES) [232], l'échelle d'affirmation de soi de Rathus [245], le Hospital Anxiety and Depression scale (HAD*) [335], le questionnaire de personnalité OCEAN* [145], et un questionnaire de retour d'expérience indiquant l'anxiété et la peur du jugement ressentis⁹.

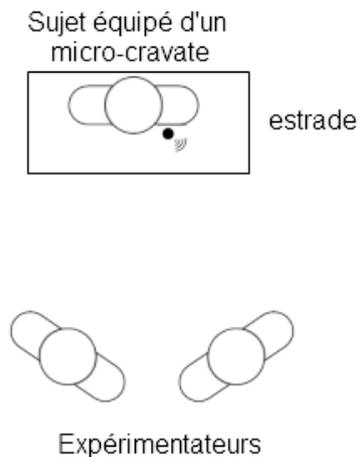


Figure 8 : Disposition de l'expérience de collecte ETCC-HH

Suite à la collecte, les données audio ont été segmentées par tâche. Le Tableau 6 récapitule les données collectées à l'aide de l'expérimentation.

Mesure	Audio	Questionnaires
Capteurs	Micro cravate, enregistrement à 16KHz (Audacity)	OCEAN, LSAS, QES, Rathus, HAD, retour d'expérience
Données	4h25min	

Tableau 6 : Récapitulatif des données du corpus ETCC-HH

⁹ Les questionnaires sont disponibles en annexe

4.3 Compare : public non-pathologique en situation stressante ponctuelle

Au cours de cette étude, nous avons participé au projet Compare¹⁰ dont l'objectif est l'étude des liens entre émotions, cognition et comportement. Nous avons collaboré avec le CIAMS, Unité de Recherche de l'UFR STAPS de l'Université Paris-Sud, sur la tâche de Prise de Parole en Public (Public Speaking Task, PST*).

Cette tâche, variante du Trier Social Stress Test (TSST*) [154], est utilisée pour éliciter du stress chez les participants en les plaçant dans une situation d'évaluation sociale. Le TSST* consiste en une période de 10 minutes d'anticipation (30 minutes en cas de prise de sang), une période de 10 minutes de préparation après instructions, une présentation orale de 5 minutes suivie d'une session de questions de 5 minutes, et une tâche de calcul mental devant un jury composé de deux personnes.

Dans le projet Compare, après une période d'anticipation de 30 minutes, les participants lisent à voix haute un texte court, reçoivent les instructions et débutent leur présentation orale de 5 minutes sans temps de préparation, puis répondent aux questions de deux juges.

Entre 2012 et 2013, 4 campagnes de collecte ont permis d'enregistrer 53 participants à cette expérimentation. Les populations sont présentées dans le Tableau 7.

Campagne	Nombre de sujets	Répartition hommes/femmes	Age moyen
1	10	5/5	28 ans
2	19	7/12	27 ans
3	10	4/6	26 ans
4	14	8/6	26 ans
Total	53	24/29	27 ans

¹⁰ Etude des liens entre COgnition, Motivation et Personnalité pour l'Adaptation et la Régulation émotionnelle par Simulation virtuelle Empathique

Collecte, Annotation et Analyse de Signaux de Stress

Tableau 7 : Description des sujets des collecte Compare

Une semaine avant la collecte, les participants ont rempli les questionnaires de personnalité suivants : évaluation des traits de personnalité (BFI* [145]), trait anxieux (STAI* [290]), alexithymie (TAS*-20 [12] [13]), narcissisme vulnérable (HSNS [127]), stratégies de coping (COPE* [46]), et évaluation des menaces et des défis [265]. Chaque participant s'est vu attribuer un horaire de passage individuel, et a été invité à se présenter 30 minutes avant afin de procéder aux mesures physiologiques basales (i.e. décrivant les valeurs moyennes pour l'individu au repos).

Le jour de la collecte, le participant était invité à fournir des échantillons de salive et à procéder aux mesures physiologiques suivantes, pour établir ses relevés basaux : conductance et hémodynamique de la peau, température, rythme cardiaque et variation du rythme cardiaque. 10 minutes avant de procéder à l'expérimentation, le participant recevait les instructions de la tâche : « Vous postulez pour un entretien d'embauche ». Il remplissait également les questionnaires d'évaluation d'état anxieux (STAI* [290]), et d'état affectif (PANAS [323]).

A l'heure convenue, le participant était accueilli par le modérateur dans la salle d'expérimentation et équipé d'un micro-cravate pour l'enregistrement de la voix, puis dirigé sur une estrade dans laquelle une plateforme de force était intégrée, et filmé par 2 caméras (une vue de plein pied et un plan rapproché sur le visage, de face) et une Kinect (vue en pied de face). Le jury de deux personnes (un homme et une femme) se trouvait face au participant derrière une table ; les expérimentateurs étaient présents dans la salle, hors du champ des caméras (le dispositif expérimental est illustré en Figure 9).



Figure 9 : Collecte de données Compare, deuxième campagne (2012)

Collecte, Annotation et Analyse de Signaux de Stress

Après lecture d'un texte à voix haute, le sujet débutait sa présentation de 5 minutes. S'il terminait avant le temps imparti, le jury l'invitait à continuer ; si le temps était écoulé, le jury l'interrompait. Suite à la présentation, les juges posaient chacun trois questions suivant un script préétabli¹¹. Le premier juge jouait un rôle négatif et critique, le second un rôle positif et encourageant. Les juges homme et femme ont alterné les rôles afin d'obtenir une répartition homogène en fonction du sexe des candidats.

Suite aux enregistrements, les participants remplissaient 3 questionnaires : évaluation de l'état anxieux (STAI* [290]) et d'état affectif (PANAS [323]), et évaluation des stratégies de coping (COPE* [46]). Ils étaient ensuite invités à discuter avec un expérimentateur neutre – non présent durant l'expérimentation – dans une pièce adjacente pour remplir un questionnaire de retour d'expérience, et à fournir des échantillons de salive post-expérimentation.

Le corpus collecté comprend de très nombreuses modalités, propriété des différents partenaires impliqués dans le projet [II]. Cette étude porte sur les données audio et la capture 3D de la Kinect. Les captures Kinect de la première campagne n'étant pas disponibles, les résultats présentés par la suite pour cette modalité portent sur les campagnes 2, 3 et 4.

Suite à la collecte, les données collectées ont été segmentées par tâche, selon les intitulés suivants : « lecture », « présentation », « questions négatives », et « questions positives ». Le Tableau 8 récapitule les données collectées à l'aide de l'expérimentation.

Mesure	Audio	Visuel 3D	Questionnaires
Capteurs	Micro cravate, enregistrement à 16KHz (Audacity)	Kinect, enregistrement 640x480 25FPS (OpenNI)	BFI, STAI, TAS-20, HSNS, Brief COPE, PANAS, auto evaluation
Données CP1	2h10min	-	
Données CP2	4h10min	4h00min ¹²	
Données CP3	2h53min	2h53min	
Données CP4	3h17min	3h17min	
Total	12h30min	10h10min	

Tableau 8 : Récapitulatif des données des corpus CP1-CP2-CP3-CP4

¹¹ Les questions des juges sont disponibles en annexe de ce document

¹² L'un des enregistrements vidéo n'est pas exploitable, d'où la différence de quantité par rapport aux données audio.

4.4 Nao Fait de l'Humour: public non-pathologique en interaction sociale humoristique

Afin d'étudier l'expression du stress dans un contexte social et l'utilisation de l'humour comme stratégie de coping, nous avons conçu une expérimentation d'interactions à but humoristique et divertissant avec des machines. Cette expérimentation a été déployée lors de la conférence Interspeech 2013 à Lyon. Cet évènement international réunissant chaque année des chercheurs du monde entier travaillant sur la parole, il s'agissait d'une excellente opportunité pour étudier l'influence des normes socioculturelles sur l'expérience et l'expression du stress [IV].

Nous avons souhaité comparer l'impact de deux types de tâches pour éliciter les rires et sourires des participants à cette étude : des tâches dites « passives », où le sujet est simplement exposé à un stimulus divertissant, et des tâches dites « actives », où le sujet doit effectuer une action. Nous avons également souhaité comparer l'impact de deux machines : un écran d'ordinateur, et le robot humanoïde Nao.

Avec les tâches passives, nous souhaitons éliciter un amusement réflexe, suite à l'observation d'un stimulus : nous avons utilisé des vidéos d'enfant riant et d'animaux dans des situations cocasses diffusées sur écran, et des blagues racontés par le robot Nao.

Avec les tâches actives, nous souhaitons éliciter un amusement fier ou embarrassé, selon la réussite de la tâche : nous avons conçu un jeu de fourchelangue, à lire sur écran ou à répéter avec le robot Nao dans différentes langues¹³ pour augmenter la difficulté. Les fourchelangues sont des phrases existant dans la plupart des langues, spécifiquement conçues pour être difficile à prononcer en alternant des phonèmes semblables. Un exemple classique en français est « Les chaussettes de l'archiduchesse sont-elles sèches ou archi-sèches ».

Une application a été développée afin de permettre à un MOz* de partiellement contrôler les différentes tâches. Une interface unifiée permet de manipuler à la fois les informations affichées sur l'écran face au sujet (i.e. vidéos et fourchelangues), et le robot Nao (i.e. énoncés, langue, mouvements et rires), et conserve un historique des interactions. Les stimuli sont sélectionnés aléatoirement par le système, le MOz* en déclenche simplement l'exécution. Il a également la possibilité d'utiliser le moteur de synthèse vocale du robot pour faire répondre Nao aux questions des participants.

¹³ Le jeu propose des phrases en français, anglais, espagnol, italien, allemand, et mandarin.

Collecte, Annotation et Analyse de Signaux de Stress

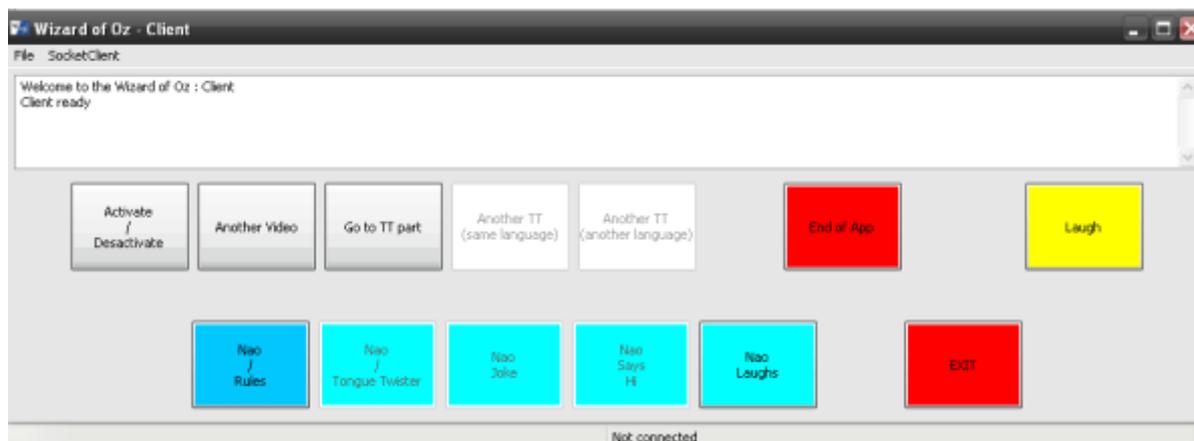


Figure 10 : Interface de contrôle MOZ de l'application IS

45 personnes ont participé à l'expérimentation durant les 3 jours de la conférence Interspeech. La population est présentée dans le Tableau 9.

Population	Nombre de sujets	Répartition hommes/femmes
Europe	27	18/9
Asie	12	11/1
Amérique	6	2/4
Total	45	31/14

Tableau 9 : Description des sujets de la collecte IS

Chaque enregistrement s'est déroulé en deux parties : une première face à un écran d'ordinateur, puis une seconde face au robot Nao. Les participants, après consentement explicite, ont été filmés à l'aide d'une webcam pour la première partie, et enregistrés à l'aide d'un micro-cravate pour la totalité de l'expérience. Les sessions duraient une dizaine de minutes.

Face à l'écran, après avoir fourni des renseignements sur leur pays d'origine et leur langue, les participants visionnaient deux vidéos tirées au hasard parmi une dizaine pour la tâche passive. Ils lisaient ensuite à haute voix 3 à 5 fourchelanges tirés au hasard, dans leur langue mais également dans une langue aléatoire, pour la tâche active.

Les participants étaient ensuite invités à parler avec le robot Nao, qui leur proposait de répéter après lui des fourchelanges en anglais ou en français. Après une ou deux répétitions, le robot leur racontait quelques blagues courtes pour la tâche passive. Durant toute l'interaction le robot pouvait rire avec des gestes, mais également faire des remarques moqueuses (e.g. « Même ma petite sœur y arriverait ! »). Le dispositif expérimental est illustré en Figure 11.



Figure 11 : Collecte de données lors d'Interspeech 2013, à Lyon (photo A. Crenn)

Suite à l'expérience, les participants ont rempli un questionnaire de personnalité [145] pour les dimensions Ouverture, Extraversion et Névrotisme [265], une échelle de mesure du Sens de l'Humour [200], et un questionnaire de retour d'expérience indiquant notamment leur tâche favorite.¹⁴

Suite à la collecte, les données ont été segmentées par tâche, selon les intitulés suivants : « vidéos », « fourchelangues lus », « fourchelangues répétés », et « blagues ». Les deux premières étiquettes correspondent à la partie sur écran et comportent des données audio et vidéo. Les deux dernières étiquettes correspondent à l'interaction avec le robot et comportent uniquement des données audio. Le Tableau 10 récapitule les données collectées à l'aide de l'expérimentation

Mesure	Audio	Video	Questionnaires
Capteurs	Micro cravate, enregistrement à 16KHz (Audacity)	Webcam, enregistrement 320x240 25FPS (WinAVI)	Sub-OCEAN, SHS, retour d'expérience
Données	3h56min	2h01min	

Tableau 10 : Récapitulatif des données des corpus IS-Comp et IS-Nao

¹⁴ Les questionnaires sont disponibles en annexe.

4.5 Récapitulatif des corpus

Corpus	Population	Protocole de collecte	Modalités enregistrées	Échelles psychologiques utilisées	Durée de données capturée
ETCC-HM	15 sociophobes et 9 témoins 11 hommes / 15 femmes âge moyen 35 ans	Exercices de TCC avec une machine : mise en situation anxiogènes simulée par des animations abstraites	Audio (micro cravate) Vidéo du visage (webcam)	OCEAN, LSAS, retour d'expérience	Audio 2h04min Vidéo 1h41min
ETCC-HH	22 sociophobes et 18 témoins 22 hommes / 19 femmes âge moyen 35 ans	Exercices de prise de parole face à un public humain	Audio (micro cravate)	OCEAN, LSAS, QES, Rathus, HAD, retour d'expérience	Audio 4h25min
CP1	10 étudiants 5 hommes / 5 femmes âge moyen 28 ans	Entretien d'embauche simulé avec 2 juges humains	Audio (micro cravate)	BFI, STAI, TAS-20, HSNS, Brief COPE, PANAS, auto evaluation	Audio 2h10min
CP2	19 étudiants 7 hommes / 12 femmes âge moyen 27 ans	Entretien d'embauche simulé avec 2 juges humains	Audio (micro cravate) Posture (kinect)	BFI, STAI, TAS-20, HSNS, Brief COPE, PANAS, auto evaluation	Audio 4h10min Posture 4h00min
CP3	10 étudiants 4 hommes / 6 femmes âge moyen 26 ans	Entretien d'embauche simulé avec 2 juges humains	Audio (micro cravate) Posture (kinect)	BFI, STAI, TAS-20, HSNS, Brief COPE, PANAS, auto evaluation	Audio 2h53min Posture 2h53min
CP4	14 étudiants 8 hommes / 6 femmes âge moyen 26 ans	Entretien d'embauche simulé avec 2 juges humains	Audio (micro cravate) Posture (kinect)	BFI, STAI, TAS-20, HSNS, Brief COPE, PANAS, auto evaluation	Audio 3h17min Posture 3h17min
ISComp	45 chercheurs de 24 pays 31 hommes / 14 femmes	Interaction humoristique face à un ordinateur : vidéos humoristiques et fourchelangues	Audio (micro cravate) Vidéo du visage (webcam)	Sub-OCEAN, SHS, retour d'expérience	Audio 2h01min Video 2h01min
ISNao	45 chercheurs de 24 pays 31 hommes / 14 femmes	Interaction humoristique avec le robot Nao : blagues et fourchelangues	Audio (micro cravate)	Sub-OCEAN, SHS, retour d'expérience	Audio 1h55min

Tableau 11 : Récapitulatif des corpus collectés au cours de l'étude. Les corpus encadrés en gras sont utilisés dans la suite de l'étude.

Chapitre 5 Annotations

5.1 Considérations générales

Afin de pouvoir utiliser les données collectées pour entraîner un système de reconnaissance automatique, il est nécessaire de les caractériser en les annotant selon les critères que l'on souhaite détecter automatiquement par la suite. Cette étape d'annotation des données est le résultat d'une réflexion autour de la question "Qu'est-ce que le stress, et comment le reconnaît-on?". L'aboutissement de cette réflexion est la conception d'un protocole d'annotation du stress suivant des critères perceptibles dans les signaux capturés au cours de nos expérimentations.

Dans le Chapitre 1 de ce manuscrit, nous avons détaillé les symptômes du stress chez l'homme. Pour un observateur externe les effets à la fois physiques et mentaux du stress ne sont pas forcément visibles, mais il peut observer leur impact sur le comportement des sujets lors de l'exécution d'une tâche. Par exemple de nombreux sujets ont expliqué avoir une sensation de bouche sèche lorsqu'ils sont stressés, mais il n'est pas envisageable d'obtenir cette information en temps réel. En revanche il est possible d'entendre le sujet déglutir de façon audible. De la même façon il n'est pas possible de connaître les processus cognitifs se déroulant dans l'esprit d'une personne stressée, mais il est possible d'observer un débit de parole accéléré ou au contraire interrompu par des pauses et des disfluences dans le discours.

Pour les raisons présentées en 3.2 l'annotation de données pour l'identification de l'expression de processus internes tels que l'expérience du stress est une tâche complexe, qui nécessite de considérer un certain nombre de questions.

Tout d'abord, qui est le plus à même d'identifier le stress d'une personne ? La réponse intuitive est l'individu lui-même puisqu'il est le seul à avoir accès à ses processus cognitifs. L'auto-annotation, par le biais de questionnaires de retour d'expérience ou par l'auto-évaluation des enregistrements réalisés pendant l'expérimentation, va fournir une information au plus proche de l'expérience de l'individu. Cette information reste subjective car influencée par le système de valeurs de l'individu qui impacte l'expression de son ressenti : une personne ayant honte de son stress hésitera à l'avouer. Mais si nous considérons que l'annotation des données a pour but d'entraîner un système à reconnaître le stress comme un observateur externe, il devient intéressant d'utiliser la perception d'observateurs humains comme guide pour l'élaboration du système.

Dans cette étude nous avons utilisé les deux formes d'annotations (qualifiées par la suite d'auto-annotation et d'annotation externe), dont nous comparons les résultats dans la suite de ce chapitre.

Outre l'identification de l'annotateur se pose la question du traitement des données à annoter : comment segmenter les enregistrements collectés ? Quelle forme d'enregistrement (audio, vidéo, audio et vidéo) fournir aux annotateurs ? Faut-il annoter les données étayées d'informations contextuelles ou hors contexte ?

Les corpus présentés dans le Chapitre 4 ont tous été constitués de la même manière : des sujets volontaires participent à une expérimentation durant laquelle ils sont enregistrés de façon continue. Chacun des corpus présentés correspond à un protocole expérimental différent, et bien qu'ils suivent le même protocole, tous les sujets d'une même expérimentation ne réagissent pas de la même façon, et des facteurs individuels viennent nuancer leurs réponses (c.f. 1.5). Ces différences peuvent être quantifiées par des « méta-annotations » à l'échelle d'un corpus ou d'un sujet. Le Tableau 12 récapitule les méta-annotations disponibles pour chacun des corpus présentés au Chapitre 4.

Corpus/ Meta	Sexe	Traits de Personnalité	Phobie	Stratégie de Gestion du Stress	Influence Culturelle	Sens de l'Humour
ETCC-HH	✓	✗	✓	✗	-	✗
ETCC-HM	✓	✓	✓	✗	-	✗
CP2	✓	✓	-	✓	-	✗
IS-Comp	✓	✓	-	✗	✓	✓
IS-Nao	✓	✓	-	✗	✓	✓

Tableau 12 : Meta informations disponibles par corpus
(✓ disponible ; ✗ non renseigné ; - non applicable)

Dans cette étude, les protocoles expérimentaux comportent tous plusieurs tâches conçues pour éliciter des réactions différentes chez les sujets. S'il est possible d'annoter le stress à l'échelle de la session expérimentale d'un sujet, cette information représentera le stress « moyen » au cours de l'expérimentation, masquant les variations dans l'intensité du stress suivant la tâche réalisée. Nous avons donc opté pour une segmentation par tâche, les enregistrements résultant durant entre 30 secondes et 5 minutes, selon les protocoles et selon les sujets.

Collecte, Annotation et Analyse de Signaux de Stress

Il est possible de segmenter de façon plus fine les enregistrements, tenant compte du contenu du discours des sujets pour identifier des « phases » : il peut s'agir du tour de parole dans un dialogue, ou bien de phrases séparées par des pauses significatives. Nous avons utilisé cette segmentation fine pour le corpus CP2, dont les tâches longues (par rapport aux autres protocoles) et l'aspect interactionnel (lors des phases de questions des deux juges) ont généré des enregistrements au cours desquels le degré d'intensité du stress varie. Le protocole de collecte Comparsé permettant une certaine variation dans le contenu du discours des participants, la segmentation par tâche a également été utilisée car elle présente l'avantage d'identifier un stressor commun à tous les sujets.

Dans le Chapitre 2 (2.2.5) nous avons vu que l'expression d'états affectifs peut prendre des formes très brèves, les affects bursts. Nous avons segmenté ces micro-expressions sur les corpus ETCC-HM, IS-Comp et IS-Nao pour évaluer leur pertinence dans la mesure du stress et de la gêne exprimée par un individu en condition de stress social.

La Figure 12 récapitule l'encapsulation des différents niveaux de segmentation.

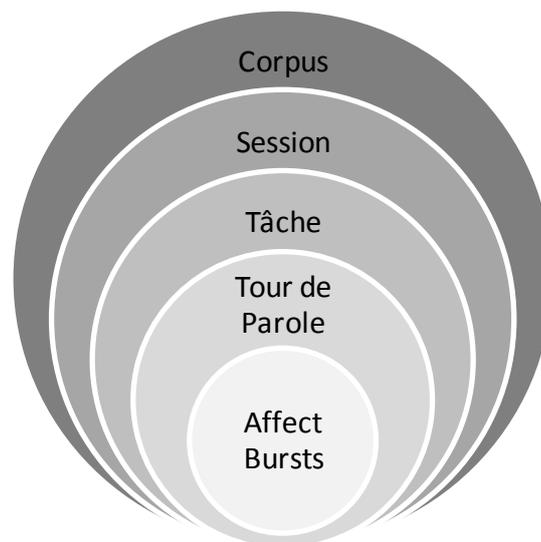


Figure 12 : Granularité de l'annotation

Les corpus présentés au Chapitre 4 comportent tous une modalité audio ; certains corpus comportent une modalité vidéo ; l'un des corpus comporte des enregistrements 3D. Si l'expression affective est multimodale, son annotation sera inévitablement influencée par les canaux d'information accessibles aux annotateurs. Dans cette étude, l'annotation par tâche sur ETCC-HM, CP2, CP3, CP4, et IS-Comp a été faite à partir d'enregistrements vidéo sonorisés ; l'annotation par tour de parole sur CP2 a été faite dans le cadre d'une étude linguistique à partir des enregistrements audio uniquement ; l'annotation d'ETCC-HH a été faite à partir d'enregistrements audio puisque ce corpus n'inclut pas de capture vidéo ; l'annotation des affects bursts sur IS-Comp a été faite sur les enregistrements audio et vidéo séparément (les enregistrements audio ont servi à l'annotation des rires, les enregistrements vidéo non sonorisés à l'annotation des sourires).

Des informations contextuelles sur les circonstances de l'enregistrement vont également influencer la perception d'un annotateur sur le stress : un observateur externe annotant l'intégralité d'un corpus aura une vision du stress relatif à l'expérimentation, tandis qu'un sujet auto-évaluant son stress aura une vision relative à son expérience personnelle du stress en général. Un annotateur ayant observé l'ensemble des enregistrements d'un sujet de façon chronologique aura une conception de l'évolution du stress du sujet durant l'expérimentation, tandis qu'un observateur traitant les enregistrements de façon aléatoire aura une conception plus générique du stress d'un enregistrement lambda. La connaissance d'informations individuelles influençant l'expérience du stress – comme la phobie sociale dans les corpus ETCC-HM et ETCC-HH – va également altérer la perception d'un observateur : le stress d'un individu identifié comme « prompt à stresser » pourra être sur-évalué. Dans cette étude nous avons opté pour une annotation chronologique des enregistrements, et pour chaque corpus les annotateurs externes ont traité l'intégralité des enregistrements.

Enfin, quel système d'annotation appliquer ? Dans l'annotation des états affectifs, deux procédés se distinguent : l'annotation catégorielle, et l'annotation dimensionnelle.

L'annotation catégorielle consiste à séparer les données en différentes classes d'émotions. Outre le fait que les différentes études sur ce point ne s'accordent pas sur le nombre ni l'intitulé des catégories à utiliser (de 4 à 11, incluant généralement les 6 émotions basiques d'Ekman [89], et des catégories plus « sociales » comme l'embarras), nous avons mentionné la complexité des expressions dans les données naturelles, qui comportent généralement un mélange d'émotions plutôt qu'une expression distincte.

L'annotation dimensionnelle est une réponse à ce problème, elle consiste à positionner les données sur un continuum axé autour d'une ou plusieurs dimensions abstraites. Les dimensions les plus fréquemment utilisées sont la « valence » (qualifiant une émotion de « positive » ou de « négative ») et l'« activation » (qualifiant le degré d'intensité d'une émotion). Il est également possible de projeter les classes catégorielles dans un espace dimensionnel [251], la limite de cette approche étant la distinguabilité des catégories considérées suivant les dimensions considérées.

Concernant l'annotation du stress dans cette étude, nous avons opté pour deux approches : une annotation discrète se rapportant à l'approche dimensionnelle indiquant un degré d'intensité du stress sur une échelle de 1 à 10 (1 pour « pas du tout stressé », 10 pour « extrêmement stressé ») pour le segment considéré, et une annotation binaire (stressé/non-stressé) se rapportant à l'approche catégorielle indiquant si le segment considéré présente ou non du stress. Une troisième approche intermédiaire, annotant trois degrés d'intensité du stress (« faible », « moyen », ou « élevé ») a également été utilisée sur le corpus IS-Comp.

5.2 Auto annotation du stress

Le stress est l'expression d'un état interne de l'individu en réaction à un évènement qu'il perçoit comme une menace. Cette perception est influencée par de nombreux facteurs propres à chacun (l'expérience de situations similaires, un évènement traumatisant, la perception de soi) mais également propres au contexte (quels sont les enjeux d'un échec dans cette situation, quels sont les paramètres contrôlables et imposés dans cette situation...).

Afin de tenir compte de l'aspect subjectif du stress, nous avons demandé aux participants d'évaluer leur propre stress dans nos différentes expérimentations, sur une échelle de Likert de 1 (« pas du tout stressé ») à 10 (« extrêmement stressé »). Cette auto-évaluation s'est faite par tâche, à posteriori : les sujets ont donc implicitement évalué leur stress pour chaque tâche de façon relative à l'ensemble de l'expérience.

5.2.1 Sur E-Thérapies

Les corpus ETCC-HM et ETCC-HH (présentés en 0) collectés dans le cadre du projet E-Thérapies portent sur un public sociophobe et une population témoin dans des tâches courtes (entre 30 et 120s) de prise de parole, similaires à celles employées en TCC*. Les participants ont évalué leur stress de 1 à 10 pour chacune des tâches, et également indiqué quelle tâche leur a semblé la plus facile et quelle tâche leur a semblé la plus difficile.

Le graphique présenté en Figure 13 illustre le stress auto-annoté pour chaque tâche pour chaque population, ainsi que le stress moyen pour l'ensemble de l'expérimentation ETCC-HM.

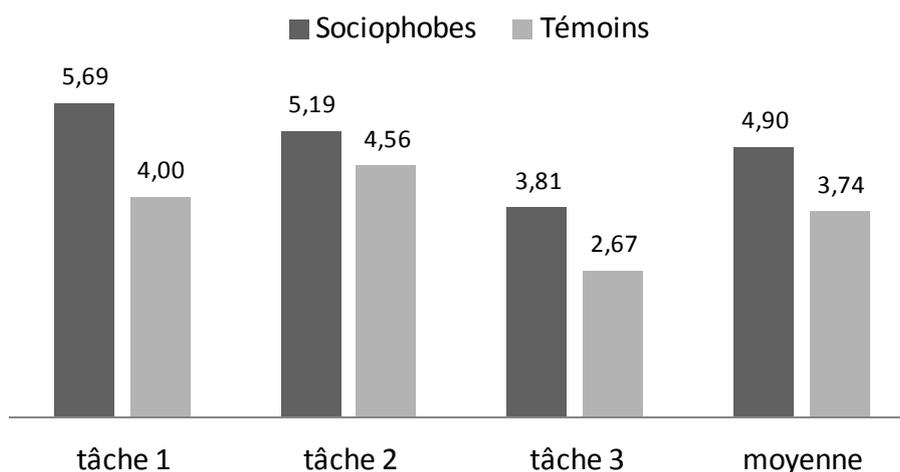


Figure 13 : Stress auto-annoté par tâche par population sur ETCC-HM

Collecte, Annotation et Analyse de Signaux de Stress

Sur l'ensemble des tâches, les sociophobes ont annoté un stress ressenti plus important que les témoins. Cette annotation concorde avec leur pathologie, qui amplifie leur perception du jugement d'autrui et déforme leur perception de leur performance. Le Tableau 13 compare la difficulté perçue des tâches et le stress ressenti pour ces tâches, pour les deux populations.

Tâche/Public	Sociophobes	Témoins
Tâche la plus facile	3	3
Tâche la moins stressante	3 (3.81)	3 (2.67)
Tâche la plus difficile	1	2
Tâche la plus stressante	1 (5.69)	2 (4.56)

Tableau 13 : Stress/Difficulté auto-annoté par tâche par population sur ETCC-HM (stress moyen auto-annoté entres parenthèses)

On constate une cohérence entre la tâche perçue comme la plus facile et celle annoté comme la moins stressante, ainsi qu'entre la tâche perçue comme la plus difficile et annotée comme la plus stressante, et ce sur les deux populations.

Si les deux populations perçoivent la 3^{ème} tâche (« *se mettre en colère pour chasser un vendeur obséquieux* ») comme la plus facile et la moins stressante, la tâche la plus difficile et la plus stressante diffère. Les sociophobes ont considéré la 1^{ère} tâche (« *convaincre un public d'adopter un comportement respectueux de l'environnement* ») comme la plus difficile et stressante, tandis que les témoins ont considéré la 2^{ème} tâche (« *inviter quelqu'un qui vous plait à un rendez-vous amoureux* ») comme telle. Le graphique présenté en Figure 14 illustre le stress auto-annoté pour chaque tâche pour chaque population, ainsi que le stress moyen pour l'ensemble de l'expérimentation ETCC-HH.

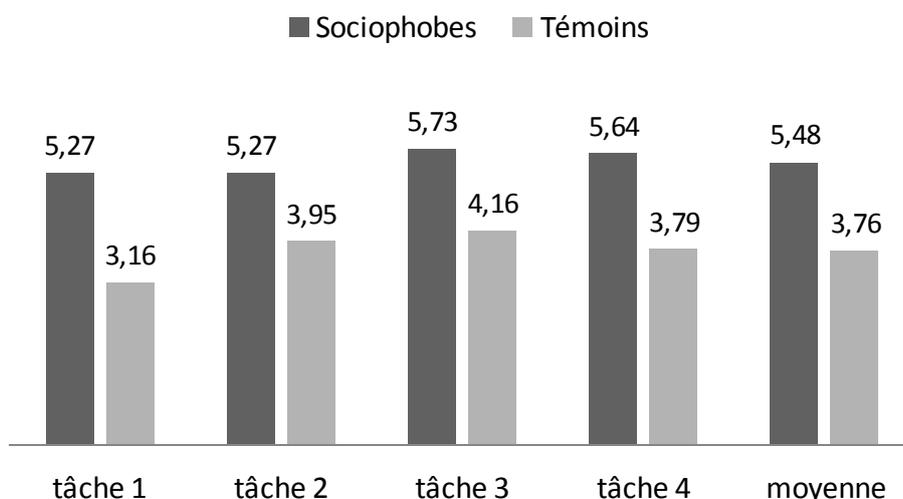


Figure 14 : Stress auto-annoté par tâche par population sur ETCC-HH

Comme pour le corpus ETCC-HM, sur l'ensemble des tâches, les sociophobes ont annoté un stress ressenti plus important que les témoins.

Le Tableau 14 compare la difficulté perçue des tâches et le stress ressenti pour ces tâches, pour les deux populations.

Tâche/Public	Sociophobes	Témoins
Tâche la plus facile	1	1
Tâche la moins stressante	1 (5.27)	1 (3.16)
Tâche la plus difficile	4	4
Tâche la plus stressante	3 (5.72)	3 (4.16)

Tableau 14 : Stress/Difficulté auto-annoté par tâche par population sur ETCC-HH (stress moyen auto-annoté entres parenthèses)

Si la 1^{ère} tâche (lecture de texte) perçue comme la plus facile par les sociophobes est effectivement la moins stressante, on constate une incohérence pour la tâche la plus stressante (la 3^{ème}, « *commenter les mœurs de reproduction des rhinocéros roses* ») qui n'est pas la plus difficile (la 4^{ème}, « *parler de soi de façon positive* »).

5.2.2 Sur Comparse

Le corpus CP2 (présenté en 4.3) a été collecté dans le cadre du projet Comparse auprès d'un public non-pathologique d'étudiants du supérieur (âge moyen 26 ans, niveau d'étude de Licence à Doctorat) sur une simulation d'entretien d'embauche impliquant une phase de prise de parole en public, et une phase de questions avec deux juges.

Les participants ont évalué leur stress à la suite de l'expérimentation sur une échelle de Likert de 1 (pas du tout stressé) à 5 (très stressé). Cette évaluation concernait l'ensemble de la session d'entretien, le détail par phase n'est pas disponible. Le graphique présenté en Figure 15 montre la répartition des auto-annotations par échelon.

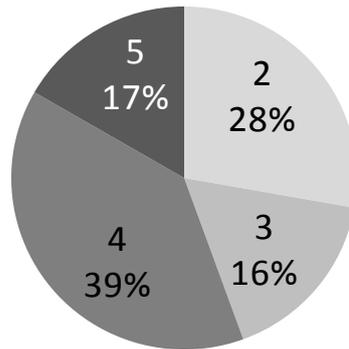


Figure 15 : Répartition des auto-annotations du stress (de 1 à 5) sur CP2

Aucun participant n'a indiqué ne pas avoir été stressé par l'expérimentation (pas d'annotation à 1) et une majorité (56%) a indiqué avoir été stressé ou très stressé (annotations 4 et 5).

5.2.3 Sur Nao Fait de l'Humour

Les corpus IS-Comp et IS-Nao (présentés en 0) ont été collectés lors d'Interspeech 2013 durant la démonstration « Nao Fait de l'Humour », auprès d'un public de chercheurs internationaux de tous horizons. Les participants ont été enregistrés en interaction avec un ordinateur et avec le robot Nao, dans des tâches visant à les faire rire, de façon amusée ou gênée.

Suite à l'expérimentation, les participants ont évalué leur malaise sur une échelle de 1 (très à l'aise) à 5 (très mal à l'aise). Cette évaluation concernait l'ensemble de l'expérimentation et le détail par tâche n'est pas disponible. Le graphique présenté en Figure 16 montre la répartition des auto-annotations de la gêne par échelon.

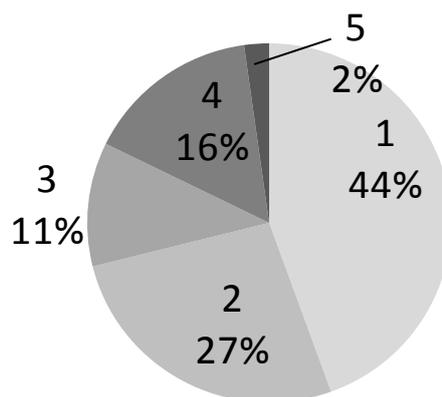


Figure 16 : Répartition des auto-annotations de gêne (de 1 à 5) sur IS

Seule une minorité de participants (18%) a indiqué s'être senti gênée ou très gênée par l'expérimentation (échelons 4 et 5).

5.3 Annotation du stress par un observateur externe

L'auto-annotation du stress est subjective. Réalisée suite à l'expérimentation, elle est impactée par les sensations immédiates induites par celle-ci. Elle est également influencée par les caractéristiques personnelles des individus. Nous avons par exemple pu constater que les sociophobes s'évaluaient comme plus stressés que les témoins face à des tâches identiques.

Une annotation des enregistrements obtenus dans nos expérimentations par des observateurs externes sera sensiblement plus proche des observations que pourra effectuer un système artificiel, puisque que dans les deux cas les processus cognitifs du sujet stressé ne seront pas accessibles et seules les manifestations externes du stress seront disponibles.

5.3.1 Protocole d'annotation du stress

Afin de guider les annotateurs externes dans leur évaluation du stress, nous avons proposé une liste de « signes » indicatifs du stress ressentis par les sujets, à rechercher dans les enregistrements à annoter. Cette liste s'appuie sur les indices du stress proposés en psychologie [209][65]. Les signes sont listés dans le Tableau 15.

Symptômes physiques	Symptômes "mentaux"
<ul style="list-style-type: none"> • profondes inspirations • déglutition • souffles • raclements de gorge • tapement de pied • tremblements 	<ul style="list-style-type: none"> • pauses marquées dans le discours (le sujet ne sais pas quoi dire) • débit de parole très rapide, pour achever la tâche stressante le plus rapidement possible • utilisation marquée de "fillers" (<i>euuh, hmmm, donc, bon, enfin, voilà</i>) • abandon de la tâche explicité vocalement ("<i>je ne sais pas quoi dire</i>", "<i>je n'y arrive pas</i>", "<i>on peut recommencer?</i>", "<i>c'est difficile!</i>") • rires de gêne • pleurs

Tableau 15 : Signes indicatifs du stress, traduits d'après [65]

Plus on trouve de marqueurs dans un enregistrement et plus le sujet semble perceptivement stressé.

Comme pour l'auto-annotation, nous avons souhaité une annotation par tâche sur une échelle de Likert de 1 (pas du tout stressé) à 10 (très stressé). Afin de normaliser les annotations, nous avons proposé les directives présentées dans le Tableau 16 pour l'attribution des notes.

Note	Description
1-4	le sujet est à l'aise réalise la tâche sans difficulté exprime de la satisfaction en fin de tâche (<i>"oui!", rire amusé</i>)
5-7	le sujet est mal à l'aise atteint l'objectif de la tâche avec plus ou moins de difficultés, exprime du soulagement en fin de tâche (<i>"ouf!", rire soulagé</i>)
8-10	le sujet est extrêmement mal à l'aise ne parvient pas à réaliser la tâche ou difficilement (<i>arrête de parler, pleur, refus d'exécuter la tâche</i>) exprime de la honte en fin de tâche

Tableau 16 : Critères d'annotation du stress

Les données présentées dans cette étude ont été annoté par deux personnes formées à l'évaluation du stress à partir de ce protocole. Elles ont également été familiarisées avec les protocoles expérimentaux ayant permis la collecte des données, et vu des cas extrême de stress et de non stress sur ces données pour leur permettre d'appréhender la variété des réactions à annoter.

Ce procédé de formation et d'annotation est long et couteux. Le « crowdsourcing », ou production participative, gagne un intérêt croissant pour l'annotation de grandes quantités de données par un grand nombre de personnes. Plutôt que de faire annoter l'intégralité d'un corpus par quelques experts, l'annotation d'une multitude de petites portions du corpus est déléguée à un grand nombre de personnes non-expertes rémunérées. Le Mechanical Turk d'Amazon est une plateforme populaire pour ce type d'annotations [36].

Les études révèlent qu'avec un nombre suffisant d'annotateurs, les résultats d'annotations consensuelles sont comparables à celles produites par des experts [289], y compris pour le domaine de l'affective computing [213] [198] [303]. La diffusion en ligne des données, en particulier des enregistrements vidéo, implique cependant un consentement spécifique des sujets d'expérimentation, et nous n'avons pas opté pour ce procédé.

5.3.2 Accord inter-annotation

Etant donnée la nature subjective des données annotées, nous avons souhaité estimer l'accord entre les deux experts qui ont fourni l'annotation externe sur les différents corpus. Nous avons choisis d'utiliser le kappa de Cohen [64] $K = \frac{P(o) - P(a)}{1 - P(a)}$ où $P(o)$ est la proportion d'accord observé entre les deux annotateurs, et $P(a)$ est la probabilité d'accord aléatoire. $P(o)$ correspond à la somme des accords p entre les deux annotateurs pour chaque graduation de l'échelle utilisée ; $P(a)$ correspond à la somme du produit des accords p entre les deux annotateurs pour chaque graduation de l'échelle utilisée : $P(o) = \sum_{i=1}^{10} \sum_{j=1}^{10} p_{ij}$ et $P(a) = \sum_{i=1}^{10} \sum_{j=1}^{10} p_{ij} p_{ij}$ où i est l'avis du 1^{er} annotateur et j l'avis du 2^{ème}.

L'échelle que nous utilisons pour évaluer le stress est une graduation progressive et linéaire: un stress de niveau 4 est plus proche d'un stress de niveau 5 que d'un stress de niveau 8. Il est nécessaire de pondérer le désaccord observé entre les annotateurs pour tenir compte de son importance. Les accords p sont donc pondérés d'un poids $w = 1 - \frac{|a-b|}{n-1}$ où n est le nombre de graduations (sur notre échelle, 10), a est la note attribuée par le 1^{er} annotateur et b est la note attribuée par le second annotateur

Avec cette pondération, $P(o) = \sum_{i=1}^{10} \sum_{j=1}^{10} w_{ij} p_{ij}$ et $P(a) = \sum_{i=1}^{10} \sum_{j=1}^{10} w_{ij} p_{ij} p_{ij}$

Dans cette étude, nous nous basons sur l'échelle d'évaluation d'accord suivant K observé proposée par Landis et Koch [168]; l'échelle est présentée en Figure 17.

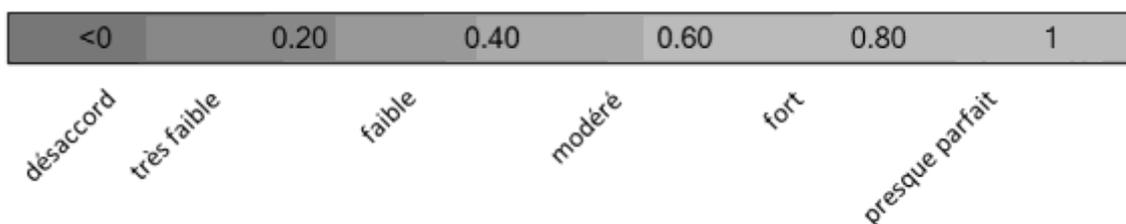


Figure 17 : Accord inter-annotateurs suivant le Kappa observé proposé par Landis et Koch [168]

Mathématiquement, le kappa peut prendre des valeurs allant de -1 à 1. En termes d'interprétations, un kappa inférieur à 0 signifie que les annotateurs ont un accord inférieur à celui qu'ils obtiendraient s'ils effectuaient une annotation au hasard. En application, un kappa négatif est rare [285].

Le kappa et ses seuils d'interprétation sont discutés dans la littérature [39] [60]. L'inconvénient généralement souligné est qu'un kappa faible n'est pas forcément significatif d'un accord systématiquement faible entre les deux observateurs mais peut également être dû à la distribution inégale des données entre les catégories annotées [101]. Cette mesure reste largement utilisée pour l'évaluation de l'accord inter-annotateur.

Collecte, Annotation et Analyse de Signaux de Stress

A des fins de comparaisons, nous présentons également le taux d'accord brut entre les deux annotateurs, c'est-à-dire le nombre de fichiers auxquels une même note a été attribuée sur le nombre total de fichiers annotés.

Sur ETCC-HM

Le corpus ETCC-HM a été annoté par deux observateurs, par tâche, sur une échelle de 1 à 10. En utilisant les notes discrètes de 1 à 10 comme entrées pour calculer $P(o)$, le kappa obtenu entre les deux annotateurs est de 0,41 sur le corpus ETCC-HM, soit un accord « modéré ». Pour comparaison, le taux d'accord brut entre les deux annotateurs est de 35%

Une échelle comprenant de multiples graduations augmente la probabilité de désaccord. Pour augmenter le taux d'accord entre les annotateurs, nous avons réduit l'échelle à 10 graduations en une échelle à trois niveaux (stress « faible », « moyen » et « élevé ») et en une échelle binaire à deux niveaux (« pas de stress », « stress »).

Graduation 10 niveaux	Graduation 3 niveaux	Graduation binaire
1	faible	pas de stress
2		
3		
4	moyen	stress
5		
6		
7		
8	fort	
9		
10		

Tableau 17 : Régression d'échelle

Suite à cette régression d'échelle et en appliquant le même type de pondération, le kappa entre les deux annotateurs est de 0.49 pour l'échelle à trois niveaux, et de 0.60 pour l'échelle binaire sur le corpus ETCC-HM. Si l'accord reste « modéré », il est effectivement plus important lorsque l'échelle d'annotation considérée comprend peu d'étiquettes. Le taux d'accord brut pour l'annotation binaire est de 75%, beaucoup plus favorable que celui sur l'annotation à dix niveaux.

Sur ETCC-HH

Le corpus ETCC-HH a été annoté par deux observateurs, par tâche, sur une échelle de 1 à 10. En utilisant les notes discrètes de 1 à 10 comme entrées pour calculer $P(o)$, le kappa obtenu entre les deux annotateurs est de 0,50 sur le corpus ETCC-HH, soit un accord « modéré ». Pour comparaison, le taux d'accord brut est de 38%. En appliquant la même régression que sur les notes du corpus ETCC-HM, le kappa entre les deux annotateurs est de 0,88 pour l'échelle à trois niveaux, et de 0,94 pour l'échelle binaire, soit des accords presque parfaits. Le taux d'accord brut pour l'annotation binaire est de 81%. L'accord entre les 2 annotateurs est bien plus fort en considérant des échelles avec peu de graduations.

Sur CP2, CP3 et CP4

Les corpus CP2, CP3 et CP4 ont été doublement annotés à l'échelle de la tâche, suivant une échelle de 1 à 10. Comme pour les corpus issus de l'expérimentation E-Thérapies, nous avons utilisé ces notes de 1 à 10, puis les notes régressées sur les échelles trois niveaux et binaires, pour évaluer l'accord inter-annotateur. Les scores sont récapitulés dans le Tableau 18.

mesure	Kappa 10-niveaux	Accord Brut 10-niveaux	Kappa 3-niveaux	Kappa binaire	Accord Brut binaire
CP2	0,40	31%	0,34	0,49	69%
CP3	0,24	13%	0,26	0,28	60%
CP4	0,09	7%	0,13	0,23	45%
Total	0,26	19%	0,23	0,37	59%

Tableau 18 : Kappa de Cohen et taux d'accord brut pour les corpus Comparse

Suivant la graduation proposée dans la Figure 17, l'accord entre les deux annotateurs est globalement faible. Seule l'annotation binaire du corpus CP2 obtient un accord modéré. Comme pour les corpus issus de l'expérimentation E-Thérapies, la régression des annotations vers des espaces de faibles dimensions améliore les accords inter-annotateurs.

Au vu du faible accord inter-annotateur obtenu sur les corpus CP3 et CP4, nous avons décidé de ne pas considérer ces corpus dans la suite de cette étude. L'évaluation des données par des annotateurs supplémentaires devrait permettre la stabilisation des annotations pour de futurs travaux.

Collecte, Annotation et Analyse de Signaux de Stress

Le corpus CP2 a également été annoté de façon très fine par tour de parole dans le cadre d'un stage dont le sujet d'étude était les marqueurs linguistiques du stress et des émotions [2]. Les segments obtenus durent en moyenne 10 secondes, les données ont été annotées en contexte par un premier annotateur suivant une échelle binaire, puis contre-annotées sans contexte par un second annotateur suivant cette même échelle. Le Kappa obtenu est de 0,87, un très bon taux d'accord. L'accord brut est également très élevé, à 94%.

Deux annotations sont disponibles sur le corpus CP2, à des granularités différentes et selon des échelles différentes : l'une est réalisée par tour de parole, l'autre est réalisée par tâche suivant une échelle à dix niveaux. De plus, l'annotation par tour de parole a été réalisée à partir des enregistrements audio uniquement, tandis que l'annotation par tâche a été réalisée à partir des enregistrements vidéo de face, et tient donc compte d'indices visuels supplémentaires.

Afin de comparer les deux annotations, nous avons ramené l'annotation par tour de parole à une annotation par tâche, en tenant compte du nombre de segments, de leur durée moyenne, et de leur durée relative par rapport à la durée de la tâche. Aucune de ces mesures ne corrèle, même de façon faible, avec les annotations par tâche établies à partir des enregistrements vidéo. La régression à des échelles trois niveaux ou binaires n'améliore pas les corrélations. Une explication est la différence d'information portée par les modalités audio et vidéo ; les silences prolongés en particulier sont un marqueur de stress proposé dans notre guide d'annotation en 5.3.1 qui n'apparaîtra pas dans les annotations audio, puisqu'elles sont réalisées uniquement sur les portions où les sujets parlent. Cette différence est explorée en 5.3.4

Sur IS-Comp

Le corpus IS-Comp a été évalué par deux annotateurs pour estimer la gêne des participants, sur une échelle à trois niveaux : « peu stressé », « moyennement stressé », « très stressé ». Le kappa obtenu entre les deux annotateurs est de 0,59, soit un accord « modéré ». Pour comparaison, le taux d'accord brut est de 74%.

Le corpus IS-Comp a été annoté selon un second protocole consistant à repérer des affect bursts sous la forme de rires et de sourires. Nous avons comparé les deux formes d'annotations afin d'identifier des possibles corrélations entre les affect bursts produits par les participants, et la gêne annotée par un observateur externe. Cette comparaison se justifie par le fait que les marqueurs affectifs comme les rires embarrassés sont un des indices proposés dans le guide d'annotation du stress proposé en 5.3.1, et réutilisé pour annoter la gêne en 5.3.5.

Collecte, Annotation et Analyse de Signaux de Stress

Afin de comparer les deux annotations, nous avons ramené l'annotation des affect bursts dans les enregistrements à une annotation par tâche, en tenant compte du nombre d'évènements, et de leur durée relative par rapport à la durée de la tâche. Aucune de ces mesures ne corrèle, même de façon faible, avec les annotations à la tâche. Une explication est que le stress ne s'exprime pas uniquement à travers les affects bursts observés, des indices plus bas niveau comme des variations dans la façon de parler ou des micro-expressions faciales sont également porteuses d'informations importantes.

5.3.3 Sur E-Thérapies

Le corpus ETCC-HM a été annoté par deux experts à partir des enregistrements audio et vidéo. L'interface graphique utilisée pour la collecte des données leur a été présentée, afin qu'ils se familiarisent avec les différents scénarios joués par les sujets. Chaque annotateur a écouté l'intégralité des enregistrements chronologiquement, c'est-à-dire en écoutant les tâches dans l'ordre pour chaque sujet. Ils n'avaient pas d'indication quand à la population à laquelle appartenaient les sujets (sociophobes ou témoins) afin de limiter le biais que présente la phobie sociale sur l'expression du stress. Ils ont annoté chaque enregistrement d'une note de 1 à 10 telle que définie dans le Tableau 16.

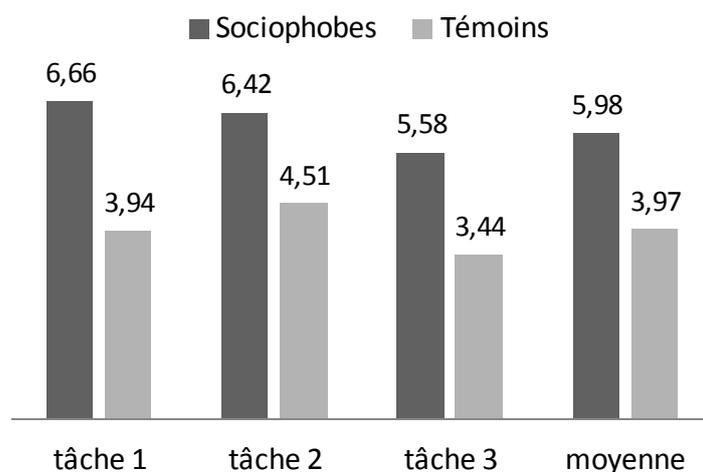


Figure 18 : Stress annoté par 2 experts par tâche par population sur ETCC-HM

Le graphique présenté en Figure 18 illustre la note moyenne attribuée par les experts sur les enregistrements ETCC-HM. Comme pour les auto-annotations présentées en 5.2.1, on constate un stress plus élevé par tâche et en moyenne sur la population phobique. Cette mesure, réalisée en aveugle quant à la population à laquelle appartenaient les sujets, indique que le stress plus prononcé rapporté par les témoins est effectivement perçu par un observateur externe.

Collecte, Annotation et Analyse de Signaux de Stress

Le taux de stress perçu induit par les différentes tâches est partiellement en accord avec les auto-annotations : la 3^{ème} tâche serait la moins stressante pour les deux populations, en revanche la 2^{ème} tâche serait la plus stressante pour les sociophobes, et la 1^{ère} plus stressante pour les témoins.

Le corpus ETCC-HH a été annoté par deux experts, à partir d'enregistrements audio uniquement, puisqu'il n'y a pas d'enregistrement vidéo pour ce corpus. Comme pour le corpus ETCC-HM, les données ont été annotées chronologiquement, et en aveugle quant à la population d'appartenance des sujets.

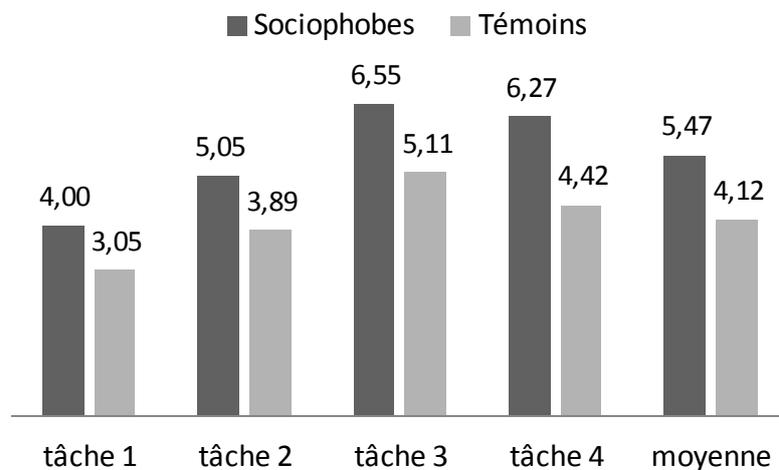


Figure 19 : Stress annoté par 2 experts par tâche par population sur ETCC-HH

Le graphique présenté en Figure 19 illustre la note moyenne attribuée par les experts sur les enregistrements ETCC-HM. Comme pour les auto-annotations présentées en 5.2.1, on constate un stress plus élevé par tâche et en moyenne sur la population phobique. Cette mesure, réalisée en aveugle quant à la population à laquelle appartenaient les sujets, indique que le stress plus prononcé rapporté par les témoins est effectivement perçu par un observateur externe. Les deux populations se comportent de façon similaire sur les différentes tâches, la 3^{ème} (« *commenter les mœurs de reproduction des rhinocéros roses* ») étant la plus stressante et la 1^{ère} (« *lire un texte à haute voix* ») étant la moins stressante.

5.3.4 Sur Compare

Par tâche

Les corpus CP2, CP3 et CP4 ont été annotés par deux experts à l'échelle de la tâche : « exposé libre », « questions du juge négatif », « questions du juge positif ». L'annotation a été faite à partir des enregistrements audio et vidéo de l'expérience capturés par la caméra située face au sujet (c.f. 4.3 pour le détail du dispositif expérimental) et les différentes tâches ont été évaluées en ordre chronologique. Cette annotation a été réalisée sur une échelle discrète de 1 à 10 comme pour les corpus E-Thérapies.

Les notes moyennes attribuées par tâche et par annotateur sont présentées dans les Figure 20, Figure 21 et Figure 22 pour les corpus CP2, CP3 et CP4, respectivement.

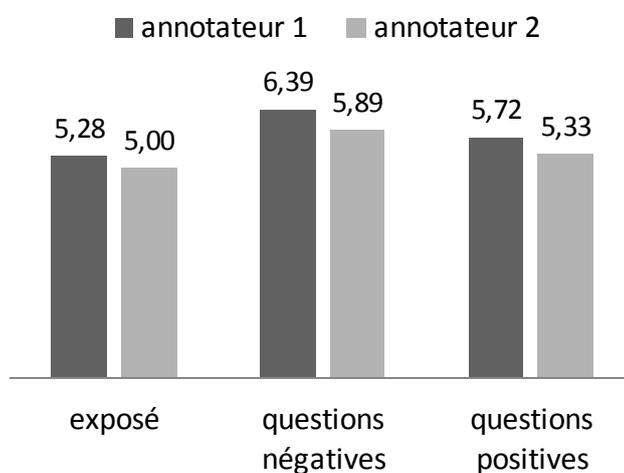


Figure 20 : Stress annoté par deux observateurs par tâche sur CP2

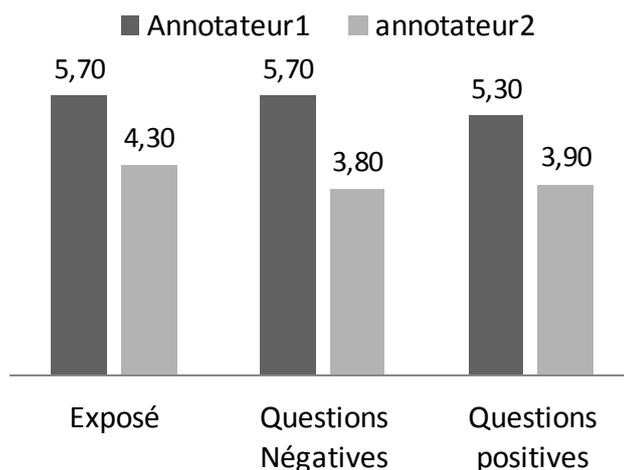


Figure 21 : Stress annoté par deux observateurs par tâche sur CP3

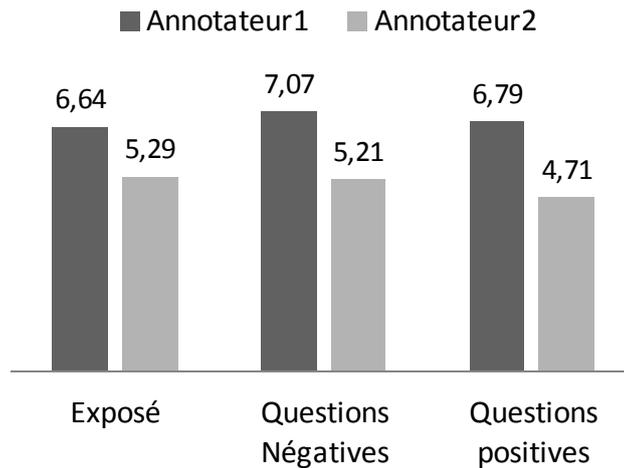


Figure 22 : Stress annoté par deux observateurs par tâche sur CP4

Le désaccord entre les annotateurs sur les corpus CP3 et CP4 est explicite dans les Figure 21 et Figure 22. Cependant, sur tous les corpus les différentes tâches de l'expérimentation ont induit des niveaux de stress similaires : l'écart-type entre les moyennes pour chaque tâche est inférieur à 0,5 pour des notes de 1 à 10.

Par tour de parole

Le corpus CP2 a également été segmenté de façon fine par tours de parole, qui ont ensuite été annotés de façon binaire en « stress » et « non-stress ». Nous avons utilisé ces annotations pour mesurer le stress (respectivement, le non-stress) produit par les différentes tâches de l'expérimentation Comparese.

Les Figure 23 et Figure 24 présentent la répartition du nombre relatif de segments « stress » et « non-stress » par tâche, et leur durée moyenne relative, respectivement : il s'agit du nombre de segments d'une catégorie par rapport au nombre total de segments pour chaque tâche, et la durée moyenne des segments par rapport à la durée totale de la tâche.

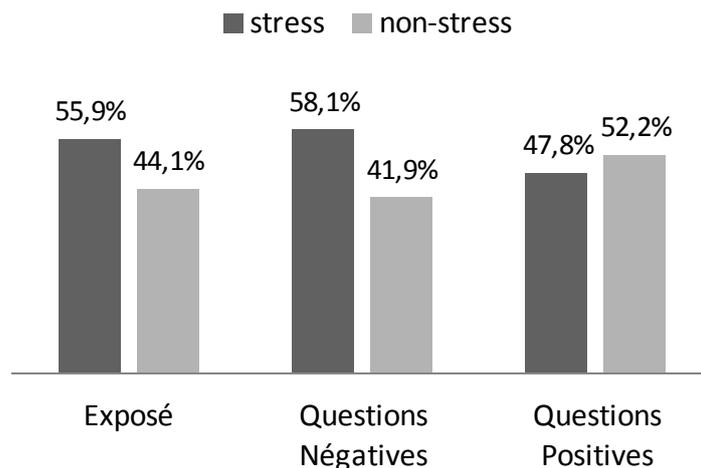


Figure 23 : Proportion de segments stress/non-stress pour chaque tâche sur CP2

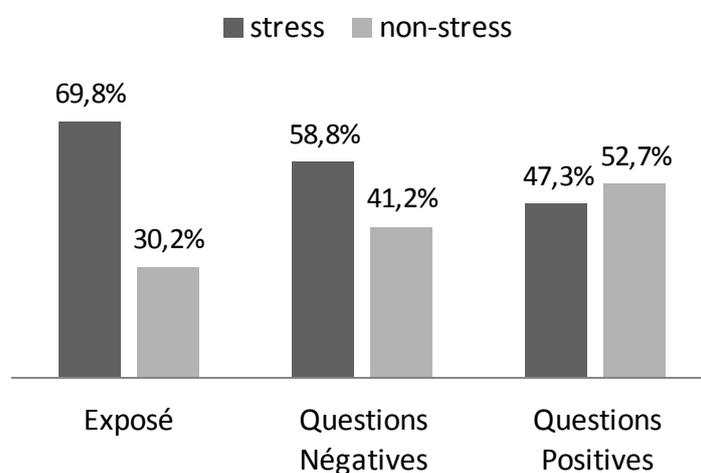


Figure 24 : Durée relative des segments stress/non-stress pour chaque tâche sur CP2

Les résultats indiquent plus de stress que de non-stress pour les tâches « exposé » et « questions négatives », quelle que soit la mesure considérée. Cette tendance s'inverse pour la tâche « question positive ». Ces chiffres sont cohérents avec le protocole expérimental, qui vise à stresser le sujet par la prise de parole libre, à le déstabiliser par des questions critiques, et à l'encourager par des questions valorisantes.

5.3.5 Sur Nao Fait de l'Humour

Gêne

Le corpus IS-Comp a été annoté par deux experts pour quantifier la gêne exprimée par les participants lors de la réalisation des deux tâches de cette partie : le visionnage de vidéos amusantes, et la lecture à voix haute de fourchelangues dans plusieurs langues. L'interface graphique utilisée pour la collecte des données leur a été présentée, afin qu'ils se familiarisent avec les différentes tâches réalisées par les sujets.

Collecte, Annotation et Analyse de Signaux de Stress

Nous qualifions ici de « gêne » le stress social induit par l'expérimentation durant laquelle les participants sont dans un lieu public et soumis au regard des passants (c.f. 0 pour le détail du dispositif expérimental). Les tâches ont été évaluées à partir des enregistrements audio et vidéo. Pour cette annotation, une échelle à trois niveaux a été utilisée, afin de qualifier la gêne exprimée par tâche de « faible », « moyenne » ou « forte ».

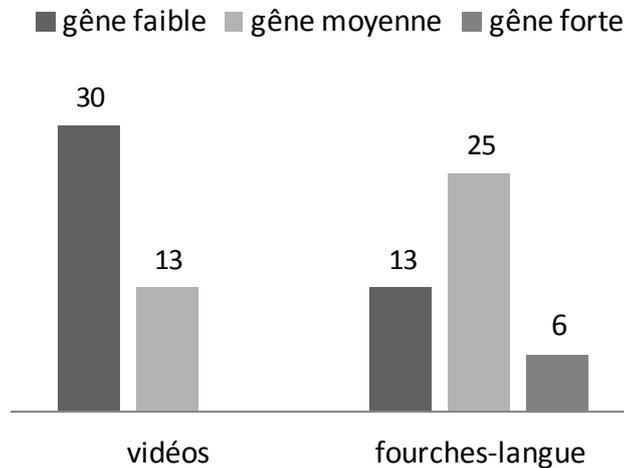


Figure 25 : Répartition des annotations de gêne par tâche sur IS-Comp

La répartition des annotations présentée en Figure 25 montre que comme attendu dans le protocole, la tâche de lecture des fourchelanges a généré une gêne plus importante que le visionnage de vidéos.

Le graphique présenté en Figure 26 distingue la note de gêne moyenne par tâche suivant l'influence culturelle des sujets. Les notes moyennes ont été calculées en substituant une valeur de 1 pour « peu gêné », 2 pour « moyennement gêné » et 3 pour « très gêné ».

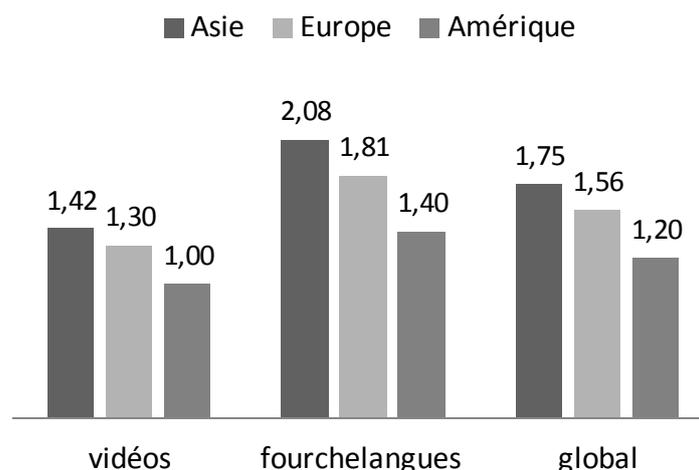


Figure 26 : Gêne moyenne par population par tâche sur IS-Comp

On constate que les différentes populations n'expriment pas la même intensité de gêne : les sujets d'influence asiatique ont globalement été annotés comme plus gênés, et les sujets d'influence américaine comme moins gênés. Cette différence reste valable en observant la gêne par tâche. La tendance constatée sur la Figure 25 indiquant la lecture des fourchelangues comme plus gênante que le visionnage des vidéos reste valable lorsqu'on distingue les résultats par population.

Affect Bursts

Les corpus IS-Comp et IS-Nao ont été annotés afin de repérer des affect bursts dans les enregistrements. Dans le contexte de cette expérimentation, nous disposons d'enregistrements audio pour le corpus IS-Nao, et audio-visuel pour le corpus IS-Comp. Nous avons cherché des Affect bursts* susceptibles de révéler la gêne ressentie par les participants : l'annotation a porté sur les rires dans les enregistrements audio, et les sourires dans les enregistrements vidéo. En particulier, l'annotation distingue les rires amusés des rires gênés (cette évaluation est faite en contexte en écoutant l'intégralité de l'enregistrement), et les sourires bouche ouverte des sourires bouche fermée.

La Figure 27 illustre le nombre de rires sur l'intégralité de l'expérimentation (corpus IS-Comp et IS-Nao cumulés).

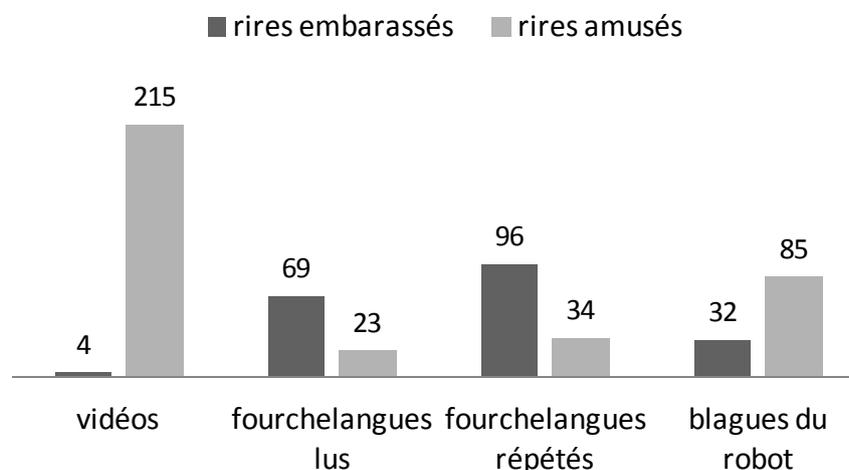


Figure 27 : Nombre de rires par tâche sur IS-Comp et IS-Nao

La distinction « rire amusé » / « rire embarrassé » donne une indication de la gêne ressentie par les participants sur les différentes tâches de l'expérimentation. Comme attendu dans le protocole, les tâches « actives » de fourchelangues ont mis les participants plus mal à l'aise que les tâches « passives » de visionnage de vidéo ou d'écoute de blagues.

La Figure 28 illustre le nombre de sourires sur la partie filmée de l'expérimentation (corpus IS-Comp).

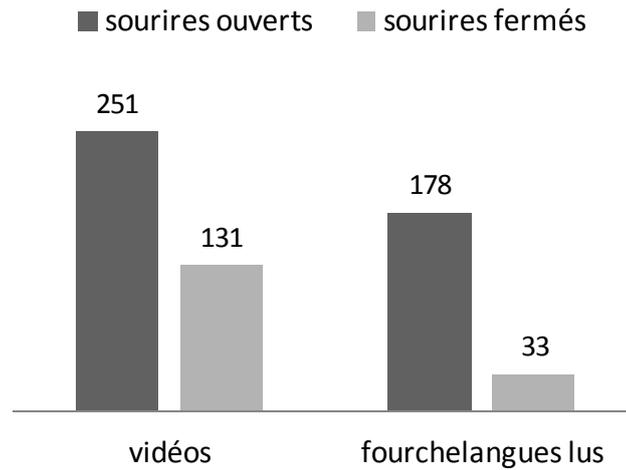


Figure 28 : Nombre de sourires par tâche sur IS-Comp

Nous avons distingué les sourires « ouverts » des sourires « fermés » selon l'hypothèse qu'un sourire poli, marqueur possible de gêne, sera moins large qu'un sourire franc amusé. Sur le corpus IS-Comp, les participants ont produit une majorité de sourires ouverts, suggérant qu'ils ont été plus amusés que gênés par l'expérimentation, et ce quelque soit la tâche observée.

5.4 Comparaison entre auto-annotation et annotation externe

Dans le Chapitre 2 nous avons présenté différents facteurs influant la perception et l'expression du stress chez des individus soumis aux mêmes stressors. Le stress est une expérience subjective qui passe en partie par les processus cognitifs d'appraisal ; l'évaluation du stress par une personne ayant accès à ces processus cognitifs (i.e. un individu évaluant son propre stress) sera différente de l'évaluation réalisée par des observateurs externes n'ayant pas accès à cette information et utilisant uniquement les signes extériorisés pour inférer le stress de cette personne. Nous comparons ici les résultats des auto-annotations et des annotations externes du stress sur nos différents corpus.

5.4.1 Sur E-Thérapies

Les corpus ETCC-HM et ETCC-HH ont été annotés par tâche sur une échelle de 1 à 10 pour l'auto-annotation comme pour l'annotation externe. Afin de comparer les deux annotations, nous nous intéressons à la fréquence d'attribution des notes, toutes tâches et toutes populations confondues.

Les courbes présentées en Figure 29 et Figure 30 illustrent la fréquence d'utilisation des étiquettes de stress sur l'échelle à dix niveaux, attribuées en auto-annotation par les sujets ou attribuées par des annotateurs externes.

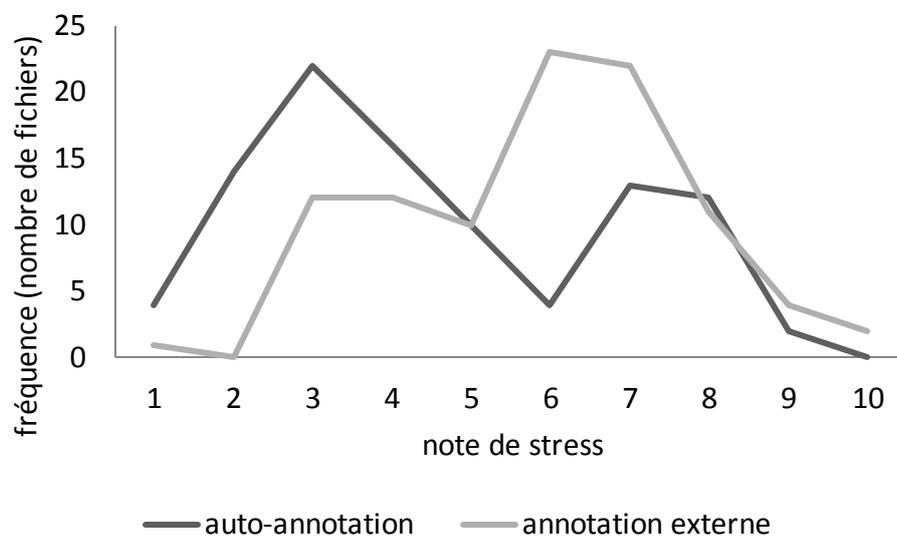


Figure 29 : Fréquence des annotations de stress par type d'annotation sur ETCC-HM

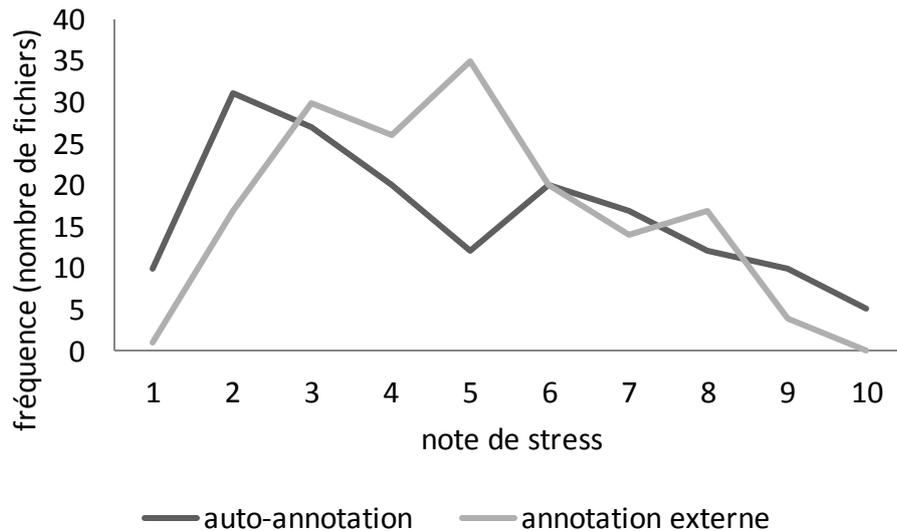


Figure 30 : Fréquence des annotations de stress par type d'annotation sur ETCC-HH

Les courbes observées sont de formes différentes : les auto-annotations sont caractérisées par deux cloches aux deux extrémités de l'échelle d'annotation, suggérant que les participants se sont évalués soit peu stressés soit assez stressés ; les annotations externes sont caractérisées par une cloche unique au milieu de l'axe, suggérant que les observateurs ont perçu un niveau de stress moyen et peu d'extrêmes.

Les courbes présentées en Figure 31 et Figure 32 illustrent la fréquence d'utilisation des étiquettes de stress sur l'échelle à dix niveaux pour les sujets phobiques uniquement, attribuées en auto-annotation par les sujets ou attribuées par des annotateurs externes.

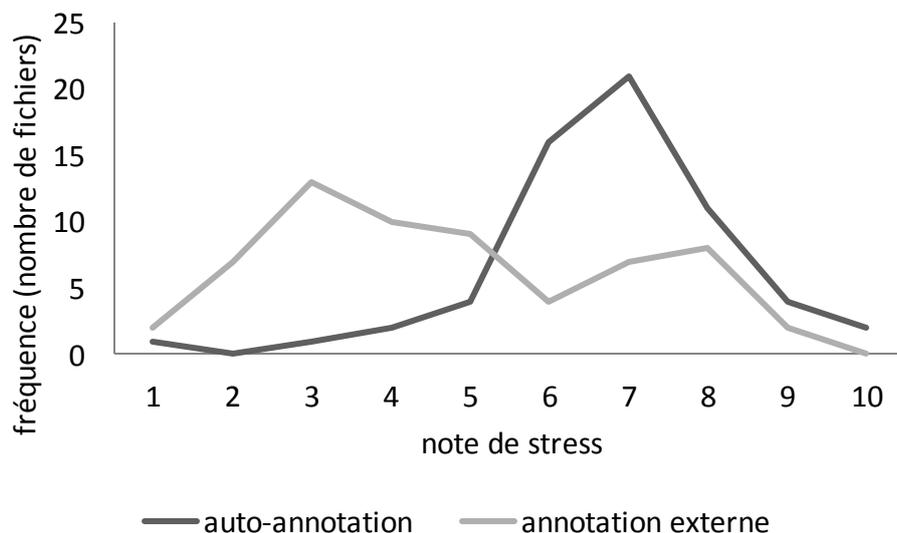


Figure 31 : Fréquence des annotations de stress de la population phobique sur ETCC-HM

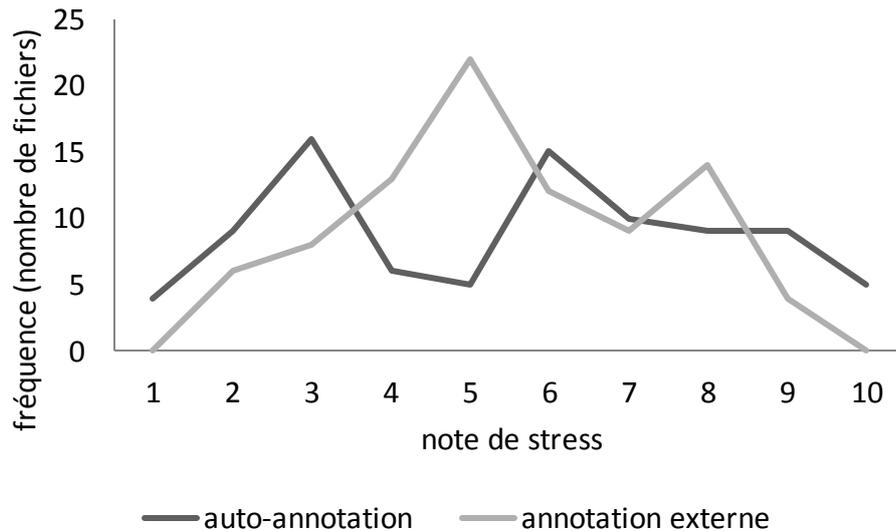


Figure 32 : Fréquence des annotations de stress de la population phobique sur ETCC-HH

Le décalage de perception est particulièrement marqué au sein de la population phobique sur le corpus ETCC-HM. Le graphe présenté en Figure 31 montre des courbes de notes situées à différentes extrémités de l'échelle. Les sociophobes se sont plus souvent évalués particulièrement stressés, là où un observateur externe les a plus souvent perçu comme peu ou modérément stressés. Ce décalage est cependant plus nuancé sur le corpus ETCC-HH, présenté en Figure 32, suggérant un impact du type de stresser utilisé. La présence d'humains dans la salle pendant l'expérimentation ETCC-HM, la différence des tâches entre les deux expérimentations, et le fait que les participants ne sont pas intégralement les mêmes dans les deux expérimentations, ne nous permettent pas de conclure sur l'influence directe d'un humain par rapport à une machine comme source de stress.

Afin de quantifier le décalage entre les deux types d'annotation, nous avons utilisé le calcul de Kappa présenté en 5.3.2. Nous avons comparé les annotations sur l'échelle à dix niveaux, puis sur les notes projetées sur des échelles trois niveaux et binaire. Les scores sont présentés dans les Tableau 19 et Tableau 20.

Echelle	Kappa	Accord brut
10-niveaux	-0,07	5%
3-niveaux	0,081	31%
Binaire	0,17	56%

Tableau 19 : Taux d'accord entre auto-annotations et annotations externes sur ETCC-HM

Echelle	Kappa	Accord brut
10-niveaux	0,15	15%
3-niveaux	0,20	41%
Binaire	0,25	62%

Tableau 20 : Taux d'accord entre auto-annotations et annotations externes sur ETCC-HH

Comme pour l'accord inter-annotateur, la réduction du nombre de graduation de l'échelle d'évaluation améliore le taux d'accord, mais celui-ci reste inexistant à faible. Cette mesure conforte l'hypothèse de différence de perception du stress, en particulier chez des populations dont la pathologie altère la régulation des états anxieux.

5.4.2 Sur Comparse

Le stress sur le corpus CP2 a été évalué sur deux granularités différentes par les sujets et par les annotateurs : les sujets ont évalué leur stress pour l'ensemble de l'expérimentation, sur une échelle de 1 à 5, tandis que les observateurs ont évalué le stress par tâche sur une échelle de 1 à 10. Afin de comparer les deux annotations, nous avons utilisé la moyenne des trois notes attribuées par les annotateurs, rapportée sur une échelle de 1 à 5, comme note pour l'ensemble de l'expérimentation.

Comme pour les corpus E-thérapies, nous avons utilisé le Kappa de Cohen pour quantifier la distance entre les deux types d'annotations, d'abord avec les notes de 1 à 10, puis avec les notes projetées sur des échelles à trois niveaux et binaire. Les scores sont présentés dans le Tableau 21

Echelle	Kappa	Accord brut
10-niveaux	0,25	20%
3-niveaux	0,05	35%
Binaire	0,61	85%

Tableau 21 : Taux d'accord entre auto-annotations et annotations externes sur CP2

L'échelle utilisée pour graduer le stress a une influence notable sur l'accord entre auto-annotation et annotation externe : lorsqu'il s'agit d'évaluer si les participants étaient stressés ou non (échelle binaire) les annotations sont raisonnablement accordées, en revanche lorsqu'il s'agit de mesurer le degré de stress des participants (échelles à dix niveaux et à trois niveaux) les annotations ne concordent pas.

5.4.3 Sur Nao Fait de l'Humour

Le même procédé a été appliqué sur le corpus IS-Comp, où les annotateurs et les sujets ont évalué leur stress sur deux granularités différentes. Nous présentons dans le Tableau 18 les scores kappa pour évaluer les annotations sur des échelles à trois niveaux et binaire.

Echelle	Kappa	Accord brut
3-niveaux	0,23	20%
Binaire	0,23	69%

Tableau 22 : Taux d'accord entre auto-annotations et annotations externes sur IS-Comp

Quelle que soit l'échelle utilisée, le kappa entre les deux annotations reste sensiblement le même et est faible. Une explication possible de cet écart de perception est l'influence de facteurs culturels sur l'évaluation du stress : les participants de l'expérimentation « Nao fait de l'Humour » venaient d'horizons différents, tandis que les deux observateurs qui ont fourni l'annotation externe étaient français. Une autre explication est le type de public recruté et le contexte de la collectes : des personnes avec un haut niveau d'éducation et dans des circonstances professionnelles auront moins de facilité à admettre leur gêne que des observateurs pour qui l'enjeu n'est pas le même.

Chapitre 6 Analyse du signal pour l'extraction d'indices du stress

6.1 Extraction d'indices

L'élaboration de systèmes de reconnaissance automatique du stress à partir de signaux émis par un individu passe par l'extraction d'indices représentatifs de ces signaux (c.f. Chapitre 3 et Figure 4). Cette extraction soulève plusieurs questions : quelles modalités observer pour y voir des signes de stress ? A quelle fréquence observer ces signes ? Quelles mesures utiliser pour caractériser au mieux ces signes ?

Dans le Chapitre 2, nous avons présenté les modalités les plus fréquemment utilisées dans la littérature pour la détection du stress. Dans cette étude nous nous intéressons aux indices mesurables à l'aide de capteurs peu invasifs, comme des micros ou des caméras. Nous présentons ici les mesures extraites à partir de la voix, du visage, et de la posture des sujets. Une étude comparative avec des mesures physiologiques collectées dans le cadre du projet Compare (c.f. 4.3) est présentée dans la seconde partie de ce chapitre.

6.1.1 Observation du signal

Dans cette étude, nous utilisons des enregistrements « longs » (de 30s à 5min) pour observer des signes de stress. Afin d'observer l'impact de la fréquence d'échantillonnage des marqueurs de stress, nous comparons trois tailles de fenêtre d'analyse. Pour chacune des modalités étudiées, nous comparons des fenêtres de 1s, 2s et 5s.

Pour tous les enregistrements, nous avons procédé à une analyse par fenêtre glissante : chaque fenêtre comporte N secondes d'enregistrement, et les fenêtres se chevauchent de façon à ce que $N/2$ secondes soient incluses dans deux fenêtres consécutives. Le procédé est illustré en Figure 33.

Pour extraire des enregistrements collectés dans cette étude les descripteurs précédemment décrits, nous avons appliqué plusieurs largeurs de fenêtres : $N = 1s$ (0.5s de chevauchement), $N = 2s$ (1s de chevauchement), et $N = 5s$ (2.5s de chevauchement).

La quantité d'information contenue dans chaque fenêtre de N secondes dépend de la fréquence de capture de l'échantillon considéré. Dans cette étude les enregistrements audio sont échantillonnés à 16KHz, les enregistrements vidéo capturés à l'aide de la webcam et les enregistrements de matrices de profondeur capturés à l'aide de la Kinect sont échantillonnés à 25 images par seconde.

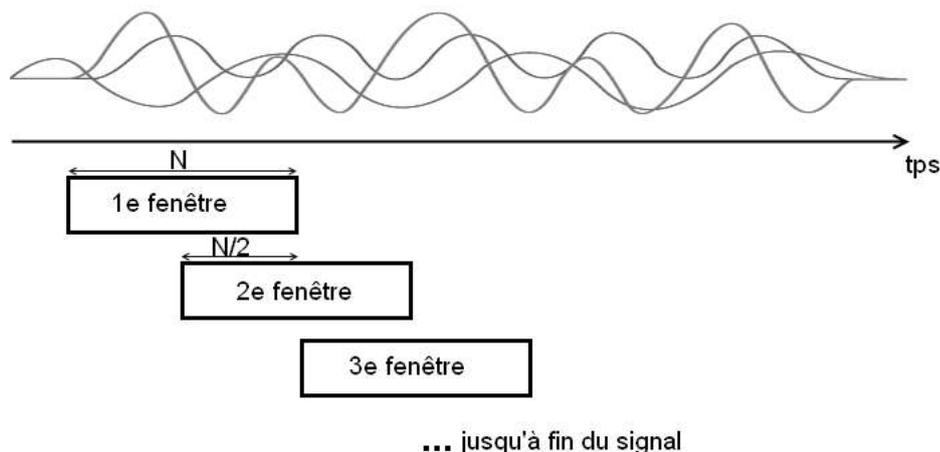


Figure 33 : Processus d'analyse par fenêtres glissantes avec 50% de recouvrement

6.1.2 Audio

Nous avons vu au Chapitre 2 que les études portant sur les caractéristiques acoustiques des voix stressées ont montré l'intérêt du rythme, des contours de l'énergie, de la fréquence fondamentale et des paramètres spectraux dans la reconnaissance du stress.

Dans cette étude, l'extraction de paramètres audio a été réalisée à l'aide de l'outil OpenSMILE [96] dans sa version 1, en utilisant le set de paramètres audio emo_IS09. Le set comprend (16 paramètres bas niveaux + 16 dérivées) x 12 mesures statistiques = 384 descripteurs quantifiant la voix. Le set est détaillé dans le Tableau 23.

Descripteurs calculés par trame (16 + 16 Δ)	Fonctionnelles statistiques(12)
<ul style="list-style-type: none"> • Taux de passage par zéro (ZCR), Δ ZCR, • Energie RMS, Δ Energie RMS, • F0 (normalisée à 500Hz), Δ F0, • Probabilité de Voix , Δ Probabilité de Voix • MFCC [1-12], Δ MFCC [1-12] 	<ul style="list-style-type: none"> • moyenne, déviation standard, • asymétrie (skewness), aplatissement (kurtosis), • extrêmes : minimum, maximum, position relative, étendue (range) • régression linéaire : décalage (offset), décalage MSE (Mean Square Error), pente (slope)

Tableau 23 : Les 384 descripteurs audio du set emo_IS09 d'OpenSMILE

Collecte, Annotation et Analyse de Signaux de Stress

Les descripteurs du set emo_IS09 reprennent les caractéristiques identifiées pour la reconnaissance du stress dans la voix. Ce set a été proposé lors du Challenge Interspeech 2009 pour la détection d'émotions dans la voix [270], et a été utilisé dans l'étude de reconnaissance des émotions dans la voix [56] [296] [299], y compris pour la reconnaissance du stress [336].

6.1.3 Visage

Nous avons vu au Chapitre 2 que les expressions faciales étaient porteuses d'informations affectives. Dans notre étude, nous souhaitons exploiter ces expressions pour la reconnaissance du stress.

Les travaux d'Ekman et le système FACS* [88] montrent que les expressions faciales sont contrôlées par les parties mobiles du visage, c'est-à-dire zones des yeux et de la bouche. L'analyse de ces zones du visage à partir d'images est d'ailleurs conjointe dans le cerveau humain pour la perception des expressions [321]. Nous nous intéressons donc particulièrement aux froncements et haussement de sourcils, et aux sourires et moues. Les indices décrivant la position et le mouvement de ces parties du visage sont largement employés pour la reconnaissance d'expressions faciales, par annotation manuelle [199], par modèles d'apparence (Active Appearance Model AAM*) [188] ou encore par motif binaire local (Local Binary Patterns LBP) [309].

Nous avons utilisé le logiciel développé par J. M. Saragih et maintenu par Kyle McDonald, FaceTracker [253], afin d'extraire des images capturées par la webcam un réseau de points marquant les saillances du visage (sourcils, yeux, nez, lèvres, et menton) grâce à un modèle d'apparence. Le modèle appliqué extrait 66 points, illustrés par le schéma sur la partie gauche de la Figure 34.

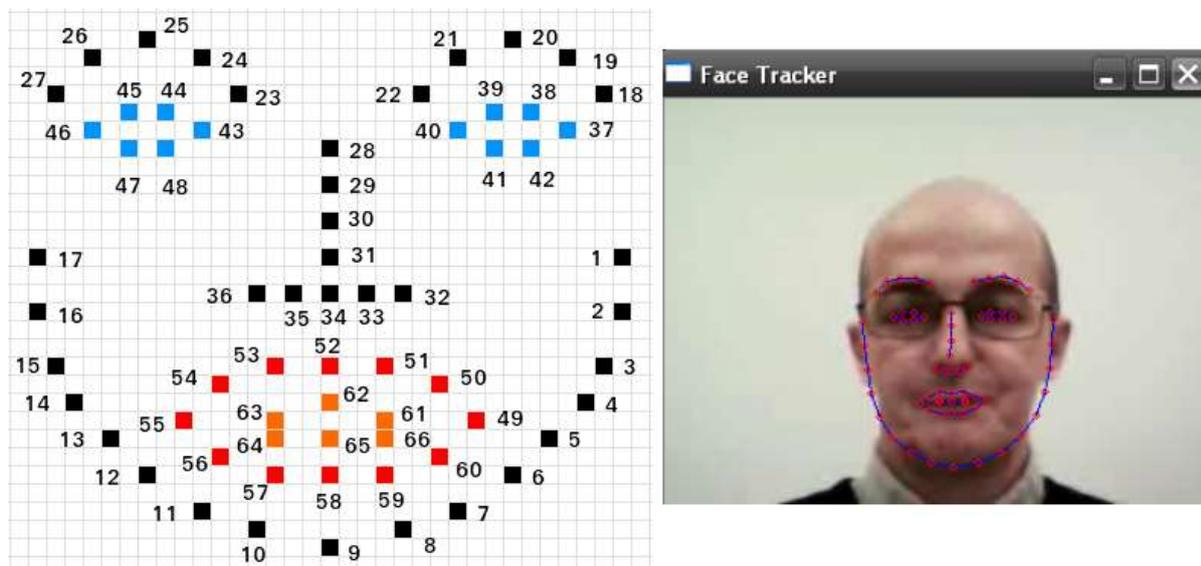


Figure 34 : Le modèle à 66 points de FaceTracker

A partir des points extraits à l'aide de FaceTracker, nous calculons pour chaque image les distances suivantes : arche des sourcils, écart entre les sourcils, écart entre les sourcils et les yeux, ouverture de la bouche, rapport entre l'épaisseur de la lèvre supérieure et l'épaisseur de la lèvre inférieure.

Les trois premières mesures quantifient les mouvements des sourcils, les deux dernières ceux de la bouche. Le calcul des distances est présenté en annexe.

Chaque individu possédant des caractéristiques faciales uniques, il est nécessaire de normaliser ces mesures de distances afin de pouvoir les comparer d'une personne à l'autre. La distance entre le visage et la caméra impacte également ces mesures : plus le visage s'éloigne, plus les points sont proches. La normalisation des valeurs calculées est également nécessaire pour les comparer d'une image à l'autre si la personne filmée s'est éloignée ou rapprochée de la caméra.

Pour normaliser les mesures, nous avons utilisé la longueur et la largeur de la « boîte englobante » du visage, c'est-à-dire le rectangle minimum contenant l'intégralité des points du réseau. Les distances mesurées en abscisse sont relativisées par la largeur de la boîte, les distances mesurées en ordonnées par la hauteur. De plus, les dimensions et la position de la boîte englobante nous donnent une indication des mouvements du visage devant la caméra.

Collecte, Annotation et Analyse de Signaux de Stress

Les mesures décrites quantifient l'expression du visage pour une image donnée. Mais les expressions faciales sont dynamiques, il est donc intéressant d'observer leurs variations sur une fenêtre de plusieurs images. Pour chaque fenêtre, nous avons M valeurs¹⁵ pour les 5 distances observées. Nous calculons les dérivées de ces distances, pour quantifier les variations d'une image à la suivante. Nous calculons également les dérivées de la longueur, la largeur, la position en abscisse et la position en ordonnées de la boîte englobante. Nous avons donc 9 mesures dérivées.

Nous calculons 7 descripteurs statistiques sur les M valeurs des 5 distances, et les 9 dérivées : moyenne, déviation standard, minimum, maximum, étendue (range), asymétrie (skewness), et aplatissement (kurtosis).

Pour chaque fenêtre nous avons donc (5 distances + 9 dérivées) x 7 mesures statistiques = 98 descripteurs quantifiant les expressions faciales. Le set est récapitulé dans le Tableau 24.

Descripteurs calculés par trame (5 + 9 Δ)	Fonctionnelles statistiques(7)
<ul style="list-style-type: none">• Arche sourcils, Δ Arche sourcils• Ecart sourcils, Δ Ecart sourcils• Ecart sourcils - yeux, Δ Ecart sourcils - yeux• Ouverture bouche, Δ Ouverture bouche• Rapport lèvres, Δ Rapport lèvres• Δ largeur et Δ hauteur boîte englobante• Δ abscisse et Δ ordonnée boîte englobante	<ul style="list-style-type: none">• moyenne, déviation standard,• asymétrie (skewness), aplatissement (kurtosis)• extrêmes : minimum, maximum, étendue

Tableau 24 : Les 98 paramètres décrivant les expressions faciales ; un 99^{ème} paramètre indique la probabilité de la présence d'un visage dans la séquence d'images observée.

Les enregistrements sur lesquels nous travaillons sont capturés dans des conditions peu contraintes : bien qu'assis devant un écran, les individus filmés sont libres de bouger la tête ou de se tourner. Par conséquent, il n'est pas toujours possible de détecter un visage de face dans les images, par exemple si l'individu est de profil, ou si son visage est partiellement hors du champ de la webcam.

¹⁵ M correspond au nombre d'images pour N secondes film, ici à 25FPS

Collecte, Annotation et Analyse de Signaux de Stress

Nous avons tenu compte de cette information dans nos mesures, en ajoutant aux 98 descripteurs une mesure de fiabilité, calculée comme suit : pour chaque image, nous utilisons le modèle HaarCascade d'OpenCV [30] pour identifier un visage vu de face. Si un visage est effectivement détecté nous extrayons les points saillants avec FaceTracker, sinon nous marquons l'image comme « non utilisable ». Lorsque N secondes de film ont été analysées, nous calculons le rapport entre le nombre d'images utilisables et le nombre total d'images dans la fenêtre. Ce rapport est le 99^{ème} descripteur de la fenêtre, et indique la fiabilité des 98 autres mesures.

6.1.4 Posture

Nous avons vu au Chapitre 2 que la posture et les gestes portent des informations affectives, en particulier les qualités d'amplitude et de célérité des mouvements.

L'un des corpus de cette étude, CP2, comporte l'enregistrement des sujets par une Kinect, dont le capteur de profondeur permet de distinguer différents objets dans une scène. Pour chaque image capturée par la Kinect, nous utilisons le framework OpenNI [221] pour générer une matrice de pixels dont les valeurs ne représentent pas des couleurs (comme pour une capture d'image classique) mais une distance au capteur. En parcourant cette matrice il est possible d'isoler un groupe de pixels suffisamment grand et situé dans une plage de distances donnée. La mise en scène de l'expérimentation Comparse (c.f. 4.3) nous permet de considérer que le plus grand groupe de plus de 5000 pixels situé entre 1,5 et 3,5m correspond à la silhouette du sujet.

Afin de qualifier les mouvements des sujets, nous utilisons l'Indice de Contraction (Contraction Index, CI*) et la Quantité de Mouvement (Quantity of Motion QoM*) proposés par Camurri [41]. Les informations décrivant les mouvements corporels (fluidité, vitesse, amplitude) ont été reliées à l'expression d'émotions (pour leur génération [248] comme pour leur identification [76]), mais pas au stress à notre connaissance.

Le CI* décrit dans quelle proportion le sujet observé est ramassé sur lui-même, en particulier à quel point ses bras sont proches de son corps. Il correspond au rapport entre l'aire de la silhouette du sujet et l'aire de la « boîte englobante » de cette silhouette (le plus petit rectangle contenant l'ensemble des pixels de la silhouette).

Pour une image donnée i , l'indice de contraction est : $CI(i) = A(b_i) - A(s_i)$, où $A(b_i)$ est l'aire de la boîte englobante de la silhouette dans l'image i , et $A(s_i)$ est l'aire de la silhouette dans l'image i .

La Figure 35 illustre la silhouette d'un sujet capturée par la Kinect, après filtrage pour ne conserver que le groupe de pixels dans la plage de profondeurs acceptable.

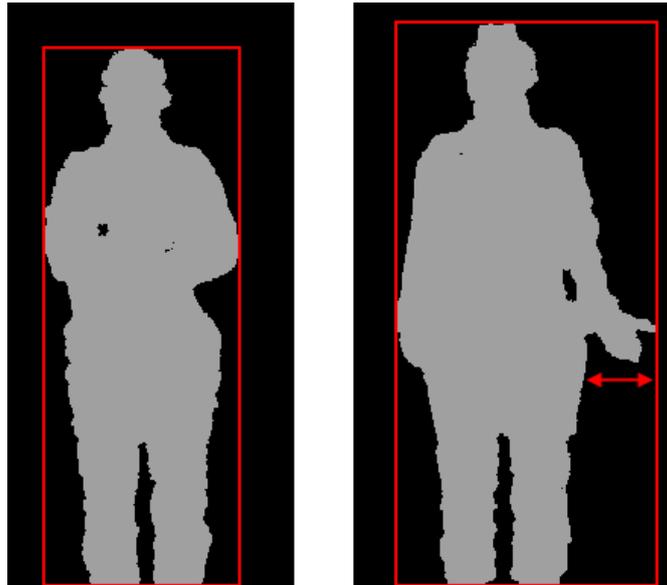


Figure 35 : Silhouette capturée par la Kinect, et boîte englobante

Sur la partie gauche le sujet garde les bras croisés devant le corps : l'aire de la silhouette et l'aire de la boîte englobante (tracée en rouge) sont proches, le CI^* est proche de 1. Sur la partie droite le sujet écarte son bras gauche, l'aire de la boîte englobante est plus grande que celle de la silhouette et le CI^* est plus petit.

Cet indice décrit la posture du sujet à un instant donné, mais comme pour les expressions faciales, l'évolution de la posture dans le temps apporte des informations supplémentaires. Nous avons donc analysé des fenêtres de plusieurs images. Pour chaque fenêtre, nous avons M valeurs du CI^* . Nous calculons la dérivée du CI^* , pour quantifier les variations d'une image à la suivante. Puis nous calculons 7 descripteurs statistiques sur les M valeurs du CI^* et de sa dérivée : moyenne, déviation standard, minimum, maximum, étendue (range), asymétrie (skewness), et aplatissement (kurtosis).

La QoM^* décrit le déplacement d'un corps dans l'espace sur une période de temps donnée ; il est calculé sur une suite d'images successives, en notant pour chaque image les pixels occupés par la silhouette pour construire la « silhouette du mouvement », une image représentant la forme et la position du corps sur la succession d'images considérée. Il s'agit d'une variante des images mouvement-énergie (Motion-Energy Images) qui exclut la silhouette de la dernière image pour ne retenir que la silhouette du mouvement.

La Figure 36 illustre la silhouette de mouvement extraite de deux séries d'images. La silhouette de gauche illustre une faible quantité de mouvements, la silhouette de droite une quantité de mouvements plus importante au niveau du bras, de l'épaule et de la tête.

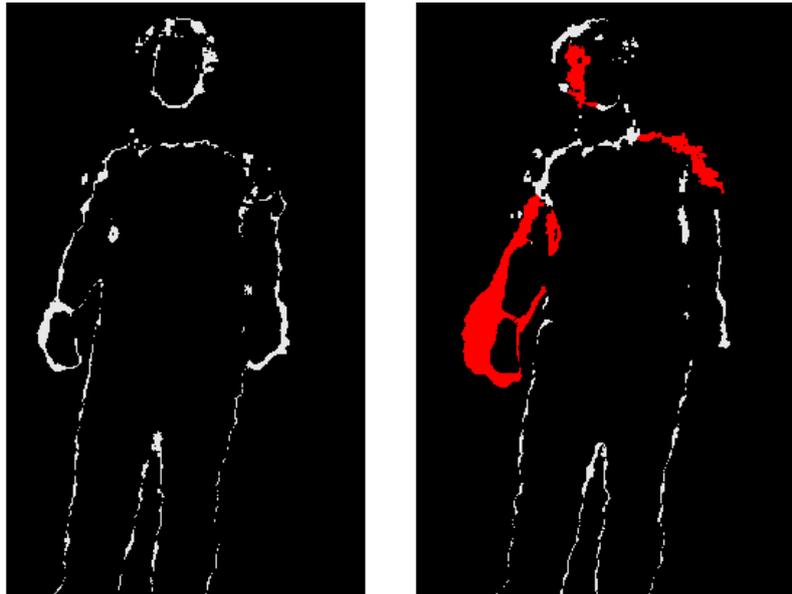


Figure 36 : Quantité de Mouvement à partir de la silhouette capturée par la Kinect

La QoM* correspond au rapport de l'aire de la silhouette de « mouvement » (les pixels blancs et rouges dans la Figure 36) sur l'aire de la silhouette « mouvement et corps » (les pixels blancs et rouges, et les pixels noirs à l'intérieur de la silhouette). Plus ce rapport est élevé, plus le mouvement aura été important pendant la série d'images considérée. L'algorithme de calcul de la QoM* est détaillé en annexe.

Pour chaque fenêtre, nous avons donc $(CI + \text{dérivé}) \times 7$ mesures statistiques + QoM = 15 mesures quantifiant la posture. Le set est récapitulé dans le Tableau 25.

Descripteurs calculés par trame (1+ 1 Δ)	Fonctionnelles statistiques(7) et descripteur calculé par fenêtre (1)
<ul style="list-style-type: none"> • CI, Δ CI 	<ul style="list-style-type: none"> • moyenne, déviation standard, • asymétrie (skewness), aplatissement (kurtosis) • extrêmes : minimum, maximum, étendue
--	QoM

Tableau 25 : Les 15 descripteurs posturaux calculés à partir de la silhouette des sujets

6.1.5 Récapitulatif des indices

Modalité	Indices
Audio	Descripteurs calculés par trame (16 + 16 Δ) Taux de passage par zéro (ZCR), Δ ZCR, Energie RMS, Δ Energie RMS, F0 (normalisée à 500Hz), Δ F0, Probabilité de Voix , Δ Probabilité de Voix MFCC [1-12], Δ MFCC [1-12]
	Fonctionnelles statistiques (12) moyenne, déviation standard, asymétrie (skewness), aplatissement (kurtosis), extrêmes : minimum, maximum, position relative, étendue (range) régression linéaire : décalage (offset), décalage MSE (Mean Square Error), pente (slope)
Visage	Descripteurs calculés par trame (5 + 9 Δ) Arche sourcils, Δ Arche sourcils Ecart sourcils, Δ Ecart sourcils Ecart sourcils - yeux, Δ Ecart sourcils - yeux Ouverture bouche, Δ Ouverture bouche Rapport lèvres, Δ Rapport lèvres Δ largeur et Δ hauteur boîte englobante Δ abscisse et Δ ordonnée boîte englobante
	Fonctionnelles statistiques (7) moyenne, déviation standard, asymétrie (skewness), aplatissement (kurtosis) extrêmes : minimum, maximum, étendue
Posture	Descripteurs calculés par trame (1+ 1 Δ) CI, Δ CI
	Fonctionnelles statistiques(7) et descripteur calculé par fenêtre (1) moyenne, déviation standard, asymétrie (skewness), aplatissement (kurtosis) extrêmes : minimum, maximum, étendue QoM

Tableau 26 : Récapitulatif des indices extraits des enregistrements de cette étude.

6.2 Etude de cas

6.2.1 Corrélations entre mesures audio, posturales, et évaluation de performance

Une première étude [II] sur les données du corpus CP2 a fait l'objet d'une publication lors de la conférence IEEE ACII (Affective Computing and Intelligent Interaction) en 2013, impliquant l'ensemble des acteurs du projet Compare.

L'objectif de cette étude était d'observer les corrélations entre les mesures de retour d'expérience (frustration, motivation, stress, difficulté perçue), l'évaluation de performance par les juges, et les différents paramètres extraits des enregistrements des multiples capteurs. Dans cette précédente étude nous nous sommes intéressés aux corrélations entre les paramètres audio, posturaux et les mesures de retour d'expérience et de performance.

L'analyse audio a été réalisée à l'aide du logiciel Praat [27] pour extraire 5 indices et en étudier la valeur globale sur l'ensemble de la session pour chaque sujet. De la même façon, la valeur globale des paramètres posturaux de contraction (CI*) et de mouvement (QoM*) a été collectée sur l'ensemble de la session pour chaque sujet. Les indices mesurés sont récapitulés dans le Tableau 27.

Audio (7 paramètres)	Posture (4 paramètres)
F0 (moyenne, étendue)	Indice de Contraction CI (moyenne, écart-type)
Intensité (moyenne, étendue)	
Jitter (moyenne)	Quantité de Mouvement QoM (moyenne, écart-type)
Shimmer (moyenne)	
Taux de silences	

Tableau 27 : Indices audio et visuels de l'étude [II]

Les mesures audio et posturales observées ont révélé des corrélations avec les mesures de retour d'expérience et de performance. La performance évaluée par les juges est corrélée au taux de silences, à l'intensité et au jitter moyen, mais à aucune des mesures de posture. La frustration est corrélée aux variations de F0 et de CI* ; la difficulté perçue est corrélée à la valeur moyenne de F0, à la valeur moyenne et aux variations de QoM* ; la motivation est négativement corrélée aux valeurs moyennes et aux variations de CI* ; le stress auto-évalué est corrélé à la valeur moyenne et aux variations de QoM*.

Ces corrélations sur 18 sujets sont des tendances indiquant une possibilité pour la détection automatique de l'expérience subjective d'un individu à partir d'indices mesurés dans sa voix et sa posture ; elles révèlent également les informations influençant l'évaluation de performance dans un exercice d'entretien d'embauche par des observateurs externes.

6.2.2 *Corrélations entre affect bursts et facteurs individuels*

Dans le Chapitre 5 (5.3.5), nous avons présenté l'annotation d'affect bursts, réalisés sur les corpus IS-Comp en audio-visuel, et IS-Nao en audio uniquement. Ces événements semblent mal corrélés avec la gêne – auto-annotée ou annotée par des observateurs externes – mais apportent des informations sur les différences interindividuelles des sujets.

Nous avons réalisé une première étude [VII] portant sur les rires amusés et embarrassés des sujets durant l'intégralité de l'expérimentation (i.e. les corpus IS-Comp et IS-Nao). L'objectif de cette étude était d'observer l'influence de facteurs comme l'origine culturelle, le sexe, la personnalité et le sens de l'humour sur la production de rires dans une interaction avec une machine et avec un robot.

Les observations ont révélé des différences culturelles (les asiatiques ont produit plus de rires embarrassés, les américains plus de rires amusés), liées au sens de l'humour (les sujets ayant un score plus élevé au questionnaire SHS* [200] ont plus ri), aux traits de personnalités (l'extraversion est positivement corrélée aux rires amusés, l'ouverture est négativement corrélée aux rires embarrassés) et à la tâche (les vidéos amusantes ont induit le plus de rires amusés, les fourchelanges le plus de rires embarrassés).

Nous avons réalisé une seconde étude [VI] portant sur les rires et les sourires des sujets durant la première partie de l'expérimentation (i.e. le corpus IS-Comp). L'objectif de cette étude était d'observer les informations apportées par les différentes modalités expressives.

Les observations ont révélé des différences culturelles (les asiatiques ont produit plus de rires ouverts, les européens plus de rires fermés) et liées au sens de l'humour (les sujets ayant un score plus élevé au questionnaire SHS* [200] ont produit moins de sourires fermés), Il apparaît également que les événements visuels (i.e. des sourires) étaient plus fréquents que les événements audio (i.e.. les rires) pendant l'interaction face à un écran, soulignant l'apport de modalités supplémentaires pour l'extraction d'informations.

Les corrélations observées dans ces deux études portent sur un petit nombre de participants pour établir des comportements culturels (45 sujets : 6 américains, 27 européens, 12 asiatiques) ; elles révèlent cependant des tendances qui indiquent des différences de comportement. Elles révèlent également une influence des traits de personnalité extraversion et ouverture, et du sens de l'humour sur les comportements des sujets. Dans le Chapitre 1, nous avons identifié l'humour comme une stratégie de coping positive face au stress : l'influence de l'humour sur la production d'affect bursts souligne l'intérêt de ces événements pour un système de reconnaissance du stress.

6.2.3 Variations des paramètres considérés pour des cas particulier de stress

Afin d'illustrer le comportement des paramètres audio, faciaux et posturaux présentés en 6.1, nous avons utilisé une analyse en composantes principales (Principal Component Analysis PCA*) des différents corpus, et de cas particuliers de sujets particulièrement stressés et particulièrement non stressés. Les fonctions de densité de probabilité du 1^{er} composant sont illustrées en Figure 37, Figure 38 et Figure 39 pour les paramètres audio, faciaux et posturaux, respectivement.

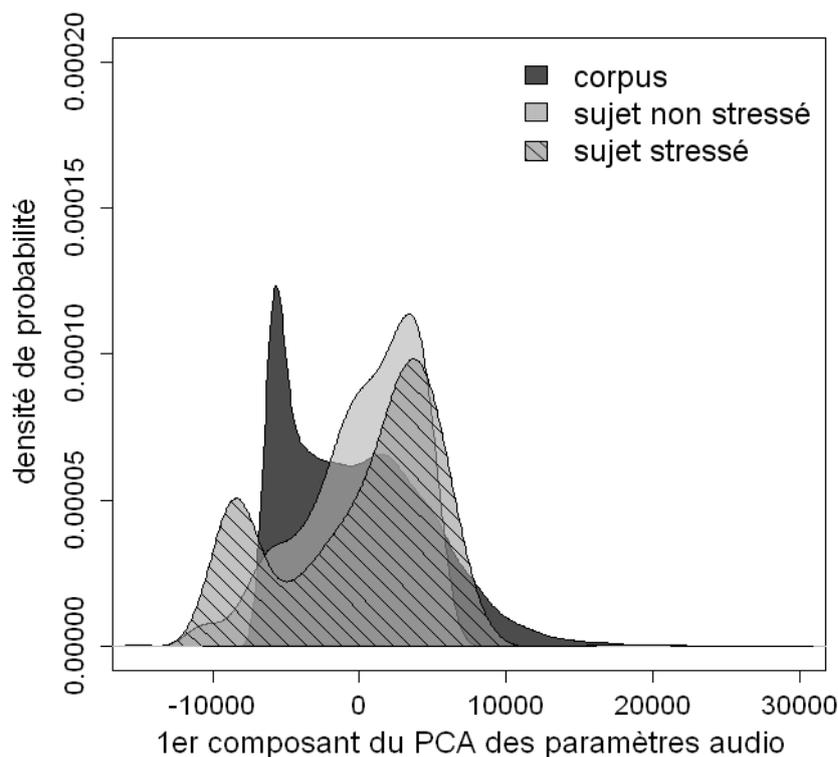


Figure 37 : Densité de probabilité du premier composant de l'Analyse en Composantes Principales (PCA) des paramètres audio sur le corpus ETCC-HH

La Figure 37 illustre la densité de probabilité du 1^{er} composant de l'analyse PCA* des paramètres audio pour les données issues du corpus ETCC-HH (courbe la plus foncée). A des fins de comparaison, un sujet phobique particulièrement stressé (courbe hachurée) et un sujet témoins particulièrement à l'aise (courbe la plus claire) sont également représentés.

La forme des courbes suggère des comportements audio comparables entre le participant stressé et le participant non stressé. La courbe générale du corpus suggère des comportements plus semblables à ceux du sujet stressé que du sujet non stressé.

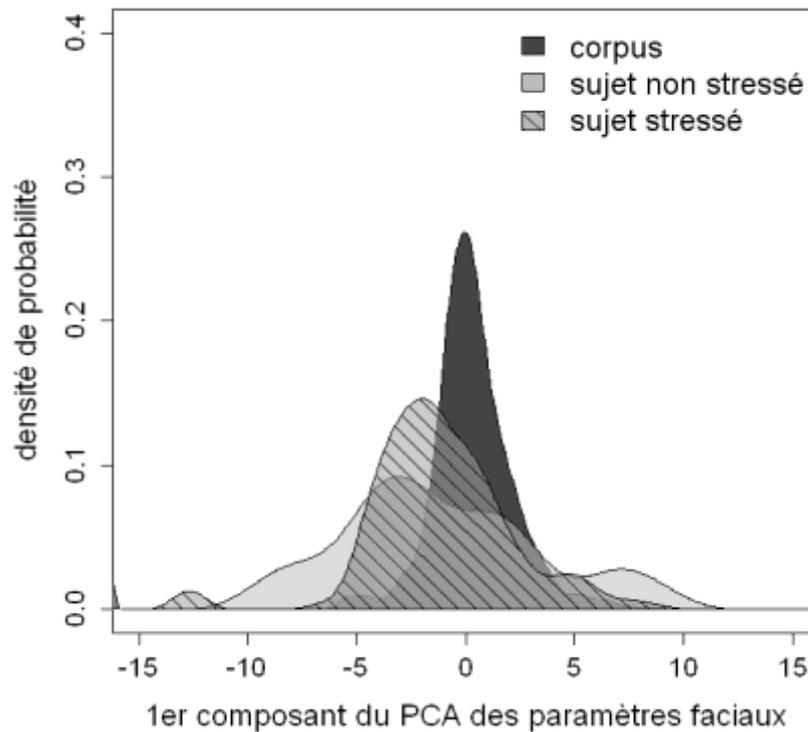


Figure 38 : Densité de probabilité du premier composant de l'Analyse en Composantes Principales (PCA) des paramètres faciaux sur le corpus ETCC-HM

La Figure 37 illustre la densité de probabilité du 1^{er} composant de l'analyse PCA* des paramètres faciaux pour les données issues du corpus ETCC-HM (courbe la plus foncée). A des fins de comparaison, un sujet phobique particulièrement stressé (courbe hachurée) et un sujet témoins particulièrement à l'aise (courbe la plus claire) sont également représentés.

La forme des courbes suggère des comportements faciaux comparables entre le participant stressé et le participant non stressé. La courbe générale du corpus suggère des comportements plus semblables à ceux du sujet stressé que du sujet non stressé.

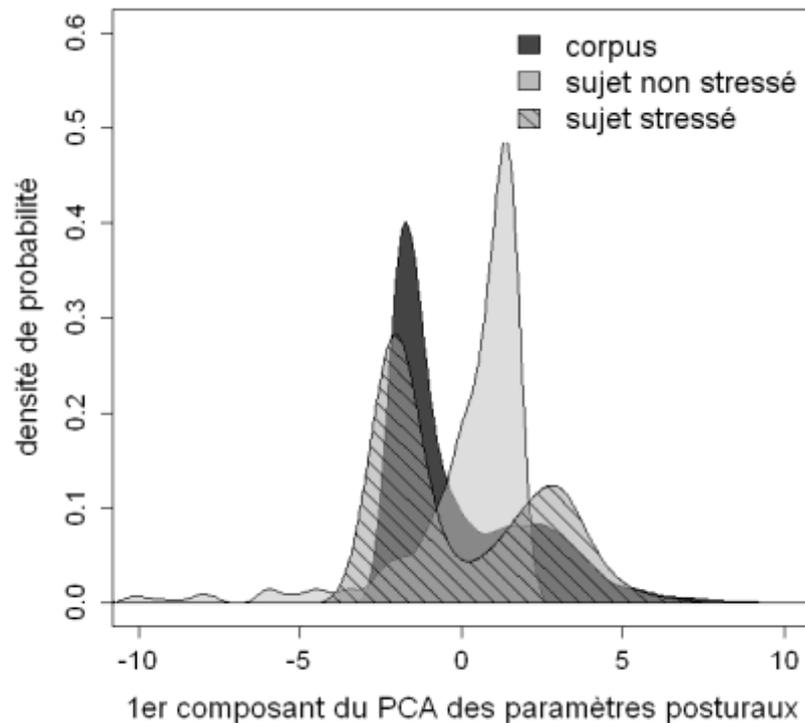


Figure 39 : Densité de probabilité du premier composant de l'Analyse en Composantes Principales (PCA) des paramètres posturaux sur le corpus CP2

La Figure 37 illustre la densité de probabilité du 1^{er} composant de l'analyse PCA* des paramètres posturaux pour les données issues du corpus CP2 (courbe la plus foncée). A des fins de comparaison, un sujet particulièrement stressé (courbe hachurée) et un sujet particulièrement à l'aise (courbe la plus claire) sont également représentés.

La forme des courbes suggère des différences de comportements posturaux entre le participant stressé et le participant non stressé. La courbe générale du corpus suggère des comportements plus semblables à ceux du sujet stressé que du sujet non stressé.

Synthèse de la Collecte, Annotation et Analyse de Signaux de Stress

Nous avons contribué à la collecte de 5 corpus de stress, auprès de 3 populations :

- Des sociophobes et un groupe de contrôle de témoins non-pathologiques dans des exercices de Thérapie Cognitive Comportementale face à une machine (corpus ETCC-HM) et face à des humains (corpus ETCC-HH)
- Des étudiants dans une simulation d'entretien d'embauche face à des humains (corpus CP2-CP3-CP4)
- Des chercheurs d'origines internationales dans des exercices de diction et dans des tâches humoristiques, face à une machine (corpus IS-Comp) et face au robot Nao (corpus IS-Nao).

Ces expérimentations illustrent la diversité du stress, suivant les contextes mais également à travers les différences interindividuelles, interculturelles, et inter-expérimentales.

Les corpus collectés dans cette étude ont été annotés pour évaluer le stress en auto-annotation, par les sujets des expérimentations, et en annotation externe, par des annotateurs experts. Nous avons constitué un guide d'annotation à destination des annotateurs pour l'évaluation du stress à partir d'indices issues de la littérature en psychologie.

Nous avons confronté les auto-annotations aux annotations externes et montré leur non-concordance, résultat confortant la subjectivité de la perception du stress.

Nous avons évalué la consensualité de trois échelles d'annotation du stress :

- une échelle graduant le degré d'intensité du stress de 1 à 10
- une échelle graduant l'intensité du stress suivant 3 niveaux (« faible », « moyen », « élevé »)
- une échelle « binaire », indiquant la présence ou l'absence de stress

L'annotation binaire étant la plus consensuelle, que ce soit entre les différents annotateurs ou entre les auto-évaluations par les sujets d'expérimentation et les évaluations des observateurs externes, nous utiliserons uniquement cette évaluation dans la suite de cette étude.

Nous avons utilisé les données des différents corpus collectés durant cette étude afin d'en extraire des indices pour caractériser la voix, les expressions faciales, et la posture des participants afin d'en identifier le stress.

Nous avons étudié ces indices en corrélation avec les différentes mesures de caractéristiques individuelles susceptibles d'influer sur l'expérience et l'expression du stress (traits de personnalité, facteurs culturels, humour, mesure de performances dans les tâches), et sur des cas particuliers de stress.

Section III Détection Automatique du Stress

Dans cette partie, nous présentons les expériences menées pour évaluer la faisabilité d'un système de reconnaissance automatique du stress à partir d'indices vocaux, faciaux et posturaux. Cette partie se divise en trois chapitres :

Le Chapitre 7 « Modèles monomodaux » présente les performances de modèles monomodaux, et compare les résultats obtenus sur les différentes populations et les différents contextes expérimentaux, afin d'évaluer s'il certaines configurations d'analyse du signal sont plus efficaces que d'autres pour la reconnaissance du stress.

Le Chapitre 8 « Fusion d'informations multimodales » présente les performances de modèles multimodaux, et compare les résultats obtenus par différentes stratégies de fusion, afin d'évaluer l'apport d'informations complémentaires pour la reconnaissance du stress.

Le Chapitre 9 « Facteurs individuels d'influence sur I » présente une étude de l'influence de trois facteurs individuels sur les performances des systèmes: l'existence d'un trouble anxieux, des stratégies de coping fonctionnelles ou dysfonctionnelles, et des traits de personnalités. Ce chapitre discute la signification des résultats obtenus pour l'élaboration de logiciels de remédiation.

Chapitre 7 Modèles monomodaux de reconnaissance du stress

7.1 Pré-traitements

Les données collectées et annotées lors des différentes expérimentations menées pendant cette thèse ont été utilisées pour entraîner des modèles de reconnaissance du stress.

Comme nous l'avons montré au Chapitre 5, l'annotation du stress est réalisée à partir d'un ensemble de signes dont la prévalence va varier d'un individu à l'autre et dont la perception est subjective ; les annotations par des observateurs externes sont modérément consensuelles, et ne sont pas en accord avec les annotations réalisées par les participants. Le stress perçu et le stress ressenti sont différents, aussi nous avons décidé de comparer les performances de modèles entraînés suivant les deux informations dans ce chapitre. Nous utilisons les annotations de l'échelle binaire à deux niveaux, c'est-à-dire que les données sont annotées comme « stressées » ou « non-stressées », car il s'agit de l'annotation la plus consensuelle sur les différents corpus.

Pour chaque modèle, les annotations de stress auto-annoté ou annoté par un tiers ont été associées aux paramètres audio, visuels ou posturaux extraits des enregistrements. Les paramètres ont été extraits par fenêtres d'analyse glissantes (c.f. 6.1.1) et l'ensemble des fenêtres d'un même enregistrement (i.e. les tâches des différentes expérimentations) portent l'étiquette « stress » ou « non-stress » attribuée à cet enregistrement.

Dans la Section I, nous avons présenté le stress comme une réaction physiologique et psychologique impactée par des facteurs individuels mais également contextuels. Afin de tenir compte de ces effets dans la conception des modèles de reconnaissance du stress, nous avons créé des modèles distincts pour les différents corpus collectés, puisqu'ils représentent des contextes différents.

Dans nos expérimentations, nous avons enregistré des populations avec des caractéristiques variées : des sociophobes et des sujets non-pathologiques, avec des traits de personnalités hétérogènes. Afin de tenir compte de cette diversité, nous avons conçu des modèles indépendants du locuteur pour chaque corpus. Nous avons procédé en découpant chaque corpus en sous-ensembles d'entraînement et de test, et en nous assurant qu'un locuteur présent dans le sous-ensemble d'entraînement n'apparaissait pas dans le sous-ensemble de test, et réciproquement. Nous avons divisé nos corpus en tiers, deux tiers des sujets servant à l'entraînement des modèles, et le tiers restant servant à l'évaluation du modèle. Le découpage est illustré en Figure 40.

Détection Automatique du Stress

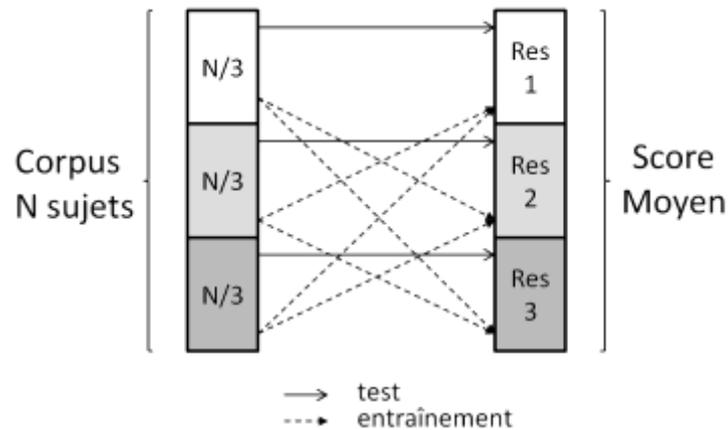


Figure 40 : Découpage des corpus en 3 sous-ensembles pour des modèles indépendants du locuteur

L'une des difficultés des données utilisées dans cette étude est que tous les locuteurs n'ont pas une production équilibrée de stress et de non-stress. Ainsi, certains sujets n'ont produit que des enregistrements annotés « stress », et ce quelle que soit le type d'annotation utilisé. Dans ces circonstances, le découpage des corpus en sous-ensembles indépendants des locuteurs a été précédé d'une étape de sélection des locuteurs suivant l'homogénéité de leur production. Cette répartition est illustrée dans la Figure 41

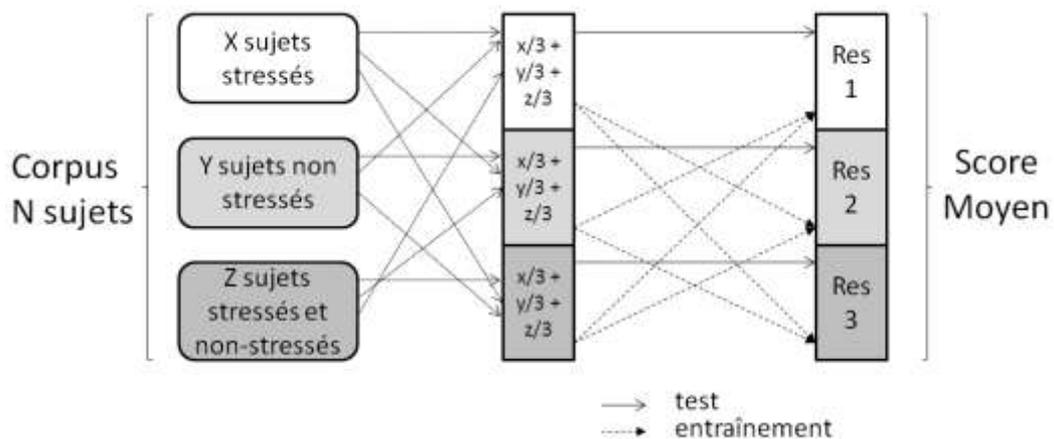


Figure 41 : Découpage des corpus en 3 sous-ensembles de sujets assurant une répartition équilibrée des annotations suivant le type de données produit par sujet

Suite à cet équilibrage par locuteur, les sous-ensembles d'entraînement et de test ont été équilibrés par instance, afin de contenir un nombre similaires de trames annotées « stress » et « no-stress », pour éviter de sur- ou de sous-apprendre l'une des classes. Dans notre étude, le stress est globalement surreprésenté dans tous les corpus.

Détection Automatique du Stress

A des fins de comparaison, nous avons également conçu des modèles mélangeant les locuteurs dans les sous-ensembles d'entraînement et de test pour un apprentissage en cross-validation, ce qui implique que les modèles apprennent en partie les locuteurs sur lesquels ils sont ensuite testés. Comme pour les modèles indépendants du locuteur (IL) nous avons appliqué un découpage des corpus en trois sous-ensembles pour les modèles dépendants du locuteur (DL).

Dans cette étude, nous avons choisi d'utiliser les SVM comme algorithme pour l'apprentissage de modèles de reconnaissance automatique du stress. L'utilisation de ces modèles a été validée pour la reconnaissance des états affectifs [55] [79] [274] et du stress [336], validant leur efficacité pour la séparation d'espaces de dimensionnalité élevée, Nous avons utilisé l'implémentation des SVM proposée par Fan, Chen et Lin dans LibSVM [98], avec un noyau RBF : ce type de noyau permet d'établir des relations non linéaires entre les paramètres et les classes (contrairement à un noyau linéaire) tout en utilisant moins de paramètres qu'un noyau polynômial [137]. Ce type de noyaux est utilisé pour la reconnaissance d'états affectifs dans l'audio [58] comme dans la vidéo [206].

Nous avons entraîné des modèles de reconnaissance du stress pour chaque corpus (sujets phobiques et témoins en interaction avec une machine : ETCC-HM ; sujets phobiques et témoins en interaction avec des humains : ETCC-HH ; sujets non pathologiques en prise de parole en public : CP2), pour chaque type d'annotation (auto-annotation, annotation par des observateurs externes), pour chaque modalité (audio, vidéo du visage, capture de posture), pour les différentes largeurs de fenêtre considérées (1s, 2s, 5s, ces durées ont été choisies suite à des expériences précédentes les indiquant comme favorable pour la détection d'états affectifs [278] et du stress[III]).

Nous comparons les résultats obtenus par ces différents modèles dans la suite de ce chapitre. Le Tableau 28 récapitule le nombre de segments de données par corpus. Les différences entre modalités audio et vidéo sont dues à certaines captures vidéo non exploitables pour le corpus ETCC-HM, et à l'inclusion de segments vidéo où les sujets ne parlent pas pour le corpus CP2.

corpus	ETCC-HH			ETCC-HM			CP2		
	1s	2s	5s	1s	2s	5s	1s	2s	5s
audio	30549	15004	5714	9872	4866	1856	13025	6161	1783
vidéo				8040	4003	1580	19644	10052	4213

Tableau 28 : Nombre de segments de 1s,2s et 5s utilisés pour construire les modèles, par corpus

Détection Automatique du Stress

Afin d'évaluer et de comparer les performances des différents modèles, nous utilisons la f-mesure moyenne (pondérée par le nombre d'instances de stress et de non-stress) obtenue par chacun des modèles. Dans le cas des modèles entraînés indépendamment du locuteur, les résultats présentés sont le score de f-mesure moyen obtenu sur les 3 sous-ensembles ; la déviation standard des scores de f-mesure sur les 3 sous-ensembles est également présentée pour illustrer la variabilité de performance intra-modèle.

L'ensemble de paramètres audio que nous utilisons est le set IS_emo09 du Challenge Interspeech 2009 sur la reconnaissance des émotions [270] ; il comprend 384 paramètres (présentés en détails en 6.1.2), ce qui rend l'espace à diviser par l'algorithme SVM complexe. Afin de simplifier cet espace, nous avons procédé à une sélection automatique de paramètres. Nous avons pour cela utilisé l'implémentation InfoGain de Weka [116] pour déterminer l'apport de chaque paramètre à la performance de classification. Nous aurions pu utiliser d'autres approches de sélection de paramètres présentées dans le Chapitre 3, mais la sélection par gain est à la fois rapide, et informative pour déterminer les paramètres les plus utiles pour la tâche de reconnaissance considérée.

Une première évaluation a montré qu'en moyenne les 100 meilleurs paramètres étaient responsables de 99% de la performance. A partir de ce résultat, nous avons décidé de comparer les performances de 3 sous-ensembles de paramètres, respectivement composés des 25 meilleurs, 50 meilleurs et 100 meilleurs paramètres sélectionnés selon la méthode InfoGain. Les paramètres sélectionnés sont présentés dans le Tableau 29. Les paramètres systématiquement retenus pour la classification du stress sont tous liés à l'énergie, quelque soit le nombre de paramètres sélectionnés. Les travaux existants sur un vocabulaire non contraint [333] [334] ont valorisé des paramètres dérivés du Teager Energy Operator pour la reconnaissance du stress dans la voix.

Nous avons également testé l'ensemble intégral des 384 paramètres du set IS_emo09, à des fins de comparaison. La suite du Chapitre 7 compare systématiquement les performances obtenues avec ces 4 ensembles de paramètres.

Détection Automatique du Stress

Nombre de paramètres sélectionnés parmi 384	Nombre de paramètres robustes aux variations inter-contextuelles	Paramètres robustes sélectionnés
25	6	Energie : moyenne, déviation standard, décalage MSE (Mean Square Error) Δ Energie : déviation standard, décalage MSE (Mean Square Error)
50	14	Energie : moyenne, déviation standard, décalage MSE (Mean Square Error), max, étendue, décalage Δ Energie : déviation standard, décalage MSE (Mean Square Error), min, étendue Probabilité de Voix : moyenne MFCC[11] : déviation standard Δ MFCC[2] : moyenne, décalage MSE (Mean Square Error)
100	41	Energie : moyenne, déviation standard, décalage MSE (Mean Square Error), max, étendue, décalage Δ Energie : déviation standard, décalage MSE (Mean Square Error), min, max, étendue Probabilité de Voix : moyenne Δ Probabilité de Voix : déviation standard, décalage MSE (Mean Square Error) Δ F0 : déviation standard, aplatissement, décalage MSE (Mean Square Error) MFCC[1] : max MFCC[2] & MFCC[11] : déviation standard, étendue, décalage MSE (Mean Square Error) MFCC[5], MFCC[7] & MFCC[11] : déviation standard, décalage MSE (Mean Square Error) MFCC[3], MFCC[6], MFCC[9] & MFCC[10] : déviation standard Δ MFCC[1] : min, déviation standard, étendue, décalage MSE (Mean Square Error) Δ MFCC[2] : déviation standard, étendue, décalage MSE (Mean Square Error)

Tableau 29 : Sélection des 25 meilleurs, 50 meilleurs et 100 meilleurs paramètres robustes aux variations interindividuelles et inter-contextuelles pour la reconnaissance du stress.

7.2 Audio

7.2.1 Voix de sujets phobiques et non-phobiques en interaction avec des humains

Au cours du projet E-Thérapies, portant sur la reconnaissance du stress chez des sociophobes, nous avons collecté le corpus ETCC-HH, constitué de données audio de sujets phobiques (22 personnes) et de sujets témoins (18 personnes) en interaction avec des humains (c.f. 4.2.2). Ces données ont été utilisées pour entraîner des modèles pour la reconnaissance du stress, suivant la méthodologie décrite en 7.1. Les scores de f-mesure obtenus par les différents modèles sont présentés dans le Tableau 30, les meilleurs résultats sont indiqués en gras.

	Taille de Fenêtre	1000		2000		5000	
Nombre de Paramètres	Annotation/Dépendance au Locuteur	DL	IL	DL	IL	DL	IL
25	Auto	0,75	0,56 <i>0,04</i>	0,77	0,53 <i>0,04</i>	0,62	0,47 <i>0,04</i>
	Externe	0,80	0,63 <i>0,04</i>	0,81	0,63 <i>0,03</i>	0,82	0,65 <i>0,06</i>
50	Auto	0,77	0,55 <i>0,04</i>	0,78	0,51 <i>0,04</i>	0,62	0,46 <i>0,04</i>
	Externe	0,80	0,56 <i>0,05</i>	0,79	0,53 <i>0,05</i>	0,81	0,57 <i>0,05</i>
100	Auto	0,79	0,52 <i>0,04</i>	0,77	0,49 <i>0,04</i>	0,61	0,44 <i>0,04</i>
	Externe	0,79	0,54 <i>0,05</i>	0,77	0,49 <i>0,05</i>	0,76	0,41 <i>0,07</i>
384	Auto	0,75	0,40 <i>0,05</i>	0,75	0,40 <i>0,05</i>	0,59	0,40 <i>0,05</i>
	Externe	0,74	0,31 <i>0,02</i>	0,73	0,31 <i>0,03</i>	0,73	0,35 <i>0,04</i>

Tableau 30 : Comparaison des scores de f-mesure pour les modèles entraînés à partir des données audio du corpus ETCC-HH (22 phobiques / 18 témoins en interaction avec une machine). DL = modèle dépendant du locuteur ; IL = modèle indépendant du locuteur ; Auto = stress auto-annoté par les sujets ; Externe = stress annoté par des observateurs. Scores moyens sur 3 sous-ensembles de données (déviations standard en italique pour les modèles IL). Les meilleurs scores sont en gras

Détection Automatique du Stress

Variabilité suivant les locuteurs

Nous constatons que les modèles dépendants du locuteur (DL) obtiennent des performances systématiquement plus élevées que les modèles indépendants du locuteur (IL), et ce quelle que soit la configuration des paramètres et annotation considérée. Cette différence est expliquée par le processus d'entraînement des modèles : en DL les locuteurs sont présents à la fois dans l'ensemble d'entraînement et de test des modèles, ce qui signifie que l'algorithme apprend certaines caractéristiques des locuteurs. L'entraînement de modèles IL évite précisément cet apprentissage, l'objectif étant de construire des modèles plus résistants aux variabilités des locuteurs en identifiant des indices du stress trans-individus. Les performances plus faibles des modèles IL peuvent indiquer que l'expression du stress est variable d'un individu à l'autre, et que l'apprentissage d'exemple de stress des sujets d'étude favorise leur reconnaissance.

Ce corpus mélange des sujets phobiques et non phobiques dans les mêmes exercices d'interaction avec des humains. Les scores présentés dans le Tableau 30 portent sur cette population mélangée. Si nous observons les performances des modèles utilisant des fenêtres de 5s¹⁶ en distinguant les sujets phobiques et non phobiques, les performances des modèles IL auto-annotés et annotés extérieurement sont plus élevées pour les sujets témoins (f-mesure moyenne 0,70) que pour les sujets phobiques (f-mesure moyenne 0,49). La phobie sociale est un exemple de caractéristique individuelle influençant la reconnaissance du stress, car il présente des caractéristiques atypiques chez les patients qui le rendent plus difficile à distinguer.

Influence de l'annotation

Nous avons vu au Chapitre 5 que le stress était annoté différemment lorsqu'il était auto-évalué par les sujets et lorsqu'il était annoté par des observateurs externes. Les deux formes d'annotations étant faiblement à modérément consensuelles, nous comparons les résultats de modèles utilisant l'une ou l'autre de ces évaluations.

Les modèles IL auto-annotés les plus performants atteignent des scores de f-mesure légèrement supérieurs au hasard ($0,56 \pm 0,04$). Nous constatons également une variabilité des résultats suivant les configurations appliquées : les modèles utilisant l'intégralité des 384 paramètres audio du set du Challenge Interspeech 2009 pour la reconnaissance des émotions (emo_IS09) obtiennent les moins bonnes performances ($0,4 \pm 0,04$), et les modèles utilisant des paramètres extraits de plus larges fenêtres d'analyse sont moins performants que les modèles utilisant des paramètres extraits de plus petites fenêtres.

¹⁶ Nous utilisons les plus larges fenêtres pour distinguer les populations, suivant l'hypothèse que la phobie est une influence constante sur le stress des sujets quelque soit l'échelle d'observation des paramètres

Détection Automatique du Stress

Les modèles annotés extérieurement sont globalement plus performants que ceux obtenus par les modèles auto-annotés. Les modèles IL annotés extérieurement les plus performants atteignent des scores de f-mesure supérieurs au hasard 0,65 ($\pm 0,06$). Nous constatons également une variabilité des résultats suivant les configurations appliquées : les modèles utilisant l'intégralité des 384 paramètres audio du set emo_IS09 obtiennent les moins bonnes performances ($0,31 \pm 0,02$). La variabilité des performances suivant la taille de fenêtre d'analyse utilisée est négligeable, à nombre de paramètres égaux.

La Figure 42 illustre les configurations de nombre de paramètres audio et de taille de fenêtres d'analyse obtenant les meilleures performances pour les modèles DL et IL auto-annotés et annotés par des observateurs externes.

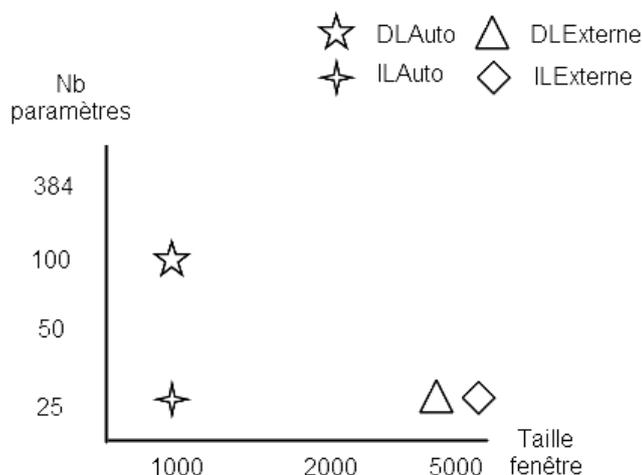


Figure 42 : Meilleures configurations pour les modèles entraînés sur les données audio du corpus ETCC-HH (22 phobiques / 18 témoins en interaction avec des humains). DL = modèle dépendant du locuteur ; IL = modèle indépendant du locuteur ; Auto = stress auto-annoté par les sujets ; Externe = stress annoté par des observateurs.

Les résultats supérieurs des modèles annotés extérieurement peuvent s'expliquer par l'annotation fortement subjective du stress par une population souffrant de troubles anxieux : une estimation exagérée du stress ne se traduira pas forcément par des signes distinctifs. Sur ce corpus, les annotations consensuelles entre auto-évaluation du stress et évaluation par les observateurs externe représentent 62% des données. Ces annotations consensuelles sont équitablement réparties entre populations phobique et témoin.

Détection Automatique du Stress

Nous avons évalué les performances des modèles IL uniquement sur ces données consensuelles, en émettant l'hypothèse qu'elles sont plus représentatives du stress et du non-stress, car moins influencées par l'évaluation subjective des sujets. Comparativement aux modèles entraînés sur l'intégralité des données, les performances en f-mesure sont de 0,63 ($\pm 0,02$), systématiquement supérieures au hasard et meilleures que les performances des modèles auto-annotés (meilleur f-mesure 0,56 $\pm 0,04$). Ce résultat confirme la subjectivité de l'auto-évaluation du stress, que ce soit par des sujets phobiques ou des témoins non pathologiques, et la difficulté pour un système utilisant des indices perceptible du stress extériorisés par les sujets de reconnaître leur ressenti interne.

Évaluation à la tâche

Les scores présentés dans le Tableau 30 sont une évaluation des performances des modèles par segment, c'est-à-dire à l'échelle des fenêtres d'analyse de 1s, 2s ou 5s. Les données ont été annotées à l'échelle de la tâche (i.e. les quatre exercices : lecture de texte à voix haute, avis sur le texte, question incongrue, et description positive de soi), d'une durée moyenne de 1m44s (déviation standard de 25s) sur cette expérimentation.

Afin d'évaluer les performances des modèles à l'échelle de la tâche, nous avons calculé le pourcentage de segments correctement identifiés par tâche, c'est-à-dire catégorisés comme stress pour une tâche annotée comme stress, et réciproquement pour le non-stress. En considérant qu'une tâche ayant plus de 50% de segments correctement catégorisés est correctement reconnue par les modèles, les performances en f-mesure des modèles sont comparables à celles des modèles sur les deux types d'annotations, dans les marges de la déviation standard observée.

Sur des tâches courtes (de l'ordre de la minute), l'évaluation du stress à la tâche ou au segment donne des résultats comparables, suggérant un stress assez constant au court d'une tâche. Les erreurs des modèles sont liées aux variabilités inter-locuteurs plutôt qu'à des erreurs commise au sein d'une même tâche, par exemple reconnaître du stress sur les trente premières secondes puis du non-stress sur la fin de la tâche.

7.2.2 Voix de sujets phobiques et non-phobiques en interaction avec une machine

Nous avons collecté un second corpus au cours du projet E-Thérapies. Le corpus ETCC-HM est constitué de données audio et vidéo de sujets phobiques (15 personnes) et de sujets témoins (9 personnes) en interaction avec une machine simulant des situations anxiogènes (c.f. 4.2.2). Ici, les données audio seules ont été utilisées pour entraîner des modèles de reconnaissance du stress. Les scores de f-mesure obtenus par les différents modèles sont présentés dans le Tableau 31, les meilleurs résultats sont indiqués en gras.

Détection Automatique du Stress

	Taille de Fenêtre	1000		2000		5000	
Nombre de Paramètres	Annotation/Dépendance au Locuteur	DL	IL	DL	IL	DL	IL
25	Auto	0,75	0,54 <i>0,03</i>	0,77	0,58 <i>0,02</i>	0,80	0,54 <i>0,03</i>
	Externe	0,84	0,66 <i>0,08</i>	0,87	0,67 <i>0,09</i>	0,91	0,67 <i>0,09</i>
50	Auto	0,81	0,55 <i>0,06</i>	0,82	0,50 <i>0,07</i>	0,75	0,36 <i>0,09</i>
	Externe	0,87	0,66 <i>0,08</i>	0,88	0,67 <i>0,09</i>	0,89	0,68 <i>0,10</i>
100	Auto	0,77	0,45 <i>0,06</i>	0,75	0,42 <i>0,07</i>	0,75	0,36 <i>0,09</i>
	Externe	0,88	0,67 <i>0,09</i>	0,88	0,67 <i>0,09</i>	0,89	0,68 <i>0,10</i>
384	Auto	0,73	0,29 <i>0,08</i>	0,73	0,28 <i>0,08</i>	0,74	0,32 <i>0,11</i>
	Externe	0,88	0,67 <i>0,09</i>	0,88	0,67 <i>0,09</i>	0,89	0,68 <i>0,10</i>

Tableau 31 : Comparaison des scores de f-mesure pour les modèles entraînés à partir des données audio du corpus ETCC-HM (15 phobiques / 9 témoins en interaction avec une machine). DL = modèle dépendant du locuteur ; IL = modèle indépendant du locuteur ; Auto = stress auto-annoté par les sujets ; Externe = stress annoté par des observateurs. Scores moyens sur 3 sous-ensembles de données (déviations standard en italique pour les modèles IL). Les meilleurs scores sont en gras.

Détection Automatique du Stress

Variabilité suivant les locuteurs

Comme pour le corpus enregistré en interaction avec des humains (c.f. 7.2.1), nous constatons des performances plus élevées pour les modèles dépendant du locuteur (DL) que pour les modèles indépendants du locuteur (IL). Ce corpus présente également un mélange de sujets phobiques et non phobiques évalués simultanément.

Si nous évaluons les deux populations séparément, nous observons un très fort écart de performance : les modèles IL entraînés sur des fenêtres de 5s reconnaissent très bien le stress et le non-stress des sujets phobiques (f-mesure 0,94) à partir des annotations externes, tandis que les performances sur les sujets non pathologiques sont largement inférieures au hasard. Ces différences sont en accord avec celles observées sur l'expérience portant sur les sujets en interaction avec des humains.

L'écart de performance est moins important, et inversé, si nous considérons les auto-évaluations du stress comme référence : le stress des sujets non pathologiques est mieux reconnu (f-mesure 0,66, soit de meilleures performances que les évaluations sans distinction de population), que celui des sociophobes (f-mesure de l'ordre du hasard).

Cette différence entre les deux annotations souligne la subjectivité de la perception du stress, en particulier par une population souffrant de trouble anxieux et ayant tendance à surévaluer son stress. La différence de performance entre les deux populations renforce l'idée d'un stress fortement individuel.

Influence de l'annotation

Comme pour le corpus ETCC-HH, les modèles IL auto-annotés les plus performants atteignent des scores de f-mesure légèrement supérieurs au hasard ($0,58 \pm 0,02$). Nous constatons également une variabilité importante des résultats suivant les configurations, et en particulier des scores largement inférieurs au hasard pour les modèles utilisant l'intégralité des 384 paramètres audio d'IS_emo09. Les meilleures performances sont obtenues avec le plus petit nombre de paramètres audio ; la variabilité suivant la taille de fenêtre d'analyse utilisée n'est pas significative.

Sur ce corpus également, les modèles annotés extérieurement performant globalement mieux que les modèles auto-annotés. Les modèles IL annoté extérieurement les plus performants atteignent des scores de f-mesure supérieurs au hasard 0,68 ($\pm 0,1$). En revanche pour ces modèles la variabilité de performance suivant les configurations est négligeable, pour les modèles DL comme pour les modèles IL.

Les configurations de nombre de paramètre et de tailles de fenêtre d'analyse obtenant les meilleures performances pour les modèles DL et IL auto-annotés et annotés par des observateurs externes sont illustrées dans la Figure 43. Pour les modèles IL et DL annotés extérieurement, les configurations donnent des résultats équivalents et le positionnement est donné à titre de comparaison avec les modèles auto-annotés.

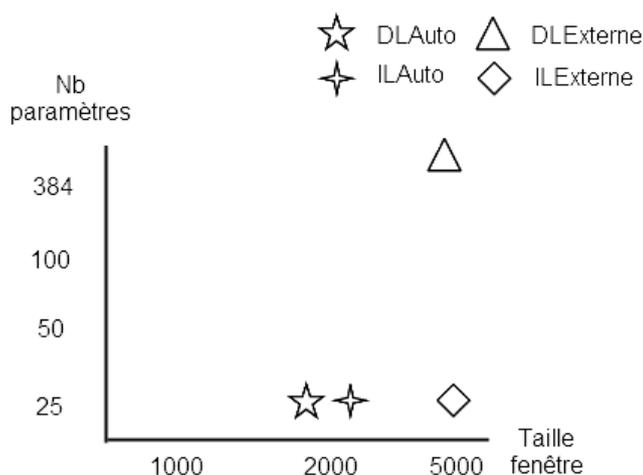


Figure 43 : Meilleures configurations pour les modèles entraînés sur les données audio du corpus ETCC-HM (15 phobiques / 9 témoins en interaction avec une machine) . DL = modèle dépendant du locuteur ; IL = modèle indépendant du locuteur ; Auto = stress auto-annoté par les sujets ; Externe = stress annoté par des observateurs.

Comme pour l'expérimentation en interaction avec des humains, la population cible de cette expérimentation souffre de troubles anxieux, et risque de surévaluer son stress, expliquant les résultats plus faible des modèles auto-annotés. Sur ce corpus, les annotations consensuelles entre auto-évaluation du stress et évaluation par les observateurs externe représentent 56% des données, équitablement réparties entre populations phobique et témoin. Comme pour les données collectées en interaction avec des humains, nous avons évalué les performances des modèles IL en ne tenant compte que des données obtenant des labels consensuels stress ou non-stress dans les deux annotations.

Les performances en f-mesure varient entre 0,70 et 0,72 suivant les tailles de fenêtre d'analyse et les paramètres audio utilisés, améliorant les performances observées pour les modèles auto-annotés comme pour les modèles annotés extérieurement. Comme pour le corpus collecté en interaction avec des humains, ce résultat souligne la subjectivité de l'auto-évaluation du stress.

Évaluation à la tâche

Comme pour l'expérimentation sur les données en interaction avec des humains, ici le stress a été annoté à la tâche (i.e. parler à un public, inviter quelqu'un à dîner, repousser les avances d'un vendeur) d'une durée moyenne de 58s déviation standard 34s), mais évalué au segment de 1s, 2s ou 5s. Afin de mesurer les erreurs intra-tâche, nous avons comme pour le corpus précédent évalué les performances à la tâche, en considérant que si plus de 50% des segments d'une même tâche étaient correctement identifiés, alors le stress de cette tâche était correctement reconnu.

Les performances observées en annotation consensuelle ne sont pas significativement supérieures à celles observées sur les annotations externes et les auto-annotations. Ce résultat conforte celui du corpus enregistré en interaction avec des humains : les différences inter-individus rendent la reconnaissance du stress plus difficile pour les modèles que les variations du stress intra-tâche sur des tâches courtes (de l'ordre de la minute).

7.2.3 Voix de sujets non pathologiques en prise de parole en public

Au cours du projet Compare, portant sur la reconnaissance du stress au cours d'un entretien d'embauche simulé, nous avons collecté le corpus CP2, constitué de données audio et vidéo 3D de sujets non-pathologiques (19 personnes) en prise de parole en public (c.f. 4.3).

Une première étude [III] sur ces données annotées sur le contenu stressé et non-stressé par tour de parole (c.f. 5.3.4) a fait l'objet d'une publication lors de la conférence IEEE ACII (Affective Computing and Intelligent Interaction) en 2013.

L'objectif de cette précédente étude était d'observer l'influence de certaines classes de paramètres audio et de la taille des fenêtres d'analyse sur la reconnaissance du stress. Les paramètres audio utilisés sont issus de travaux sur la reconnaissance des émotions effectués au LIMS I [299] et sont récapitulés dans le Tableau 32.

Détection Automatique du Stress

Descripteurs calculés par trame (30)	Fonctionnelles statistiques (80)
MFCC 1-13	Moyennes MFCC; moyennes MFCC voisées; moyennes MFCC non-voisées; moyennes Δ MFCC (13x4 = 52)
Qualité vocale	Rolloff_5 médian; rolloff_95 médian; nombre d'harmoniques; barycentre médian (4)
Energie	étendue; déviation standard; fréquence des pics; shimmer médian; shimmer maximal; déviation standard du shimmer (6)
Ratio Voix/Silence	Durée moyenne des parties voisées; ratio voisé/non voisé; déviation standard de la durée des parties voisées; durée des silences; fréquence des silences; (5)
F0	médiane; étendue; ratio; déviation standard; jitter médian; jitter maximal; déviation standard du jitter; inter médiane; inter max; inter déviation standard; intra médiane; intra max; intra déviation standard; (13)
LPC 1-12	-
ZCR	-

Tableau 32 : Paramètres audio utilisés pour la reconnaissance du stress sur le copurs de sujets en prise de parole en public dans l'étude [III]

Dans cette précédente étude, nous avons comparé les performances de 54 modèles SVM entraînés de façon dépendante du locuteur, en utilisant des fenêtres d'analyse de 0,5s à 5s et en entraînant des modèles distinct par familles de paramètres : paramètres liés à la fréquence fondamentale (F0), paramètres liés au spectre et cepstre (MFCC et LPC), paramètres liés à l'énergie, paramètres liés au ratio voix/silence, et paramètres liés au jitter et shimmer. La meilleure performance (f-mesure 0,73) a été obtenue en utilisant l'intégralité des 110 paramètres audio sur des fenêtres d'analyse de 5000ms. Cette précédente étude a également mis en valeur des comportements variables selon les classes de paramètres et les tailles de fenêtre, illustrés en Figure 44.

Détection Automatique du Stress

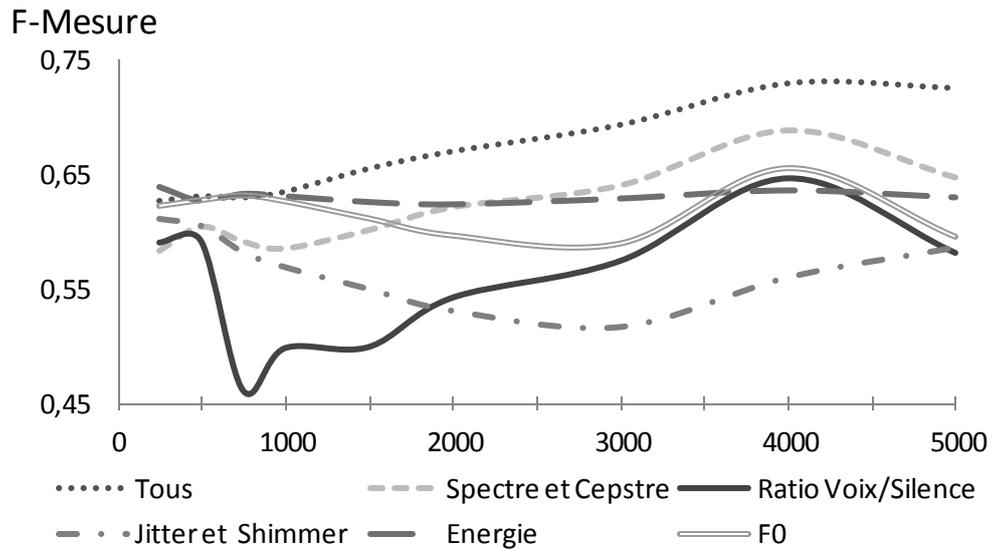


Figure 44 : Scores de f-mesure pour les modèles SVM audio dépendants du locuteur sur les annotations par tour de parole du stress chez les sujets en prise de parole en public, Entraînement en cross-validation 10-folds.

Détection Automatique du Stress

Dans cette thèse, nous utilisons le set de paramètres audio emo_IS09 du Challenge Interspeech2009 [270] présentés dans le Tableau 23, comme pour les corpus issus du projet E-Thérapies. Ici seules les données audio du corpus CP2 ont été utilisées pour entraîner des modèles de reconnaissance du stress. Les scores de f-mesure des différents modèles sont présentés dans le Tableau 33.

Nombre de Paramètres	Taille de Fenêtre	1000		2000		5000	
		DL	IL	DL	IL	DL	IL
25	Auto	0,73	0,74 <i>0,05</i>	0,69	0,70 <i>0,10</i>	0,65	0,66 <i>0,15</i>
	Externe	0,54	0,35 <i>0,10</i>	0,47	0,31 <i>0,09</i>	0,40	0,28 <i>0,09</i>
50	Auto	0,81	0,72 <i>0,07</i>	0,71	0,69 <i>0,12</i>	0,65	0,66 <i>0,16</i>
	Externe	0,46	0,32 <i>0,08</i>	0,45	0,31 <i>0,09</i>	0,40	0,28 <i>0,08</i>
100	Auto	0,84	0,67 <i>0,13</i>	0,66	0,66 <i>0,14</i>	0,64	0,66 <i>0,17</i>
	Externe	0,53	0,32 <i>0,06</i>	0,44	0,43 <i>0,13</i>	0,40	0,28 <i>0,08</i>
384	Auto	0,87	0,66 <i>0,14</i>	0,87	0,66 <i>0,15</i>	0,85	0,66 <i>0,17</i>
	Externe	0,77	0,29 <i>0,14</i>	0,78	0,43 <i>0,13</i>	0,72	0,24 <i>0,05</i>

Tableau 33 : Comparaison des scores de f-mesure pour les modèles entraînés à partir des données audio du corpus CP2 (19 personnes en prise de parole en public). DL = modèle dépendant du locuteur ; IL = modèle indépendant du locuteur ; Auto = stress auto-annoté par les sujets ; Externe = stress annoté par des observateurs. Scores moyens sur 3 sous-ensembles de données (déviations standard en italique pour les modèles IL). Les meilleurs scores sont en gras.

Variabilité des locuteurs

Comme pour les modèles vocaux précédemment évalués, nous constatons globalement de meilleures performances pour les modèles entraînés dépendamment du locuteur (DL) que pour les modèles entraînés indépendamment (IL).

Ce corpus a été enregistré auprès de sujets non sociophobes, mais d'autres facteurs individuels peuvent influencer l'expérience du stress. Nous nous sommes intéressés aux traits de personnalité des participants, en particulier à deux des traits du modèles OCEAN : extraversion et névrotisme, ainsi qu'au trait anxieux (i.e. une tendance générale à ressentir de l'anxiété) et à l'alexithymie, la difficulté à identifier et décrire ses propres émotions. Nous avons séparé les sujets en deux groupes suivant leurs scores aux tests de personnalités (BFI [145], TAS-20 [12] et STAlt [290]) et comparé les performances des modèles suivant les groupes.

Suivant l'annotation considérée, les traits de personnalité observés ont une influence plus ou moins notable sur les performances des modèles. Les modèles entraînés à partir des annotations sont sensibles à l'extraversion, au névrotisme et à l'alexithymie. Sur les sujets présentant ces traits les modèles utilisant des fenêtres de 5s¹⁷ atteignent des performances en f-mesure supérieures à 0,80 (0,88 pour les extravertis, 0,87 pour les névrotiques, et 0,81 pour les alexithymiques), tandis que sur les sujets ne présentant pas ces traits, les performances sont plus faibles (0,69 pour les non alexithymiques, 0,68 pour les non névrotiques, et 0,55 pour les introvertis).

Les modèles entraînés à partir des annotations externes sont surtout sensibles au trait anxieux : sur les sujets anxieux les performances en f-mesure atteignent 0,68, tandis que sur les sujets non anxieux elles sont inférieures au hasard. Les traits névrotiques et alexithymiques montrent également une différence de performance mais les scores restent inférieurs au hasard pour les deux groupes.

Les traits de personnalités observés influencent l'expérience du stress ressenti par les sujets, et son expression perceptible par des observateurs externes. Dans cette expérimentation, les traits influençant la reconnaissance du stress ressenti et du stress perceptible sont différents, soulignant la subjectivité de l'évaluation du stress.

¹⁷ Nous utilisons des fenêtres de 5s, comme pour la phobie nous émettons l'hypothèse que les traits de personnalité ont une influence constante sur le stress des sujets quelque soit l'échelle d'observation des paramètres

Détection Automatique du Stress

Nous nous sommes également intéressés aux stratégies de coping des participants : nous avons distingué les sujets en deux groupes, suivant leur tendance à appliquer des stratégies fonctionnelles (e.g. recadrage positif de la situation, humour) ou dysfonctionnelles (e.g. dénis, désengagement). Quelle que soit l'annotation considérée, les deux groupes obtiennent des performances similaires, supérieures au hasard.

Dans le contexte de cette expérimentation, les stratégies de coping basées sur l'action (e.g. la planification d'une solution au problème stressant, ou le recours à des substances psychotropes comme échappatoire) ne sont pas disponibles pour les sujets sans interrompre l'expérimentation. Seules les stratégies de coping cognitives sont accessibles, et leur influence sur la reconnaissance du stress n'est pas notable.

Influence de l'annotation

Sur ce corpus de sujets non-pathologiques, nous n'observons pas les mêmes tendances que sur les corpus issus du projet E-Thérapies : les modèles entraînés à partir des auto-annotations du stress réalisées par les sujets de l'expérimentation produisent ici de meilleurs résultats que les modèles entraînés à partir d'annotations externes.

Les modèles IL auto-annotés obtiennent des scores systématiquement supérieurs au hasard, et sont plus performants que les modèles équivalents sur les corpus E-Thérapies présentés précédemment. Les résultats présentent cependant une forte variabilité (scores moyens entre 0,66 et 0,74, déviations standard entre 0,05 et 0,17), rendant les différences de performances suivant les configurations de nombre de paramètres et de taille de fenêtre d'analyse non significatives.

Les performances des modèles IL annotés extérieurement sont systématiquement inférieures au hasard. Ce résultat peut s'expliquer par le procédé d'annotation : les observateurs ont utilisé les vidéos des candidats pour mesurer le stress. Or ici nous utilisons uniquement les données audio pour construire un modèle apprenant cette annotation. Les annotateurs ont pu utiliser des informations à la fois audio et visuelles pour leur évaluation perceptive, non reflétées dans les paramètres audio. Nous évaluons l'apport d'informations posturales dans la suite de cette étude.

Les modèles DL auto-annotés et annotés extérieurement obtiennent de meilleures performances en utilisant l'intégralité des 384 paramètres audio du set IS_emo09. Les configurations de nombre de paramètres et de tailles de fenêtres réalisant les meilleures performances pour les différents types de modèles sont illustrées en Figure 45 ; les modèles IL sont placés uniquement à titre de comparaison, du fait de leur fortes déviations standards.

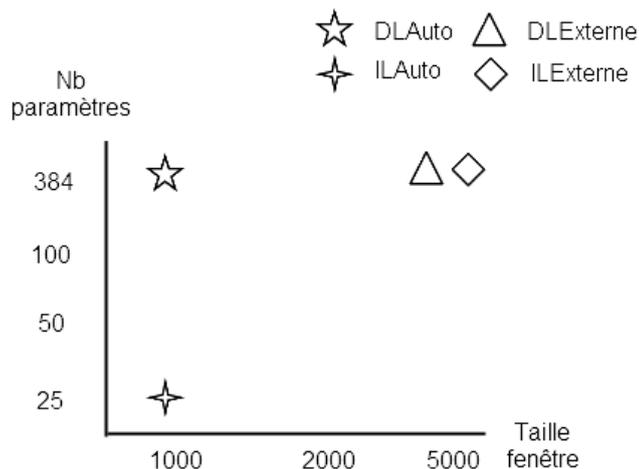


Figure 45 : Meilleures configurations pour les modèles entraînés sur les données audio du corpus CP2 (19 personnes en prise de parole en public) . DL = modèle dépendant du locuteur ; IL = modèle indépendant du locuteur ; Auto = stress auto-annoté par les sujets ; Externe = stress annoté par des observateurs.

Sur ce corpus, l'auto-annotation et l'annotation externe sont plus consensuelles que sur les corpus portant sur des sujets pathologiques. En annotation binaire, utilisée pour entraîner les modèles de reconnaissance automatique du stress, 85% des données sont consensuelles, une portion importante du corpus.

Comme pour les corpus collectés auprès des sociophobes, nous avons évalué les performances des modèles IL uniquement sur ces données consensuelles : les performances en f-mesure varient entre 0,71 et 0,74 suivant la taille de fenêtre d'analyse et les paramètres audio utilisés, soit des résultats comparables à ceux des modèles auto-annotés les plus performants, et fortement supérieurs à ceux des modèles annotés extérieurement. Ce résultat illustre la difficulté de l'évaluation du stress cognitif des sujets par des observateurs externes dans le cas d'un public non pathologique.

Évaluation à la tâche

Les scores présentés dans le Tableau 33 sont une évaluation des performances des modèles par segment, c'est-à-dire à l'échelle des fenêtres d'analyse de 1s, 2s ou 5s. Les données ont été annotées à l'échelle de la tâche (i.e. l'exposé, les questions négatives, et les questions positives ; durée moyenne de 3m15s pour l'exposé et 1m40s pour les phases de question, en temps de parole cumulé) sur cette expérimentation.

Nous avons calculé le pourcentage de segments correctement identifiés par tâche, c'est-à-dire catégorisés comme stress pour une tâche annotée comme stress, et réciproquement pour le non-stress. En considérant qu'une tâche ayant plus de 50% de segments correctement catégorisés est correctement reconnue par les modèles, les performances sont très fortement améliorées.

Détection Automatique du Stress

Les modèles annotés extérieurement obtiennent des performances supérieures au hasard sur les fenêtres de 1s et 2s, alors que l'évaluation par segment donne des scores systématiquement inférieurs au hasard (la meilleur f-mesure était de $0,43 \pm 0,13$). Les performances en f-mesure des modèles auto-annotés sont en moyenne supérieures à 0,85, similaires à celles des modèles DL. Ces performances présentent toujours une forte variabilité (déviations standard moyenne de 0,33), car le stress de certains sujets reste très mal reconnu malgré cette méthode.

Sur cette expérimentation, le stress ne semble pas constant au cours des tâches, et les modèles de reconnaissance peuvent identifier du stress et du non-stress au sein d'une même tâche à différents moments, par exemple entre le début et la fin de l'exposé, ou d'une question à l'autre dans la phase de questions négatives comme dans celle des questions positives (chaque phase comporte trois questions, voir le script des juges présenté en annexe).

La granularité de l'annotation et la granularité de la reconnaissance du stress sont des facteurs influents sur les performances de reconnaissance. Les tâches de cette expérimentation sont plus longues que celles évaluées dans le projet E-Thérapies, particulièrement la phase d'exposé, et présentent plus d'échanges verbaux avec les membres du jury que les enregistrements avec les sociophobes, durant lesquels les expérimentateurs maintiennent un contact minimal. Ces différences contextuelles expliquent la plus forte variabilité des performances (visible dans les déviations standards plus importantes) par rapport aux corpus précédents.

7.3 Vidéo

7.3.1 Expressions faciales de sujets phobiques et non-phobiques en interaction avec une machine

Les vidéos de 13 sujets phobiques et 8 sujets témoins du corpus ETCC-HM, collecté lors du projet E-Thérapies sur la reconnaissance du stress des sociophobes, ont été utilisées afin d'entraîner des modèles de reconnaissance du stress, à partir des expressions faciales des sujets.

Le nombre limité de paramètres faciaux (comparativement aux paramètres audio) ne nous a pas incité à comparer des sous-ensembles de paramètres comme nous l'avons fait pour les paramètres audio. Les résultats présentés dans le Tableau 34 sont les performances en f-mesure de modèles entraînés à partir des 99 paramètres. Nous avons cependant fait varier la taille des fenêtres d'analyse entre 1s, 2s et 5s, comme pour les modèles audio.

Taille de Fenêtre	1000		2000		5000	
Annotation/Dépendance au Locuteur	DL	IL	DL	IL	DL	IL
Auto	0,82	0,44 <i>0,00</i>	0,79	0,45 <i>0,05</i>	0,77	0,38 <i>0,03</i>
Externe	0,87	0,69 <i>0,07</i>	0,88	0,69 <i>0,07</i>	0,90	0,67 <i>0,09</i>

Tableau 34 : Comparaison des scores de f-mesure pour les modèles entraînés à partir des données faciales du corpus ETCC-HM (15 phobiques / 9 témoins en interaction avec une machine). DL = modèle dépendant du locuteur ; IL = modèle indépendant du locuteur ; Auto = stress auto-annoté par les sujets ; Externe = stress annoté par des observateurs. Scores moyens sur 3 sous-ensembles de données (déviation standard en italique pour les modèles IL). Les meilleurs scores sont en gras.

Variabilité suivant les locuteurs

Comme pour les données audio, les modèles entraînés dépendamment du locuteur (DL) atteignent des performances plus élevées que les modèles entraînés indépendamment du locuteur (IL), et ce qu'elle que soit l'annotation considérée.

Comme nous l'avons vu précédemment sur les modèles audio, ce corpus comporte un mélange de sujets sociophobes et de sujets témoins non pathologiques. Dans cette modalité également, la phobie sociale influe fortement sur les performances des modèles de reconnaissance automatique du stress : lorsque nous distinguons par populations les scores des modèles utilisant des fenêtres de 5s, les performances en f-mesure des modèles annotés extérieurement sont favorables aux sujets témoins (0,63, contre un score proche du hasard pour les phobiques) tandis que les modèles auto-annotés favorisent largement les sujets phobiques (0,91, contre un score largement inférieur au hasard pour les témoins).

Influence de l'annotation

Les performances sont variables suivant l'annotation utilisée pour entraîner les modèles. Comme pour les modèles audio, nous constatons des performances supérieures des modèles annotés extérieurement. Les performances des modèles auto-annotés sont par ailleurs systématiquement inférieures au hasard.

Nous pouvons expliquer cette différence par deux facteurs. Tout d'abord, l'auto-annotation réalisée par les phobiques est biaisée par leur pathologie, leur faisant évaluer leur stress comme particulièrement élevé. Nous avons constaté ce biais dans le Chapitre 5 en comparant les annotations des phobiques et les annotations externes : là où un annotateur évalue un stress faible à moyen, les phobiques évaluent un stress moyen à fort. Nous avons également constaté ce biais dans l'évaluation des performances par population, qui montre une meilleure reconnaissance du stress des témoins, dont l'évaluation est plus consensuelle.

Le second facteur est lié au procédé d'annotation : les observateurs ont utilisé les enregistrements audio et vidéo ont pu baser leur évaluation sur des indices à la fois audio et visuels. Les scores obtenus par les modèles faciaux sont d'ailleurs comparables à ceux des modèles audio : le meilleur modèle IL annoté extérieurement entraîné à partir de la voix atteint un score de f-mesure de 0,68 ($\pm 0,10$), et les modèles faciaux atteignent des performances de 0,69 ($\pm 0,07$). Dans la suite de cette étude, nous observons l'utilisation de ces informations combinées.

La différence de performances suivant les tailles de fenêtre d'analyse est négligeable sur les deux types d'annotations, ne valorisant pas une fenêtre en particulier.

Comme pour les modèles audio entraînés à partir de ce corpus, nous avons évalué les performances des modèles IL en ne tenant compte que des données ayant obtenu un label stress ou non-stress consensuel en auto-annotation et en annotation externe. Ces données représentent 56% de ce corpus, avec des proportions équitables de sujets phobiques et de témoins.

Comme pour les données audio, les modèles IL sont plus performants sur les données consensuelles : les scores varient entre 0,66 et 0,72 suivant la fenêtre d'analyse appliquée, soit des scores supérieurs à ceux obtenus par les modèles auto-annotés, et comparables à celles des modèles annotés extérieurement. Ce résultat souligne la subjectivité de l'auto-évaluation du stress, difficilement perceptible dans les données.

Évaluation à la tâche

De même que pour les modèles audio, nous avons évalué les performances des modèles de reconnaissance du stress entraînés à partir d'indices faciaux à la tâche plutôt qu'au segment. Cette évaluation est justifiée par la granularité de l'annotation, réalisée à la tâche (i.e. les différents exercices de TCC* face à la machine : parler à un public, inviter quelqu'un à diner, repousser les avances d'un vendeur) d'une durée moyenne de 58s (déviatoin standard 25s), tandis que les segments évalués durent 1s, 2s ou 5s.

En plaçant le seuil de bonne reconnaissance du stress d'une tâche à la classification correcte de 50% de ses segments, les performances des modèles auto-annotés sont en moyenne de 10 points de pourcentage, soit des scores au dessus du hasard, mais sont fortement variables (déviatoin standard supérieure au gain observée). Les scores des modèles annotés extérieurement sont diminués de 4 points de pourcentage en moyenne, ce qui reste dans les marges des déviatoin standards observées dans l'évaluation par segment.

Cette méthode d'évaluation des performances n'influe pas sur les scores observés de façon significative, de même que pour les modèles audio. Ce résultat conforte notre observation précédente sur les performances audio : sur des tâches courtes au court desquelles le stress varie peu, les modèles sont plus influencés par les différences inter-individus ou inter-tâches que les variations intra-tâches.

7.3.2 Posture de sujets non pathologiques en prise de parole en public

Les enregistrements vidéo de 18 sujets du corpus CP2, collecté lors du projet Compare, portant sur la reconnaissance du stress au cours d'un entretien d'embauche simulé, ont été utilisés pour entraîner des modèles de reconnaissance du stress.

Nous n'utilisons que 15 paramètres pour caractériser la posture des sujets dans ce corpus, une sélection de paramètres comme celle appliquée aux caractéristiques audio présente peu d'intérêts et les résultats présentés dans le Tableau 35 sont les performances en f-mesure de modèles entraînés à partir des 15 paramètres.

Détection Automatique du Stress

Taille de Fenêtre	1000		2000		5000	
Annotation/Dépendance au Locuteur	DL	IL	DL	IL	DL	IL
Auto	0,52	0,39 <i>0,13</i>	0,61	0,60 <i>0,09</i>	0,65	0,44 <i>0,08</i>
Externe	0,60	0,51 <i>0,07</i>	0,62	0,52 <i>0,05</i>	0,62	0,51 <i>0,04</i>

Tableau 35 : Comparaison des scores de f-mesure pour les modèles entraînés à partir des données posturales du corpus CP2 (19 personnes en prise de parole en public). DL = modèle dépendant du locuteur ; IL = modèle indépendant du locuteur ; Auto = stress auto-annoté par les sujets ; Externe = stress annoté par des observateurs. Scores moyens sur 3 sous-ensembles de données (déviations standard en italique pour les modèles IL). Les meilleurs scores sont en gras.

Variabilité suivant les locuteurs

Comme pour les données audio et faciales, les modèles entraînés dépendamment du locuteur (DL) atteignent des performances plus élevées que les modèles entraînés indépendamment du locuteur (IL), et ce quelle que soit l'annotation considérée.

Comme nous l'avons vu sur les données audio, si cette expérimentation ne porte pas sur la phobie, d'autres différences individuelles peuvent influencer sur la reconnaissance du stress des sujets. Nous nous sommes intéressés aux mêmes facteurs que pour les modèles audio dans cette modalité : les traits de personnalité extraversion, névrotisme, anxiété et alexithymie.

Contrairement aux modèles audio, ces traits n'ont pas une influence notable sur les performances de modèles entraînés à partir des annotations des observateurs externes, et une influence non favorable sur les performances des modèles entraînés à partir des auto-annotations car les scores obtenus par les différents groupes sont inférieurs au hasard. Si les traits de personnalités sont un facteur de discrimination intéressant pour la reconnaissance du stress dans l'audio, ils ne le sont pas pour l'utilisation d'indices posturaux.

Le petit nombre d'indices posturaux utilisé (15 paramètres) comparativement au nombre de paramètres audio peut également expliquer ce résultat. Par ailleurs les écarts de performances entre modèles DL et modèles IL annotés extérieurement sont relativement faibles (de l'ordre de 0,10) comparativement aux différences observées avec les modèles audio, expliquant l'absence d'influence notable des traits de personnalité sur les résultats des modèles entraînés à partir des annotations externes.

Détection Automatique du Stress

Influence de l'annotation

Comme pour les résultats observés pour les modèles audio, les performances des modèles annotés extérieurement sont globalement supérieures aux modèles auto-annotés. Les modèles IL auto-annotés atteignent des performances supérieures au hasard uniquement lorsqu'entraînés à partir d'indices extraits d'une fenêtre de 2000ms.

Les performances des modèles entraînés IL extérieurement restent de l'ordre du hasard, et les variations suivant la taille de fenêtre d'analyse sont négligeable.

Les performances des modèles DL annotés extérieurement sont également faiblement supérieures au hasard (f-mesure 0,62) : ce résultat suggère que la posture d'un individu ne suffit pas à indiquer son stress en elle-même, mais plutôt que la posture « stressée » est variable d'un individu à l'autre. Cette hypothèse est cohérente avec la réponse « Fight or Flight » du stress : certaines personnes vont réagir de façon presque violente en prenant des postures plus actives et menaçantes pour combattre la source du stress, tandis que d'autres vont réagir par des postures plus passives et contraintes pour échapper à la source du stress. L'observation des résultats des modèles par individus révèle en effet que le stress de certaines personnes est très bien reconnu dans leur posture, tandis que pour d'autres il est systématiquement mal identifié.

Comme pour les données audio, nous nous sommes intéressés aux performances des modèles IL sur les données ayant obtenus des labels consensuels entre auto-annotation et annotation externe, représentant 85% de ce corpus. Sur les données posturales, cette sélection des données produit des performances similaires à celles des modèles annotés extérieurement, de l'ordre du hasard.

Ce résultat souligne à la fois la difficulté de l'évaluation du stress cognitif des sujets par des observateurs à partir d'indices extériorisés, et les annotateurs utilisent les informations posturales en complément d'autres informations plutôt qu'individuellement pour évaluer le stress.

Évaluation à la tâche

Comme pour les modèles audio conçus à partir de cette expérimentation, nous avons évalué les modèles posturaux à la tâche plutôt qu'au segment, afin de nous rapprocher de la granularité de l'annotation. Sur les modèles audio, cette évaluation a montré une variabilité du stress au cours des tâches, d'une longueur moyenne de 3min15s pour l'exposé et 1m40s pour les phases de question.

En utilisant le même seuil que précédemment, c'est-à-dire en estimant qu'une tâche était correctement reconnue si 50% de ses segments au moins étaient correctement reconnus, les performances des modèles posturaux évalués à la tâche sont inférieures au hasard, qu'elle que soit l'annotation considérée.

Détection Automatique du Stress

Ce résultat, et les performances peu élevées des modèles posturaux de manière générale, peut s'expliquer par le fait que d'autres facteurs que le stress vont influencer la posture des sujets. Sur les phases de question notamment, les sujets présentent des attitudes corporelles différentes s'ils écoutent en silence une question du jury ou s'ils y répondent. La posture peut compléter d'autres informations pour la reconnaissance du stress ; dans la suite de cette étude nous évaluons la fusion des informations audio et posturales de ce corpus.

7.4 Performances des modèles monomodaux

Dans ce chapitre, nous avons présenté les performances de modèles de reconnaissance automatique du stress à partir d'indices extraits d'une seule modalité: audio (indices caractérisant la voix), ou vidéo (indices caractérisant les expressions faciales ou la posture, selon les corpus utilisés).

Nous avons comparé les performances de modèles entraînés à partir des auto-évaluations du stress ressenti des sujets d'expérimentations, et à partir des annotations externes du stress perçu par des observateurs: globalement les scores des modèles utilisant les annotations externes sont plus élevés, à l'exception des modèles audio entraînés sur les données de sujets non-pathologiques en prise de parole en public (corpus CP2). Cette différence peut être le résultat des différences de population ou des différences d'exercices entre les expériences E-Thérapies et Compare. Les paramètres utilisés semblent plus adaptés pour percevoir le stress exprimé par les sujets et perçu par les annotateurs, que le stress ressenti par les sujets et indiqué dans leurs auto-évaluations.

Ces résultats sont à nuancer par les différences interindividuelles à la fois chez les sujets et chez les annotateurs, qui vont influencer l'évaluation du stress ; cette influence est visible dans le stress plus élevé annoté par les phobiques ou dans l'accord entre les annotateurs (c.f. Chapitre 5), mais également par les écarts de performances des modèles évalués sur des sous-populations. Par exemple sur les expérimentations portant sur des sujets sociophobes, nous avons constaté des performances très différentes entre phobiques et témoins. A titre expérimentale nous avons conçu un modèle de reconnaissance du stress à partir des enregistrements des sujets phobiques uniquement, et évalué ce modèle sur les enregistrements des sujets témoins uniquement, en équilibrant les données entre stress et non stress. Les scores sont très inférieurs au hasard, et nous faisons le même constat lorsque les témoins sont utilisés en entraînement et les phobiques en test. Cette expérience suggère une expression du stress différente entre les sujets phobiques et non phobiques, ne favorisant pas un modèle unique de reconnaissance du stress.

Détection Automatique du Stress

Les modèles présentés ont été constitués à partir de paramètres vocaux, faciaux ou posturaux. Les différentes expériences ne révèlent pas une modalité systématiquement plus porteuse d'information, mais plutôt des variations expressives suivant les sujets: certains individus vont favoriser leur voix comme canal d'information pour traduire leur stress, d'autre vont favoriser leurs expressions faciales. Par ailleurs, les comportements vocaux, faciaux et posturaux "stressés" vont également varier d'un individu à l'autre : certaines personnes vont avoir un comportement extrêmement expressif et extérioriser leur état, tandis que d'autres vont au contraire avoir un comportement très fermé et intérioriser leur état.

Les modes de communication affectifs sont également différents d'un individu à l'autre : certaines personnes vont exprimer leur état dans leur voix, par exemple en parlant plus vite ou plus fort sous le coup du stress, tandis que d'autres vont garder une voix similaire mais traduire leur état par des mimiques faciales différentes. A titre expérimental, nous avons étudié les performances des modèles par modalité et par sujet, et constaté que d'une personne à l'autre la modalité obtenant les meilleures performances varie, mais également que l'écart de performances entre les modalités audio et vidéo est plus ou moins important suivant les individus.

Ces résultats portent sur un petit nombre de sujets, mais indiquent une expressivité du stress prenant des formes fortement individuelles, Les performances systématiquement plus élevées des modèles entraînés dépendamment du locuteur montrent que l'apprentissage du comportement spécifique d'un individu favorise la reconnaissance de son stress. Une étude menée sur un nombre comparable de locuteurs enregistrés via un smartphone [187] a montré des résultats concordants : les performances de modèles de mélanges gaussiens de reconnaissances du stress sont meilleures lorsque le modèle est adapté au locuteur.

Les résultats des modèles entraînés à partir de l'annotation du stress par des observateurs externes sont influencés par le processus d'annotation : les annotateurs ont utilisé des informations à audio et visuelles pour estimer le stress des sujets. Une annotation distincte par modalité pourra permettre de mieux distinguer l'expressivité des sujets, et de concevoir des modèles plus précis par modalités.

Nous avons cherché à identifier des indices robustes à cette variabilité interindividuelle : nous avons confronté les sous-ensembles de paramètres audio sélectionnés suivant les différents corpus, afin d'identifier des caractéristiques vocales systématiquement retenues (ou éliminées) pour la classification du stress sur nos données. Nous avons cherché les paramètres systématiquement présents dans les sous-ensembles de paramètres établis sur les différents corpus et les différents sous-ensembles de sujets. Les paramètres consensuels obtenus en sélectionnant les 25 meilleurs, 50 meilleurs et 100 meilleurs sont présentés dans le Tableau 29.

Détection Automatique du Stress

De façon similaire, nous avons établis une liste des paramètres plus instables, en cherchant les paramètres systématiquement éliminés des sous-ensembles de 100 paramètres. La liste complète est présentée dans le Tableau 36, et montre que les paramètres cepstraux (MFCC* et Δ MFCC*) sont presque systématiquement rejetés pour la détection du stress. Ces indices sont fréquemment utilisés pour l'identification du locuteur [185] [214] [269]. Pour des modèles indépendants du locuteur, ces informations n'améliorent pas la reconnaissance du stress.

Classe	Descripteurs statistiques
Energie RMS	max relatif, min relatif, asymétrie, aplatissement
Δ Energie RMS	max relatif, min relatif, asymétrie
F0	min, min relatif
Δ F0	moyenne
ZCR	min, max relatif, min relatif, moyenne, pente, décalage, aplatissement
Δ ZCR	max relatif, min relatif, moyenne, pente, décalage, asymétrie, aplatissement, min
Probabilité de Voix	min, max relatif, min relatif, pente
Δ Probabilité de Voix	max relatif, min relatif, moyenne, pente, décalage, asymétrie
MFCC [1-12]	<i>Variable suivant la bande Mel considérée</i>
Δ MFCC [1-12]	<i>Variable suivant la bande Mel considérée</i>

Tableau 36 : Paramètres audio moins robustes aux variations interindividuelles et inter-contextuelles pour la reconnaissance du stress

Les variations de configuration appliquées (i.e. nombre de paramètres audio sélectionnés, taille des fenêtres d'analyse) ont une influence globalement faible ou non significative et variable d'un corpus à l'autre sur les performances des modèles IL, ne nous permettant pas d'identifier une configuration systématiquement plus favorable que les autres. Cette faible variabilité peut cependant indiquer que des modèles utilisant un petit nombre de paramètres peuvent être utilisés pour la détection du stress, situation favorable pour la conception de logiciels de remédiation : un nombre réduit de paramètres sera plus rapidement calculé et nécessitera moins de ressources.

Chapitre 8 Fusion d'informations multimodales pour la reconnaissance du stress

Sur l'ensemble des données collectées dans cette étude, deux des corpus présentent un aspect multimodal exploitable pour l'analyse automatique d'indices du stress : le corpus ETCC-HM, qui comporte des données audio et faciales enregistrées pendant des exercices de TCC* de sociophobes et de sujets témoins en interaction avec une machine, et le corpus CP2, qui comporte des données audio et posturales enregistrées pendant une simulation d'entretien d'embauche s'apparentant à une prise de parole en public auprès de sujets non pathologiques.

Dans le Chapitre 7, nous avons présenté les performances de modèles SVM entraînés à partir de ces données de façon monomodale. Dans ce chapitre, nous nous intéressons à la performance de modèles SVM entraînés en fusionnant les deux modalités de chacun des corpus. Suite aux résultats obtenus par les modèles monomodaux, nous appliquons la fusion uniquement aux modèles indépendants du locuteur (IL), et n'utilisons que les sous-ensembles de 25, 50 et 100 paramètres audio ; l'intégralité des paramètres faciaux et posturaux est utilisée. Le Tableau 37 récapitule le nombre de segments utilisés pour la fusion ; le nombre de segments fusionné est inférieur aux segments audio et vidéo (c.f. 7.1, Tableau 28) suite à l'alignement des données et à l'élimination de sujets dont la capture vidéo n'est pas exploitable.

corpus	ETCC-HM			CP2		
fenêtre	1s	2s	5s	1s	2s	5s
fusion	7623	3762	1435	11571	4855	1465

Tableau 37 : Nombre de segments de 1s, 2s et 5s utilisés pour la fusion, par corpus

Dans la suite de ce chapitre, nous comparons deux stratégies de fusion : une fusion précoce, et une fusion tardive.

La fusion précoce consiste en l'alignement temporel des indices extraits des différentes modalités, afin d'entraîner un modèle – dans notre étude, un modèle SVM – utilisant l'intégralité des paramètres de façon équipondérée sans distinction quant à leur provenance audio ou visuelle (i.e. faciale ou posturale).

La fusion tardive consiste à exploiter les résultats des modèles monomodaux dans des calculs combinant les décisions de plusieurs modalités pour arriver à une décision multimodale. Il s'agit d'une fusion de haut niveau qui ne prend pas en compte les valeurs des paramètres utilisés pour entraîner les modèles monomodaux mais la catégorie attribuée – dans notre étude, stress ou non-stress – à chaque segment de données évalué.

Détection Automatique du Stress

Cette combinaison peut se faire de différentes façons (c.f. 3.5), nous avons choisi de comparer trois procédés dans cette étude : une fusion par votes simples qui choisit la catégorie ayant obtenu une majorité de votes, une fusion par votes pondérés qui nuance le vote de chaque modalité par un poids de confiance (nous avons utilisé la f-mesure des modèles monomodaux comme pondération), et une fusion par cascade de spécialistes qui utilise une modalité différente – dans notre étude, audio ou visuelle – pour évaluer la probabilité qu'un segment de données appartienne à l'un ou l'autre des catégories. Nous avons utilisé la f-mesure des modèles monomodaux pour choisir le « spécialiste » pour chaque catégorie.

Nous avons choisi d'aligner les informations issues des modalités audio et visuelles suivant la même granularité : les paramètres audio et visuels sont calculés sur des segments de même taille (1s, 2s ou 5s) et combinés en fusion précoce, et les décisions des modèles monomodaux portant sur des segments de même taille (1s, 2s ou 5s) sont combinées en fusion tardive. Nous aurions pu combiner les informations à différentes fréquences temporelles (e.g. les informations visuelles calculées sur 5 secondes combinées aux informations audio calculées toutes les secondes sur cette même période), pour observer si les indices du stress sont perceptibles sur des durées différentes suivant les modalités. Cependant, l'étude des performances monomodales au Chapitre 7 n'a pas valorisé de fenêtre d'analyse explicitement plus efficace, aussi nous n'avons pas approfondi cet alignement.

8.1 Fusion de la voix et des expressions faciales de sujets phobiques et non-phobiques en interaction avec une machine

Les vidéos de 13 sujets phobiques et 8 sujets témoins du corpus ETCC-HM, collecté lors du projet E-Thérapies sur la reconnaissance du stress des sociophobes, ont été utilisées afin d'entraîner des modèles de reconnaissance du stress, à partir des expressions faciales et de la voix des sujets, suivant la méthodologie décrite en 7.1.

Les 99 paramètres faciaux, et les sous-ensembles de 25 meilleurs, 50 meilleurs et 100 meilleurs paramètres audio ont été utilisés pour une reconnaissance bimodale du stress. Ces paramètres sont récapitulés dans le Tableau 38. Les paramètres audio sélectionnés sont les mêmes que pour les modèles monomodaux audio, et varient suivant le sous-ensemble de sujet étudié, aussi il n'est pas possible d'en fournir une liste. Nous avons vu au Chapitre 7 que les paramètres systématiquement sélectionnés sont liés à l'énergie (c.f. 7.4).

	Paramètres audio (25, 50 ou 100 parmi)	Paramètres faciaux (99)
Paramètres par échantillon	<ul style="list-style-type: none"> • Taux de passage par zéro (ZCR), Δ ZCR, • Energie RMS, Δ Energie RMS, • F0 (normalisée à 500Hz), Δ F0, • Probabilité de Voix , Δ Probabilité de Voix • MFCC [1-12], Δ MFCC [1-12] 	<ul style="list-style-type: none"> • Arche sourcils, Δ Arche sourcils • Ecart sourcils, Δ Ecart sourcils • Ecart sourcils - yeux, Δ Ecart sourcils - yeux • Ouverture bouche, Δ Ouverture bouche • Rapport lèvres, Δ Rapport lèvres • Δ largeur et Δ hauteur boîte englobante • Δ abscisse et Δ ordonnée boîte englobante
Fonctionnelles statistiques par segment de 1, 2 ou 5s	<ul style="list-style-type: none"> • moyenne, déviation standard, • asymétrie (skewness), aplatissement (kurtosis), • extrêmes : minimum, maximum, position relative, étendue (range) • régression linéaire : décalage (offset), décalage MSE (Mean Square Error), pente (slope) 	<ul style="list-style-type: none"> • moyenne, déviation standard, • asymétrie (skewness), aplatissement (kurtosis) • extrêmes : minimum, maximum, étendue

Tableau 38 : Paramètres audio et faciaux utilisés pour la reconnaissance bimodale du stress sur le corpus ETCC-HM (13 sujets phobiques et 8 sujets témoins en interaction avec une machine).

Détection Automatique du Stress

Le Tableau 39 présente les scores obtenus par les modèles suivant les différentes configurations de paramètres, de fenêtres d'analyse, et le type de fusion appliqué. Les meilleures performances pour les modèles entraînés à partir des auto-annotations des sujets et pour les modèles entraînés à partir des évaluations externes sont indiquées en gras.

	Fenêtre	1000				2000				5000			
Paramètres	Mode Fusion	P	V	W	C	P	V	W	C	P	V	W	C
25+99	Auto	0,42 <i>0,12</i>	0,30 <i>0,03</i>	0,53 <i>0,04</i>	0,52 <i>0,06</i>	0,42 <i>0,10</i>	0,32 <i>0,05</i>	0,54 <i>0,05</i>	0,55 <i>0,09</i>	0,40 <i>0,06</i>	0,38 <i>0,03</i>	0,54 <i>0,03</i>	0,53 <i>0,05</i>
	Externe	0,68 <i>0,04</i>	0,64 <i>0,10</i>	0,72 <i>0,06</i>	0,69 <i>0,06</i>	0,70 <i>0,02</i>	0,70 <i>0,07</i>	0,74 <i>0,05</i>	0,73 <i>0,06</i>	0,76 <i>0,21</i>	0,74 <i>0,05</i>	0,76 <i>0,05</i>	0,75 <i>0,05</i>
50+99	Auto	0,40 <i>0,10</i>	0,35 <i>0,04</i>	0,56 <i>0,05</i>	0,55 <i>0,02</i>	0,39 <i>0,08</i>	0,41 <i>0,00</i>	0,57 <i>0,03</i>	0,55 <i>0,06</i>	0,38 <i>0,10</i>	0,49 <i>0,02</i>	0,51 <i>0,02</i>	0,51 <i>0,02</i>
	Externe	0,67 <i>0,05</i>	0,65 <i>0,09</i>	0,73 <i>0,05</i>	0,70 <i>0,06</i>	0,69 <i>0,02</i>	0,71 <i>0,06</i>	0,75 <i>0,04</i>	0,74 <i>0,06</i>	0,71 <i>0,08</i>	0,75 <i>0,05</i>	0,77 <i>0,04</i>	0,76 <i>0,05</i>
100+99	Auto	0,36 <i>0,01</i>	0,44 <i>0,03</i>	0,52 <i>0,06</i>	0,53 <i>0,05</i>	0,35 <i>0,01</i>	0,47 <i>0,03</i>	0,54 <i>0,06</i>	0,54 <i>0,06</i>	0,35 <i>0,00</i>	0,49 <i>0,02</i>	0,51 <i>0,02</i>	0,51 <i>0,02</i>
	Externe	0,66 <i>0,07</i>	0,69 <i>0,06</i>	0,73 <i>0,04</i>	0,73 <i>0,04</i>	0,66 <i>0,07</i>	0,71 <i>0,06</i>	0,75 <i>0,04</i>	0,74 <i>0,05</i>	0,67 <i>0,06</i>	0,75 <i>0,05</i>	0,77 <i>0,04</i>	0,76 <i>0,05</i>

Tableau 39 : Comparaison des scores de f-mesures des modèles entraînés à partir des données fusionnées audio et faciales du corpus ETCC-HM (13 phobiques / 8 témoins en interaction avec une machine). P = fusion précoce ; V = fusion par votes simples ; W = fusion par votes pondérés ; C = fusion en cascade; Auto = stress auto-annoté par les sujets ; Externe = stress annoté par des observateurs. Scores moyens sur 3 sous-ensembles de données (déviations standard en italique). Les meilleurs scores sont en gras.

Détection Automatique du Stress

La méthode de fusion utilisée influe fortement sur les performances observées : la fusion tardive par votes simples obtient des scores systématiquement inférieurs au hasard. Ce résultat s'explique par le processus du vote simple : il s'agit d'une décision majoritaire, or nous utilisons un nombre pair de modèles ce qui permet des cas ambigus où l'un des modèles attribue l'étiquette d'une catégorie, et l'autre modèle celle de la catégorie opposée. La décision d'une étiquette définitive est alors impossible. Sur cette expérimentation, la fusion par votes simple a rencontré de 3% à 44% d'ambiguïtés suivant la configuration considérée.

La fusion tardive par votes pondérés permet de lever les ambiguïtés rencontrées par la fusion par votes simples, en attribuant une pondération à chacun des modèles, limitant fortement la probabilité d'une décision impossible. Nous avons utilisé les performances des modèles monomodaux pour pondérer la fusion : les décisions des modèles audio, globalement plus performant que les modèles faciaux, ont plus pesé sur le vote final. Nous constatons que cette méthode s'est avérée globalement la plus efficace sur ces données ($0,54 \pm 0,06$ à $0,57 \pm 0,03$ en auto-annotation, et $0,76 \pm 0,05$ à $0,77 \pm 0,04$ en annotation externe).

La fusion par cascade de spécialistes obtient des scores comparables à ceux de la fusion par votes pondérées, les différences de performances étant inférieures à la déviation standard observée. La fusion en cascade favorise les performances de systèmes où une modalité est plus adaptée qu'une autre pour la reconnaissance d'une catégorie de données. Dans cette expérimentation, la reconnaissance du stress et du non stress par les modèles audio et visuel produit des résultats comparables à des méthodes non discriminantes des performances par catégorie. Ce résultat suggère que les indices du stress et du non stress sont portés par les deux canaux d'informations.

La fusion précoce obtient des performances différentes selon le type d'annotations considéré : pour les modèles annotés extérieurement, la fusion précoce obtient des scores au dessus du hasard pour les modèles utilisant plus de 25 paramètres audio (et toujours 99 paramètres faciaux), et les écarts de performances avec les meilleurs modèles (en fusion pondérée) sont inférieurs à la déviation standard observée. En revanche pour les modèles auto-annotés les performances sont inférieures au hasard.

Qu'elle que soit l'annotation utilisée, les scores obtenus par les méthodes de fusion les plus efficaces sont supérieurs à ceux obtenus par les modèles monomodaux faciaux, et comparables à ceux obtenus par les modèles monomodaux audio. Les variations de performances suivant les paramètres sélectionnés, et suivant la taille de fenêtre d'analyse sont, à méthodes de fusion égales, négligeable.

Détection Automatique du Stress

Variabilité suivant les locuteurs

Nous constatons que comme pour les modèles monomodaux entraînés à partir de la voix ou des expressions faciales des sujets, les modèles utilisant l'annotation externe du stress comme étiquetage des données obtiennent des performances plus élevées que ceux utilisant l'auto-évaluation du stress par les sujets. Ce résultat conforte l'hypothèse d'une évaluation subjective du stress par les sujets, influençant les performances des modèles.

Nous avons comparé les performances des modèles par fusion pondérée sur les sujets phobiques et sur les témoins. Les tendances observées sur les modèles monomodaux sont confirmées en fusion pondérée : les modèles annotés extérieurement reconnaissent mieux le stress des sujets phobiques, et les modèles auto annotés reconnaissent mieux le stress des sujets témoins. La reconnaissance du stress dépend également de facteurs individuels influençant l'expressivité des sujets, en particulier la phobie sociale semble produire des indices forts perçus par les annotateurs et impactant les performances des modèles.

8.2 Fusion de la voix et de la posture de sujets non pathologiques en prise de parole en public

Les enregistrements audio et vidéo de 18 sujets du corpus CP2, collectés lors du projet Compare, portant sur la reconnaissance du stress au cours d'un entretien d'embauche simulé, ont été utilisés pour entraîner des modèles de reconnaissance du stress ressenti, à partir de la posture et de la voix des sujets, suivant la méthodologie décrite en 7.1.

Les 15 paramètres posturaux, et les sous-ensembles de 25 meilleurs, 50 meilleurs et 100 meilleurs paramètres audio ont été utilisés pour une reconnaissance bimodale du stress. Ces paramètres sont récapitulés dans le Tableau 40. Les paramètres audio sélectionnés sont les mêmes que ceux présentés pour les modèles monomodaux audio, et sont variables suivant le sous-ensemble de sujets considéré aussi il n'est pas possible d'en établir une liste ; nous avons vu au Chapitre 7 que les paramètres consensuels sur les différents sous-ensembles sont liés à l'énergie.

	Paramètres audio (25, 50 ou 100 parmi)	Paramètres posturaux (15)
Paramètres par échantillon	<ul style="list-style-type: none"> Taux de passage par zéro (ZCR), Δ ZCR, Energie RMS, Δ Energie RMS, F0 (normalisée à 500Hz), Δ F0, Probabilité de Voix, Δ Probabilité de Voix MFCC [1-12], Δ MFCC [1-12] 	<ul style="list-style-type: none"> Indice de Contraction (CI)
Fonctionnelles statistiques par segment de 1, 2 ou 5s	<ul style="list-style-type: none"> moyenne, déviation standard, asymétrie (skewness), aplatissement (kurtosis), extrêmes : minimum, maximum, position relative, étendue (range) régression linéaire : décalage (offset), décalage MSE (Mean Square Error), pente (slope) 	Sur le CI : <ul style="list-style-type: none"> moyenne, déviation standard, asymétrie (skewness), aplatissement (kurtosis) extrêmes : minimum, maximum, étendue
		Quantité de Mouvement (QoM)

Tableau 40 : Paramètres audio et posturaux utilisés pour la reconnaissance bimodale du stress sur le corpus CP2 (18 sujets en prise de parole en public).

Détection Automatique du Stress

Les résultats de ces modèles fusionnés sont présentés dans le Tableau 41, suivant les différentes configurations de paramètres, de fenêtres d'analyse, et le type de fusion appliqué. Les meilleures performances pour les modèles entraînés à partir des auto-annotations des sujets et pour les modèles entraînés à partir des évaluations externes sont indiquées en gras.

	Fenêtre	1000				2000				5000			
Paramètres	Mode Fusion	P	V	W	C	P	V	W	C	P	V	W	C
25+15	Auto	0,68 <i>0,10</i>	0,24 <i>0,10</i>	0,74 <i>0,03</i>	0,37 <i>0,14</i>	0,67 <i>0,11</i>	0,59 <i>0,28</i>	0,79 <i>0,07</i>	0,61 <i>0,27</i>	0,65 <i>0,13</i>	0,21 <i>0,11</i>	0,74 <i>0,09</i>	0,34 <i>0,07</i>
	Externe	0,48 <i>0,15</i>	0,15 <i>0,05</i>	0,48 <i>0,01</i>	0,48 <i>0,04</i>	0,44 <i>0,13</i>	0,16 <i>0,03</i>	0,49 <i>0,01</i>	0,49 <i>0,02</i>	0,29 <i>0,10</i>	0,23 <i>0,11</i>	0,53 <i>0,02</i>	0,54 <i>0,03</i>
50+15	Auto	0,69 <i>0,09</i>	0,23 <i>0,10</i>	0,74 <i>0,05</i>	0,36 <i>0,13</i>	0,67 <i>0,10</i>	0,60 <i>0,27</i>	0,80 <i>0,09</i>	0,62 <i>0,26</i>	0,65 <i>0,13</i>	0,20 <i>0,12</i>	0,75 <i>0,10</i>	0,34 <i>0,08</i>
	Externe	0,45 <i>0,16</i>	0,13 <i>0,05</i>	0,48 <i>0,01</i>	0,47 <i>0,04</i>	0,43 <i>0,14</i>	0,16 <i>0,03</i>	0,49 <i>0,01</i>	0,49 <i>0,02</i>	0,29 <i>0,10</i>	0,22 <i>0,10</i>	0,53 <i>0,02</i>	0,54 <i>0,03</i>
100+15	Auto	0,66 <i>0,12</i>	0,22 <i>0,09</i>	0,75 <i>0,08</i>	0,36 <i>0,12</i>	0,66 <i>0,12</i>	0,61 <i>0,27</i>	0,82 <i>0,12</i>	0,62 <i>0,25</i>	0,65 <i>0,13</i>	0,20 <i>0,12</i>	0,76 <i>0,10</i>	0,35 <i>0,07</i>
	Externe	0,43 <i>0,14</i>	0,14 <i>0,05</i>	0,48 <i>0,01</i>	0,48 <i>0,03</i>	0,43 <i>0,13</i>	0,21 <i>0,11</i>	0,56 <i>0,12</i>	0,49 <i>0,01</i>	0,29 <i>0,10</i>	0,23 <i>0,10</i>	0,53 <i>0,02</i>	0,54 <i>0,03</i>

Tableau 41 : Comparaison des scores de f-mesures des modèles entraînés à partir des données fusionnées audio et posturales du corpus CP2 (18 personnes en prise de parole en public). P = fusion précoce ; V = fusion par votes simples ; W = fusion par votes pondérés ; C = fusion en cascade ; Auto = stress auto-annoté par les sujets ; Externe = stress annoté par des observateurs. Scores moyens sur 3 sous-ensembles de données (déviations standard en italique). Les meilleurs scores sont en gras.

Nous avons appliqué les quatre méthodes de fusion présentées précédemment (fusion précoce, fusion tardive par votes simples, fusion tardive par votes pondérés, et fusion tardive en cascade) sur les données audio et posturale collectées dans cette expérimentation.

Détection Automatique du Stress

Les tendances observées sur les performances des modèles entraînés sur les données collectés auprès de populations phobiques sont confirmées sur ce corpus de sujets non pathologiques. La fusion par votes simples produits des scores systématiquement inférieurs au hasard. Ce résultat est ici aussi lié à l'utilisation d'un nombre pair de modèles en entrée de la fusion, ce qui rend possible des résultats ambigus : sur ces données, le modèle a rencontré de 55% à 70% de segments ambigus suivant les configurations de paramètres et de tailles de fenêtres d'analyse utilisées. Les informations portées par les modèles audio et posturaux sont fortement contradictoires prises sans tenir compte de la fiabilité de ces modèles.

La fusion par votes pondérés résout le problème des ambiguïtés de la fusion par votes simples, en pondérant les résultats de chacun des modèles monomodaux (modèle audio ou modèle postural) par leurs performances individuelles. Les meilleures performances sont globalement atteintes en fusion pondérée ($0,79 \pm 0,07$ à $0,82 \pm 0,12$ en auto-annotation, $0,53 \pm 0,02$ à $0,56 \pm 0,12$ en annotation externe). Ce résultat, en comparaison aux performances des modèles conçus en fusion par votes simples, indique que les informations apportées par les modèles audio et par les modèles posturaux ne sont pas équivalentes.

La fusion en cascade produit des résultats relativement faibles pour les modèles auto-annotés. Les performances de cette méthode sont liées aux performances des différents modèles fusionnés sur les différentes catégories de données classifiées. Si les performances du modèles spécialiste choisi pour la catégorie de donnée la moins bien reconnue sont faibles, les performances globales vont être impactées négativement. Sur ces données, le modèle postural est choisi pour la reconnaissance du non-stress, tandis que le modèle audio est choisi pour la reconnaissance du stress. Or nous avons vu au Chapitre 7 que les modèles posturaux auto-annotés (en moyenne pondérée sur les deux classes) étaient faiblement performants comparativement aux modèles audio, obtenant des résultats de l'ordre du hasard. Ce déséquilibre entre les deux modalités explique les mauvais résultats en cascade. Sur les modèles annotés extérieurement, les performances des modèles monomodaux audio et posturaux sont comparables, et les modèles fusionnés en cascade performant de façon comparable aux meilleurs scores obtenus en fusion pondérée.

Les écarts de performances entre les modèles auto-annotés et les modèles annotés extérieurement que nous avons observés sur les modèles audio se répercutent sur les modèles fusionnés. Les modèles entraînés à partir des auto-évaluations des sujets obtiennent des résultats élevés, surpassant ceux des modèles monomodaux pour les modèles conçus en fusion pondérée.

Détection Automatique du Stress

En revanche les modèles entraînés à partir des annotations externes obtiennent des scores faibles, de l'ordre du hasard pour les meilleures performances en fusion par votes pondérés et en fusion en cascade. La variabilité des résultats, visible dans les déviations standards, est moins importante que celle des résultats monomodaux. La fusion d'informations audio et posturales permet une reconnaissance du stress plus uniforme sur l'ensemble des sujets, comparativement à la reconnaissance monomodale.

Ces résultats suggèrent que suivant les individus observés, le stress est exprimé différemment et plus perceptible dans la voix ou dans la posture. Lors de l'étude des performances des modèles monomodaux, nous avons observé des résultats variables d'un sujet à l'autre supportant cette hypothèse : pour certains le stress était mieux reconnu par les modèles audio, et pour d'autres mieux reconnu par les modèles posturaux. L'utilisation de ces informations combinées permet d'obtenir des résultats plus homogènes à l'échelle du corpus, mais peut également dégrader les résultats pour un sujet donné : pour un individu dont le stress est mieux reconnu à travers sa posture, la fusion des données posturales et des données audio va introduire des erreurs de reconnaissance.

Variabilité suivant les locuteurs

Nous avons évalué les performances des modèles en fusion pondérée suivant les traits de personnalité des sujets (extraversion, névrotisme, alexithymie, anxiété), et suivant les stratégies de coping favorisées (stratégies fonctionnelles comme la relativisation d'une situation stressante, ou dysfonctionnelles comme le déni de la situation).

Avec les modèles monomodaux, le stress des sujets distingués par leurs stratégies de coping est reconnu de façon similaire. Avec les modèles en fusion pondérée, les modèles auto-annotés ne montrent pas de différence, mais les modèles entraînés à partir de l'annotation externe du stress obtiennent des performances plus élevées et au dessus du hasard sur les sujets favorisant des stratégies dysfonctionnelles.

L'influence des traits de personnalités observée sur les performances des modèles monomodaux audio se retrouve dans les performances des modèles conçus en fusion pondérée : les sujets extravertis, névrotiques et alexithymiques sont mieux reconnus que leurs contreparties par les modèles entraînés à partir des auto-évaluations, tandis que les modèles annotés extérieurement reconnaissent mieux le stress des sujets non anxieux que des sujets anxieux.

Ces différences de performance interindividuelles, liées à la fois aux caractéristiques psychologiques des individus, et à leurs caractéristiques expressives, illustrent l'aspect fortement individuel du stress.

8.3 Apports de la fusion

Dans ce chapitre, nous avons présenté les performances de modèles de reconnaissance automatique du stress à partir de la fusion d'informations audio et faciales, sur des sujets phobiques et témoins en interaction avec un machine pour des exercices de TCC*, et d'informations audio et posturales, sur des sujets non pathologiques en prise de parole en public pour un entretien d'embauche simulé.

Nous avons comparé les résultats de fusion précoce, tardive par votes simples, tardive par votes pondérés et tardive en cascade de spécialistes. L'utilisation d'un nombre pair de modalités rend la fusion par vote simple inopérante, du fait de nombreuses décisions ambiguës. Les autres modes de fusions atteignent des performances comparables ou supérieures à celles des modèles monomodaux sur les mêmes sonnées, en particulier la fusion par votes pondérés qui lève les ambiguïtés rencontrées par la fusion par votes simples.

Les variations de configuration appliquées (i.e. nombre de paramètres audio sélectionnés, taille des fenêtres d'analyse) ont une influence faible sur les performances des modèles fusionnés, qui se révèlent plus résistant que les modèles monomodaux à ces variations.

Comme pour les modèles monomodaux, nous avons comparé les performances de modèles entraînés à partir des auto-évaluations du stress des sujets d'expérimentations, et à partir des annotations externes. Les résultats divergent selon les corpus, reflétant les résultats des modèles monomodaux. Ces disparités illustrent les comportements différents des sujets phobiques et des sujets non-pathologiques, dont l'expérience et l'expression du stress sont fortement dissemblables, et souligne l'importance des facteurs individuels sur la reconnaissance du stress.

L'utilisation d'information provenant de plusieurs modalités pour la reconnaissance du stress homogénéise les performances à l'échelle du corpus par rapport aux modèles monomodaux. Ce résultat est lié aux différences d'expressivités d'un individu à l'autre : d'une personne à l'autre, le stress est mieux reconnu par les modèles audio ou par les modèles visuels, avec des écarts de performances variables entre les deux types de modèle. Ces différences interindividuelles sont concordantes avec les études existantes sur la fusion d'informations multimodale pour la reconnaissance des états affectifs [319]. En utilisant les informations des deux modalités, les modèles fusionnés reconnaissent le stress consensuellement identifié dans l'audio et dans les informations faciales ou posturales, minimisant l'impact des différences d'expressivité interindividuelles.

Détection Automatique du Stress

Sélection de paramètres en fusion

Nous avons cherché à identifier des paramètres robustes à ces différences interindividuelles, en appliquant une sélection de paramètres suivant le gain de performances apporté par chaque paramètre à la reconnaissance du stress pour les modèles conçus en fusion précoce. Etant donné les différents sets de paramètres sur les données enregistrées auprès des sociophobes, et auprès des sujets en prise de parole en public, nous avons étudié la sélection des paramètres séparément sur les deux corpus.

Le Tableau 42 présente les paramètres sélectionnés sur les données audio et faciale des sujets phobiques et témoins de l'expérimentation ETCC-HM. Nous avons effectué la sélection des 100 meilleurs paramètres à partir des combinaisons de 100 paramètres audio et 99 paramètres faciaux, soit 50% des paramètres. Les paramètres présentés sont ceux systématiquement sélectionnés quels que soit la taille de fenêtre d'analyse appliquée et le sous-ensemble de sujets étudié. Nous avons distingué les paramètres favorisant les modèles auto-annotés des modèles annotés extérieurement.

Paramètres audio et faciaux robustes sur le corpus ETCC-HM		
	Paramètres audio	Paramètres faciaux
Modèles auto-annotés	Energie RMS : max, déviati on standard, étendue , décalage MSE (Mean Square Error) Δ Energie RMS : min, max, étendue, déviation standard, décalage MSE (Mean Square Error)	Arche des sourcils : max Distance sourcils-yeux : max , moyenne Écart sourcils : max Ouverture Bouche : min Δ Largeur boîte englobante : max
Modèles annotés extérieurement	Energie : moyenne, déviati on standard, min, max, étendue MFCC[1] : max MFCC[4] : moyenne, min, max, décalage MFCC[8] : étendue, déviation standard, décalage MSE (Mean Square Error) MFCC[9] : moyenne MFCC[10] : moyenne Δ MFCC[4] : déviation standard, décalage MSE (Mean Square Error) Δ MFCC[11] : déviation standard, étendue, décalage MSE (Mean Square Error)	Distance sourcils-yeux : max Δ Largeur boîte englobante : max Δ Hauteur boîte englobante : asymétrie

Détection Automatique du Stress

Tableau 42 : Paramètres robustes aux variations de configurations pour la reconnaissance du stress par les modèles conçus en fusion pondérée sur le corpus ETCC-HM (13 phobiques / 8 témoins en interaction avec une machine). Les paramètres en gras sont communs aux deux annotations.

Les paramètres audio et faciaux sont représentés de façon équilibrée dans le set utilisé pour la sélection (99 paramètres faciaux pour 100 paramètres audio) ; les modèles utilisent une combinaison de paramètres dans des proportions différentes suivant les annotations : les modèles auto-annotés utilisent 67% de paramètres faciaux, tandis que les modèles en annotation externe n'utilisent que 15% de paramètres faciaux. Cette sélection suggère que les annotateurs basent leur évaluation du stress en premier lieu sur la voix des sujets et la complète à partir des expressions faciales, tandis que le stress ressenti par les sujets se traduit dans leur voix et sur leur visage.

Les paramètres audio communs aux deux ensembles sont liés à l'énergie RMS, confortant le résultat de la sélection de paramètres uniquement audio présentée au Chapitre 7, et les résultats d'autres études sur la reconnaissance du stress dans la voix. Les paramètres faciaux communs sont liés à la partie haute du visage, et à la variation de taille du visage dans l'image, que l'on peut rapprocher des mouvements de tête des sujets puisqu'ils restaient assis face à la caméra au court de l'expérimentation. Les paramètres audio supplémentaires sélectionnés dans l'annotation externe sont tous liés à l'analyse cepstrale du signal.

A titre de comparaison, nous avons effectué une sélection de paramètres similaires en ne conservant que les 50 meilleurs, soit 25% du set de 100 paramètres audio et 99 paramètres faciaux. Cette sélection confirme l'utilisation de paramètres audio et visuels dans des proportions variables suivant l'annotation utilisée : les modèles auto-annotés utilisent 40% de paramètres faciaux, tandis que les modèles annotés en externe ne conservent aucuns des paramètres faciaux, et uniquement des paramètres audio liés aux MFCC.

Détection Automatique du Stress

Le Tableau 43 présente les paramètres sélectionnés sur les données audio et posturales des sujets non pathologiques de l'expérimentation CP2. Nous avons effectué la sélection des 50 meilleurs paramètres à partir des combinaisons de 100 paramètres audio et 15 paramètres posturaux, soit environ 50% des paramètres. Les paramètres présentés sont ceux systématiquement sélectionnés quels que soit la taille de fenêtre d'analyse appliquée et le sous-ensemble de sujets étudié.

Paramètres audio et posturaux robustes sur le corpus CP2		
	Paramètres audio	Paramètres posturaux
Modèles auto-annotés	<p>F0 : moyenne, déviation standard, aplatissement, asymétrie, décalage, décalage MSE (Mean Square Error)</p> <p>Δ F0 : aplatissement, décalage, pente</p> <p>Probabilité de Voix : moyenne, déviati</p> on standard, max, étendue, aplatissement, décalage, décalage MSE (Mean Square Error) <p>Δ Probabilité de Voix : min, max, étendue, déviati</p> on standard, décalage MSE (Mean Square Error)	<p>CI : max, min, moyenne</p>
Modèles annotés extérieurement	<p>F0 : étendue, déviati</p> on standard, décalage MSE (Mean Square Error) <p>Δ F0 : étendue, max</p>	<p>CI : max, moyenne</p>

Tableau 43 : Paramètres robustes aux variations de configurations pour la reconnaissance du stress par les modèles conçus en fusion pondérée sur le corpus CP2 (18 sujets en prise de parole en public). Les paramètres en gras sont communs aux deux annotations.

Les paramètres posturaux représentent 15% des paramètres du set utilisé pour la sélection (15 paramètres posturaux pour 100 paramètres audio) ; Les modèles auto-annotés sélectionnent des paramètres audio et posturaux en proportions comparables au set intégral de 115 paramètres (14% de paramètres posturaux) ; en revanche les modèles en annotation externe utilisent 33% de paramètres posturaux, suggérant que les observateurs aillant annoté le stress on tenu compte de la posture des sujets dans leurs évaluations.

Contrairement aux paramètres sélectionnés sur les sujets phobiques et témoins, et sur les paramètres sélectionnés en audio uniquement, les paramètres audio sélectionnés ici ne sont pas liés à l'énergie, mais à la fréquence fondamentale (F0). Les modèles auto-annotés sont également favorisés par des informations concernant la probabilité de voix dans l'enregistrement. Cette sélection peut s'expliquer par l'annotation bimodale effectuée par les observateurs, qui ont pu observer du stress dans les moments où les sujets parlent comme dans ceux où ils écoutent les questions du jury en silence.

Détection Automatique du Stress

Les indices posturaux sélectionnés dans les deux annotations sont liés à l'indice de contraction (CI) du corps, que l'on peut rapprocher des mouvements des bras par rapport au corps dans cette expérimentation, puisque les sujets gardaient une position fixe pendant tout l'enregistrement.

A titre de comparaison, nous avons également effectué une sélection des 25 meilleurs paramètres, soit environ 25% du set de 115 paramètres. Cette sélection confirme l'utilisation de paramètres audio et visuels dans des proportions variables suivant l'annotation utilisée : les modèles auto-annotés utilisent 7% de paramètres posturaux, tandis que les modèles annotés en externe utilisent 25% de paramètres posturaux. L'indice de contraction maximum est systématiquement sélectionné, quelque soit le nombre de paramètres sélectionnés et l'annotation considérée.

Cette étude de sélection des paramètres par des modèles fusionnant des informations audio et visuelles pour la reconnaissance du stress soulignent l'intérêt d'une approche multimodale. La variabilité expressive des sujets – qui selon les individus favorise l'expression audio ou visuelle de leurs états – et la difficulté d'anticiper cette variabilité rendent les performances de modèles monomodaux plus instables à l'échelle d'un ensemble de personnes enregistrées dans un même contexte anxiogène. Dans une approche multimodale, les modèles de reconnaissance automatique utilisent des paramètres à la fois audio et visuels dans des proportions variables suivant le contexte, et obtiennent des résultats plus homogènes à l'échelle d'un corpus (déviations standard moyennes de 0,04 en fusion pondérée, contre 0,07 en monomodal audio sur le corpus ETCC-HM ; de 0,08 contre 0,13 sur le corpus CP2).

Chapitre 9 Facteurs individuels d'influence sur la reconnaissance du stress

Dans les Chapitre 7 et Chapitre 8, nous avons constaté que les performances des modèles de reconnaissance du stress, qu'il s'agisse de modèles monomodaux conçus à partir de paramètres audio, faciaux ou posturaux, ou de modèles conçus en fusionnant ces paramètres, sont variables d'un sujet à un autre. Cette variabilité souligne l'aspect individuel du stress, et la difficulté d'établir un modèle de reconnaissance « universel ». Dans ce chapitre, nous étudions comment des informations caractérisant les individus peuvent améliorer les performances de reconnaissance du stress.

Nous utilisons les données des expérimentations E-thérapies en interaction avec des humains (enregistrements audio de 22 phobiques et 18 témoins) et en interaction homme machine (enregistrements audio et faciaux de 13 phobiques et 8 témoins dans des exercices de TCC*), et les données de l'expérimentation Comparse (enregistrements audio et posturaux de 18 sujets non pathologiques dans un entretien d'embauche factice). Afin d'uniformiser les comparaisons, nous utilisons les 100 paramètres audio et 15 paramètres posturaux ou 99 paramètres faciaux utilisés précédemment dans le Chapitre 7, extraits à partir de fenêtres d'analyse de 5000ms. Les modèles fusionnés présentés sont conçus en fusion pondérée, qui s'est avérée la méthode la plus efficace dans le Chapitre 8.

9.1 Phobie Sociale

Le projet E-Thérapie présenté au Chapitre 4 porte sur la reconnaissance du stress chez des individus souffrant de phobie sociale, afin d'évaluer la faisabilité d'outils de remédiation pour la Thérapie Cognitive Comportementale (TCC) utilisant l'état des patients pour adapter les exercices. Dans cet objectif, des sujets phobiques et un groupe de contrôle ont été enregistrés dans des exercices de TCC* menés par des humains ou par une machine, afin de capturer des expressions de stress de ces deux populations.

Une première étude a été conduite avec les membres du Centre Emotion de la Pitié Salpêtrière [330] sur le corpus ETCC-HH, qui comporte les enregistrements audio de 22 phobiques et 18 témoins en interaction avec des humains. Cette analyse visait à étudier les différences vocales entre les sociophobes et le groupe de contrôle, et les corrélations entre les différentes échelles psychométriques (mesure de la dépression, mesure de la phobie sociale, mesure des traits de personnalités), le questionnaire d'auto-évaluation du trouble anxieux (peur du jugement, difficulté à exprimer un commentaire) et les valeurs moyennes de paramètres audio.

Détection Automatique du Stress

Les paramètres audio utilisés dans cette précédente étude ont été extraits à l'aide du logiciel Praat [27] sur les enregistrements audio segmentés par tâche, et la valeur moyenne de ces paramètres a été évaluée à l'échelle de la tâche. Les paramètres sont présentés dans le Tableau 44.

Paramètres audio moyennés par tâche
Jitter: mesure des perturbations à court terme de la fréquence fondamentale.
Shimmer : mesure des perturbations à court terme de l'amplitude du signal sonore.
Ratio Harmoniques/Bruit (HNR)
Durée du signal,
Etendue des valeurs de la fréquence fondamentale
Taux de passage par zéro (ZCR)
Taux de silences
Etendue des valeurs de l'intensité

Tableau 44 : Paramètres audio extraits des données du corpus ETCC-HH (22 phobiques / 18 témoins en interaction avec des humains) pour l'étude des comportements vocaux des populations.

L'étude a révélé des corrélations entre les échelles psychométriques et les paramètres audio observés, variables suivant les tâches de l'expérimentation. Chez les sujets phobiques, la mesure de dépression (échelle HAD* [335]) est corrélée à l'étendue de l'intensité pendant la lecture, et la mesure d'extraversion (échelle BFI* [145]) est corrélée à la l'étendue de l'intensité pendant la question incongrue et la description positive de soi, ainsi qu'au jitter pendant la description positive de soi. Chez les sujets témoins, la mesure de dépression (HAD*) est corrélée à l'étendue de la fréquence fondamentale pendant la description positive de soi, l'extraversion (BFI*) est corrélée à la durée de lecture, et la mesure de phobie sociale (échelle LSAS* [183]) est également corrélée à la durée de lecture.

L'étude a également révélé des corrélations entre l'auto évaluation du stress et les paramètres audio observés, uniquement chez les sujets phobiques : la peur du jugement, la difficulté éprouvée et le stress ressenti sont corrélés au jitter sur la tâche de commentaire de texte, et le stress est également corrélé au HNR ; sur la tâche de question incongrue l'effet de surprise, la durée de réponse et la peur du jugement sont corrélées à l'étendue de la fréquence fondamentale, la peur du jugement et le stress ressenti sont corrélés au jitter et le stress est également corrélé au shimmer.

Détection Automatique du Stress

Ces différences de corrélations observées entre les sujets des deux populations montrent des comportements vocaux distincts entre les sujets phobiques et les sujets témoins dans des exercices d'affirmation de soi, suggérant un lien entre le groupe d'appartenance d'un individu et l'analyse de ses paramètres vocaux. Dans l'étude, l'appartenance à l'un ou l'autre des groupes permet de prédire les variations du shimmer sur toutes les tâches de l'expérimentation, ainsi que du jitter et du HNR sur certaines des tâches.

Dans le Chapitre 7, nous avons présenté les performances de modèle de reconnaissance du stress entraînés à partir de ces mêmes données, en utilisant les indices audio du set du Challenge Interspeech 2009 [270] pour la reconnaissance des émotions (c.f. 6.1.2). Ces modèles obtiennent des performances supérieures au hasard pour les configurations de paramètres audio et de fenêtres d'analyse les plus efficaces (f-mesure $0,65 \pm 0,06$ pour 25 paramètres audio extraits de fenêtres de 5s). Nous avons vu que ces performances étaient variables suivant la population observée : le stress des sujets phobiques est moins bien reconnu (f-mesure 0,49 pour le modèle utilisant des fenêtres de 5s) que celui des sujets témoins (f-mesure 0,70 pour le même modèle).

Nous avons appliqué la même distinction des performances par populations sur les données audio et faciales collectées pendant l'interaction des sujets phobiques et des témoins avec une machine. Les Figure 46 et Figure 47 illustrent les scores sur les deux populations des modèles monomodaux audio et faciaux, et du modèle en fusion pondéré des deux modalités, conçus à partir des paramètres extraits sur des fenêtres d'analyse de 5s. La f-mesure est calculée sur l'intégralité des données et non sur un sous-ensemble de sujets comme pour l'évaluation des modèles par corpus, aussi nous ne présentons pas de déviation standard.

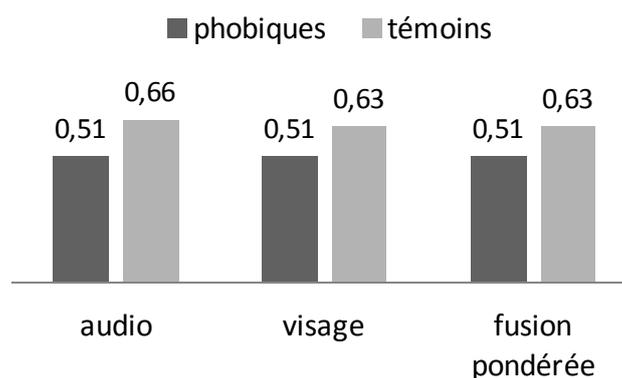


Figure 46 : Scores de f-mesure par type de modèle et par population à partir des auto-annotations du stress sur ETCC-HM (13 sujets phobiques et 8 sujets témoins en interaction avec une machine).

Détection Automatique du Stress

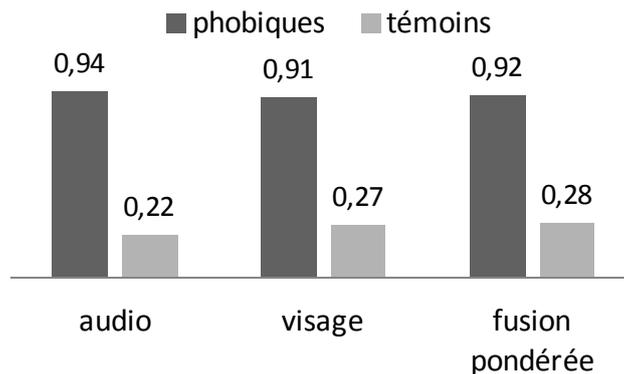


Figure 47 : Scores de f-mesure par type de modèle et par population à partir des annotations externes du stress sur ETCC-HM (13 sujets phobiques et 8 sujets témoins en interaction avec une machine).

Comme nous l'avons vu dans l'évaluation des modèles monomodaux et des modèles fusionnés, la phobie sociale est un facteur influent sur la reconnaissance du stress. Le type d'indices utilisé (audio, faciaux ou les deux informations combinées) a une influence négligeable sur le déséquilibre des performances entre les populations. Les sujets témoins sont systématiquement mieux reconnus par les modèles auto-annotés, et les sujets phobiques systématiquement mieux reconnus par les modèles annotés extérieurement.

Les scores très élevés de reconnaissance du stress des sujets phobiques pour les modèles auto-annotés peuvent s'expliquer par deux facteurs. Ces sujets ont exprimé plus de stress que les sujets témoins, qui ont exprimé du stress et du non-stress de façon plus homogène. Les performances des modèles sont généralement meilleures sur la classe stress que sur la classe non-stress, ce qui favorise les bonnes reconnaissances sur les sujets phobiques. Par ailleurs, le protocole expérimental – des exercices de TCC pour l'expression d'émotions – a été conçu en ciblant spécifiquement les sociophobes pour induire du stress. Ces sujets ont exprimé un stress particulièrement perceptible pour les annotateurs comparativement aux sujets témoins, favorisant leur reconnaissance par les modèles entraînés à partir de ces observations.

Les scores des modèles auto-annotés reflètent l'évaluation exagérée du stress par les sujets phobiques, et plus modérée par les sujets témoins. Au Chapitre 5, nous avons comparé les annotations des observateurs et les auto-évaluations des sujets, révélant un écart de perception : les sujets phobiques se sont évalué comme moyennement à fortement stressé, tandis que les annotateurs les ont perçus comme faiblement à moyennement stressés (c.f. 5.4.1). Cette évaluation exagérée défavorise les modèles de reconnaissance automatique du stress. Les scores des modèles auto-annotés restent relativement proches du hasard y compris pour les sujets témoins, supportant l'hypothèse d'un stress cognitif dont les signes ne sont pas directement extériorisés par les sujets.

Détection Automatique du Stress

Le projet E-Thérapie a permis la collecte de deux corpus suivant des protocoles expérimentaux différents : le corpus ETCC-HH porte sur des interactions des sujets avec des humains, tandis que le corpus ETCC-HM porte sur les interactions des sujets avec une machine. Dans la littérature, les TCC *in virtuo* sont évaluées comme mieux acceptées par les patients que les TCC *in vivo*, car les exercices avec une machine sont perçus comme moins menaçant que ceux avec des humains [107], tout en produisant des résultats comparables dans le traitement du trouble anxieux [167].

Les exercices utilisés dans les deux protocoles expérimentaux sont différents : dans les interactions avec les humains les sujets étaient amenés à exprimer des avis personnels sur des thèmes les mettant mal à l'aise (e.g. texte sur le mariage homosexuel, description de positive soi), tandis que dans les interactions avec la machine les sujets étaient amenés à exprimer des émotions (e.g. chamer quelqu'un et l'inviter à dîner, s'énerver contre un vendeur obséquieux). Les résultats des deux expérimentations ne sont donc pas directement comparables.

Cependant 9 sujets phobiques ont participé aux deux collectes, nous avons observé leurs réactions à titre informatif : en moyenne le stress auto-évalué par les sujets est plus élevé sur l'interaction avec la machine que sur celle avec les humains, mais les résultats ne sont pas homogènes sur l'ensemble des sujets. 4 des sujets ont été plus stressés par l'interaction avec des humains, 5 par celle avec la machine. Le stress post-expérimentation est équivalent pour 8 sujets sur les 9 : ils ont soit été stressés après les deux types d'exercices, soit non stressés.

Ces évaluations portent sur un petit nombre de sujets avec un niveau de phobie sociale non homogène, sans comparaison avec un groupe de contrôle, et ne révèlent pas de tendances indiquant un protocole plus ou moins efficace pour éliciter du stress chez les sociophobes.

Les résultats de l'étude menée en collaboration avec les membres du centre émotion ont révélé des différences vocales entre les populations phobiques et témoins enregistrées en interaction avec des humains. Afin d'estimer la faisabilité d'un système identifiant automatiquement la présence d'une pathologie phobique chez ses utilisateurs, nous avons conçu des modèles de reconnaissance de la phobie sociale à partir des paramètres audio et faciaux sur les données enregistrées auprès des sujets phobiques et témoins en interaction avec une machine. Nous avons évalué ces modèles dans des modèles monomodaux et dans un modèle en fusion par votes pondérés, à partir des indices extraits de fenêtres d'analyse de 5s.

Détection Automatique du Stress

Les résultats évalués indépendamment du locuteur sont supérieurs au hasard mais présentent une très forte variabilité interindividuelle. Il semble préférable d'utiliser les échelles d'évaluation psychologiques présentées au Chapitre 2 pour établir un diagnostic de phobie préalablement à l'utilisation d'un système de reconnaissance automatique du stress comme outil de remédiation, plutôt que de laisser le système l'identifier.

9.2 Traits de personnalité

Le projet Compare, présenté au Chapitre 4 porte sur l'évaluation de l'influence de traits psychologiques individuels sur la régulation des états affectifs, dont le stress. Dans ce but, des sujets non pathologiques ont été enregistrés suivant de multiples modalités durant un entretien d'embauche simulé conçu pour éliciter du stress chez les participants, en biaisant l'attitude des examinateurs.

Nous avons présenté au Chapitre 6 une étude réalisée en collaboration avec les membres du projet portant sur les liens entre l'auto-évaluation de leur performance par les sujets (frustration, motivation, difficulté perçue de la tâche et stress ressenti) et les indices audio et posturaux capturés au cours de l'expérimentation auprès de 18 sujets (c.f. 6.2.1). Cette étude [11] a montré des corrélations entre frustration et variations de la fréquence fondamentale (F0) et de l'indice de contraction (CI*), entre difficulté perçue et F0 moyenne, quantité de mouvement (QoM*) moyenne et variations de QoM, entre motivation et CI moyen et variations du CI, et entre stress auto-évalué et valeur moyenne et variations de QoM.

Nous évaluons ici l'influence des traits de personnalité des sujets sur les performances des modèles de reconnaissance du stress. Nous observons quatre informations de profil : les traits extraversion et névrotisme issus du modèle OCEAN, l'alexithymie (i.e. la difficulté à identifier et décrire ses émotions), et le trait anxieux (par opposition à l'état anxieux, réaction ponctuelle).

Nous avons distingué les 18 participants du projet en deux groupes pour chacun des traits observés, à partir de leurs résultats aux échelles d'évaluations psychométriques (BFI* [145] pour les traits extraversion et névrotisme, TAS-20 [12] pour le trait alexithymique, et STAI [290] pour le trait anxieux). Les performances des modèles monomodaux audio et posturaux, et le modèle en fusion pondéré conçus à partir de fenêtres d'analyse de 5s, sont comparées sur les deux groupes. La f-mesure est calculée sur l'intégralité des données et non sur un sous-ensemble de sujets comme pour l'évaluation des modèles par corpus, aussi nous ne présentons pas de déviation standard.

Détection Automatique du Stress

Dans le Chapitre 7, nous avons vu une influence significative de l'extraversion, du névrotisme et de l'alexithymie sur les performances des modèles de reconnaissance du stress auto-évalué par les sujets, favorisant les individus plus extravertis, névrotiques et alexithymiques que les individus ne présentant pas ces traits. Dans le Chapitre 8 nous avons constaté que l'influence de ces traits se retrouvait dans les performances des modèles en fusion pondérée. Ces résultats sont illustrés dans la Figure 48 (extraversion), la Figure 49 (névrotisme), et la Figure 50 (alexithymie).

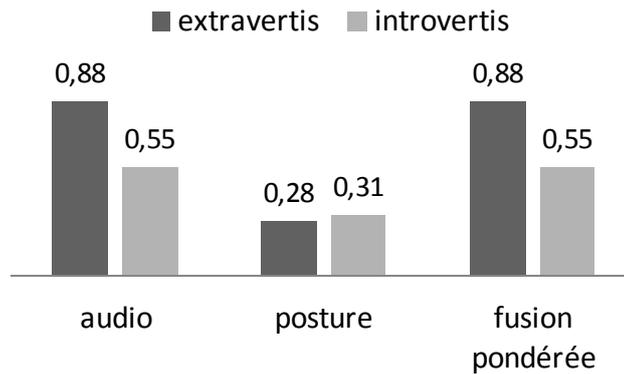


Figure 48 : Influence de l'extraversion sur les performances des modèles auto-annotés sur les données du corpus CP2 (18 sujets en prise de parole en public).

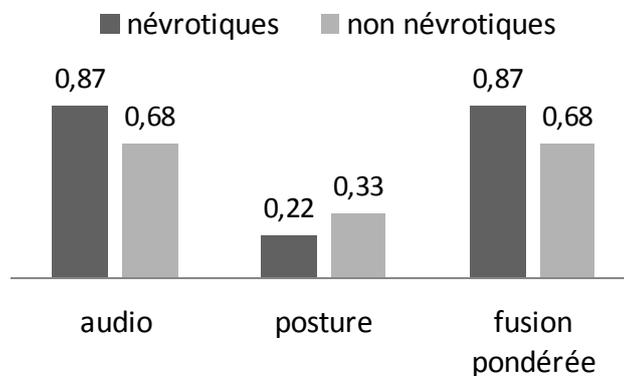


Figure 49 : Influence du névrotisme sur les performances des modèles auto-annotés sur les données du corpus CP2 (18 sujets en prise de parole en public).

Détection Automatique du Stress

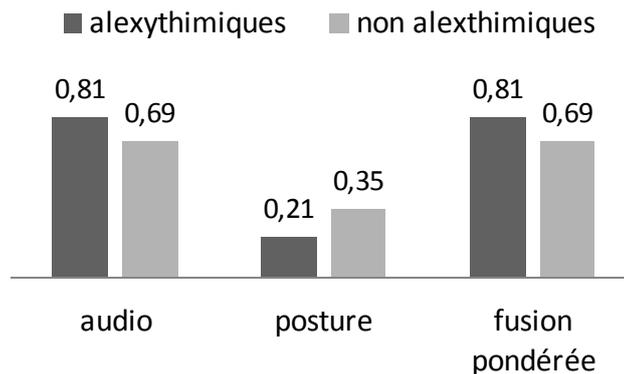


Figure 50 : Influence de l'alexithymie sur les performances des modèles auto-annotés sur les données du corpus CP2 (18 sujets en prise de parole en public).

Trois des quatre traits de personnalité observés ont une influence sur la reconnaissance du stress auto-évalué mais pas sur celle du stress évalué par les annotateurs : l'extraversion, le névrotisme et l'alexithymie semblent influencer sur l'expérience du stress chez les sujets, plutôt que sur son extériorisation visible par un observateur. Dans le contexte expérimental du projet Compare, la tâche proposée – un entretien d'embauche – n'encourage pas l'extériorisation du stress : un candidat pourrait choisir de masquer ses doutes et inquiétudes face à un recruteur pour ne pas pénaliser ses chances d'embauche.

Le trait anxieux n'a pas d'impact notable sur les performances des modèles auto-annotés, mais influe sur les performances des modèles annotés extérieurement. Les résultats sont illustrés par la Figure 51. Nous avons vu au Chapitre 7 que le modèle audio favorise la reconnaissance du stress des sujets anxieux. Le stress de ces sujets est particulièrement perceptible pour les annotateurs dans les données audio. Les données posturales sont moins sensibles à ce trait, et le modèle postural comme le modèle en fusion pondéré présente des écarts de performances moindre entre les deux populations.

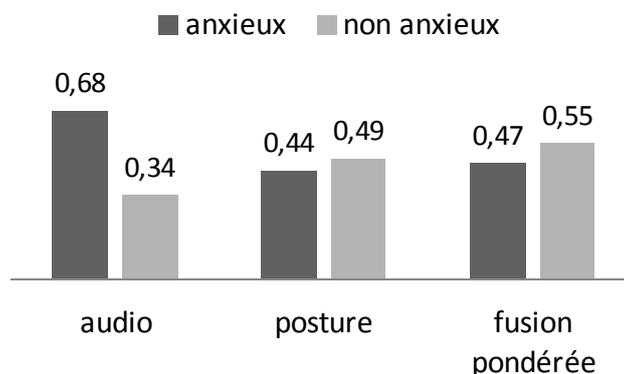


Figure 51 : Influence du trait anxieux sur les performances des modèles annotés extérieurement sur les données du corpus CP2 (18 sujets en prise de parole en public).

Détection Automatique du Stress

L'un des objectifs du projet Compare est la conception d'un logiciel plaçant les utilisateurs en interaction avec un agent conversationnel animé (ACA), pour s'entraîner à passer des entretiens d'embauche et ainsi réduire leur stress lors d'un entretien réel. Etant donnée l'influence des traits de personnalité sur le stress ressenti par les sujets, un tel système pourrait bénéficier d'une connaissance du profil des utilisateurs pour adapter les simulations.

La reconnaissance des traits de personnalités du modèle OCEAN (Ouverture, Conscienciosité, Extraversion, Agréabilité, Névrotisme) à partir de la voix a fait l'objet du challenge Interspeech 2012 [57] [212] [273]. Des études antérieures se sont également intéressées à la reconnaissance de ces traits dans la voix [256] [211] et les expressions faciales [103] ou en combinant ces deux modalités [238].

Nous avons cherché à reconnaître les traits de personnalités ayant une influence sur le stress des individus (Névrotisme, Extraversion [265], Anxiété, Alexithymie) à partir des données audio et posturales collectées lors de l'expérimentation Compare, sur les sujets du corpus CP2 (18 sujets).

Sur les quatre traits de personnalité étudiés, le névrotisme (résultat présentés en Figure 52) et l'anxiété (résultat présenté en Figure 53) sont les mieux reconnus. Comme pour la reconnaissance du stress, les modèles en fusion pondéré sont plus performants que les modèles monomodaux, présentant des scores de f-mesure plus élevés et une déviation standard des performances interindividuelles moins importante. Ce résultat illustre la variabilité de l'expressivité d'un individu à l'autre, certain favorisant l'expression vocale et d'autres l'expression corporelle ; la fusion de ces informations atténue ces différences.

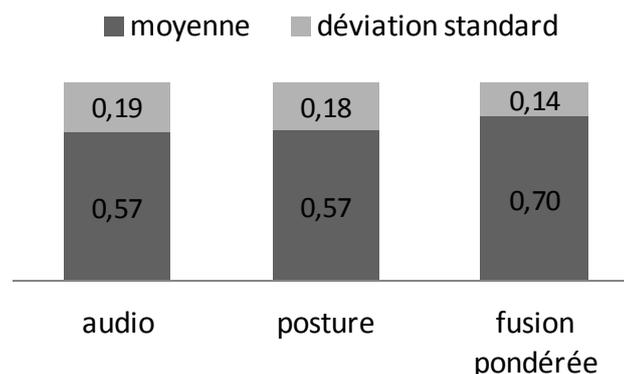


Figure 52 : Reconnaissance du névrotisme par des modèles audio, posturaux, et en fusion pondérée sur le corpus CP2 (18 sujets en prise de parole en public). Score moyen et déviatiion standard sur 3 sous-ensembles de sujets.

Détection Automatique du Stress

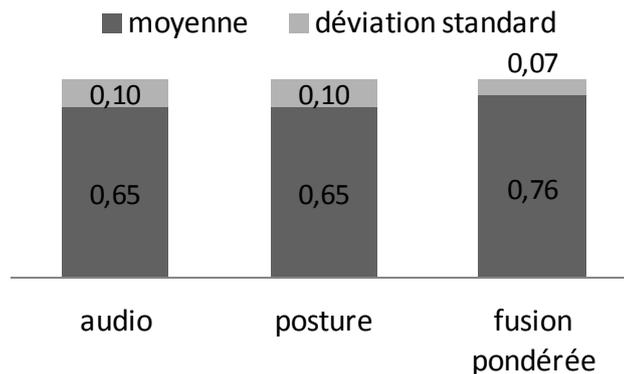


Figure 53 : Reconnaissance de l'anxiété par des modèles audio, posturaux, et en fusion pondérée sur le corpus CP2 (18 sujets en prise de parole en public). Score moyen et déviation standard sur 3 sous-ensembles de sujets.

Les scores de détection de l'extraversion et de l'alexithymie, facteurs influant sur le stress auto-évalué des sujets, sont inférieurs au hasard. L'utilisation des échelles psychométrique pour l'évaluation de ces traits de profil est préférable à un système automatique.

Nous avons également évalué l'influence du coping, ensemble de stratégies pour la régulation du stress, sur la reconnaissance du stress. Les participants à cette expérimentation ont répondu au questionnaire BRIEF COPE* [46] qui évalue 14 catégories de stratégies de coping. Pour simplifier la comparaison, nous avons regroupé les sujets en deux groupes : ceux appliquant des stratégies fonctionnelles (e.g. acceptation, réévaluation positive de la situation, humour) les aidant à surmonter leur stress, et ceux appliquant des stratégies dysfonctionnelles (e.g. dénis, culpabilité) ne leur permettant pas de surmonter leur stress.

La reconnaissance du stress auto-évalué par les sujets est indifférente au type stratégies de coping, le stress des deux groupes étant reconnu de façon similaire. En revanche, le modèle de reconnaissance du stress annoté extérieurement conçu en fusion pondérée des paramètres audio et posturaux est plus performant sur les sujets adoptant des stratégies dysfonctionnelles: le stress des sujets adoptant des stratégies fonctionnelles est reconnu à 49%, tandis que celui des sujets adoptants des stratégies dysfonctionnelles est reconnu à 57%. Ces sujets appliquent des stratégies ne les aidant pas à surmonter leur stress, et vont exprimer un stress plus perceptible pour les annotateurs, et plus reconnaissable pour le modèle.

Détection Automatique du Stress

La reconnaissance automatique du type de stratégie de coping à travers les données audio et posturale collectée produit des résultats inférieurs au hasard, l'utilisation des échelles de mesure psychométrique est donc préférable pour mesurer cette information. Si l'influence de cette caractéristique sur la reconnaissance du stress est limitée, il s'agit d'une information importante pour la conception d'un outil d'entraînement à la gestion du stress : le système pourra, par des exercices de remédiation, proposer aux sujets moins bien armés des stratégies de coping plus efficaces.

9.3 Modèles de reconnaissance du stress tenant compte des informations de profil

Les sous-chapitres 9.1 et 9.2 ont montré l'influence de certains facteurs individuels sur les performances des modèles de reconnaissance du stress. Les résultats de reconnaissance automatique de la phobie sociale, des stratégies de coping et des traits de personnalités obtenus à partir des données collectées dans les expérimentations Comparse et E-Thérapies sont au mieux de l'ordre du hasard. Nous disposons cependant des réponses des sujets aux échelles de mesures psychologiques de ces différents facteurs de profil.

Nous avons utilisé ces informations afin de concevoir des modèles de reconnaissance du stress utilisant à la fois des paramètres extraits des enregistrements audio et visuels (qu'il s'agisse de capture du visage ou de la posture), et des paramètres de profil caractérisant les individus.

9.3.1 Sujets phobiques et non-phobiques

Dans le projet E-Thérapies, nous avons enregistré 13 sujets phobiques et 8 sujets témoins en interaction avec une machine, constituant le corpus ETCC-HM. Nous avons collecté auprès des sujets phobiques et témoins les scores de l'échelle de diagnostic de la phobie sociale (LSAS* [183]). Nous avons ajouté ces scores aux paramètres audio et faciaux extraits des enregistrements par des fenêtres d'analyse de 5s, et entraîné de nouveaux modèles de reconnaissance automatique du stress auto-évalué et annoté par des observateurs externes. Les résultats des modèles entraînés en fusion par vote pondéré sont présentés en Figure 54 et Figure 55.

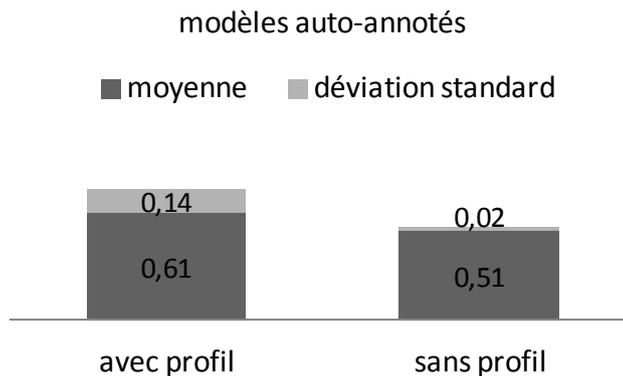


Figure 54 : Reconnaissance du stress en fusion pondéré d'informations audio et faciales avec et sans informations de profil à partir des auto-annotations sur le corpus ETCC-HM (13 phobiques et 8 témoins en interaction avec une machine). Score de f-mesure moyen et déviation standard sur 3 sous-ensembles de sujets.

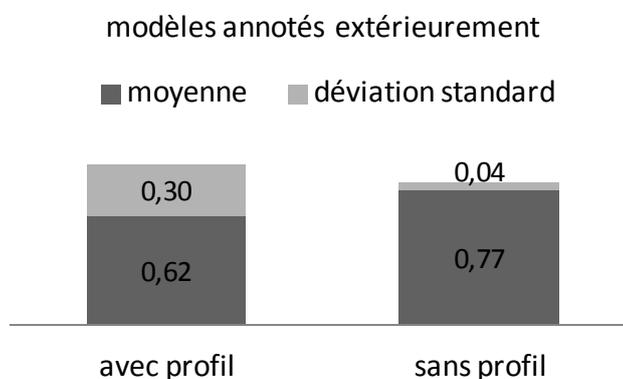


Figure 55 : Reconnaissance du stress en fusion pondéré d'informations audio et faciales avec et sans informations de profil à partir des annotations externes sur le corpus ETCC-HM (13 phobiques et 8 témoins en interaction avec une machine). Score de f-mesure moyen et déviation standard sur 3 sous-ensembles de sujets.

L'ajout des informations de profil dans les paramètres utilisés pour entraîner les modèles de reconnaissance du stress rend les performances fortement variables suivant les individus. Ce résultat reflète les performances variables des modèles sur les sujets phobiques et les sujets témoins présentées au début de ce chapitre et suggère une expression du stress différente entre les deux populations. L'utilisation d'un modèle unique du stress, y compris avec des informations quant à la phobie des sujets résulte en des performances fortement variables. Le petit nombre de sujets, et le nombre de sujets phobiques plus important que le nombre de témoin, ne nous permettent pas de concevoir des modèles distincts pour les deux populations.

Détection Automatique du Stress

L'utilisation conjointe des données audio collectées auprès de sujets en interaction avec une machine (corpus ETCC-HM) et auprès de sujets en interaction avec des humains (corpus ETCC-HH) n'améliore pas les performances de reconnaissance du stress. Ce résultat suggère que d'autres facteurs que la phobie sociale influent sur la reconnaissance du stress, notamment le contexte.

Nous avons vu que les protocoles expérimentaux des deux collectes sont différents, le corpus en interaction avec une machine implique l'expression d'émotions, ce qui n'est pas le cas pour le corpus en interaction avec des humains. Ce contenu affectif, et l'objet du stress (i.e. un humain ou une machine), affectent également la reconnaissance automatique du stress.

9.3.2 Sujets non-pathologiques

Dans le cadre de l'expérimentation Compare, nous avons enregistré 18 personnes en prise de parole en public, constituant le corpus CP2. Nous avons collecté auprès des sujets les scores des échelles de mesure de l'alexithymie (TAS* [12]), du trait anxieux (STAI* [290]) et des traits de personnalité (BFI* [145]). Nous avons ajouté ces scores aux paramètres audio et posturaux extraits des enregistrements, et entraîné de nouveaux modèles de reconnaissance automatique du stress auto-évalué et annoté par des observateurs externes. Les résultats des modèles entraînés en fusion par vote pondéré sont présentés dans les Figure 56 et Figure 57.

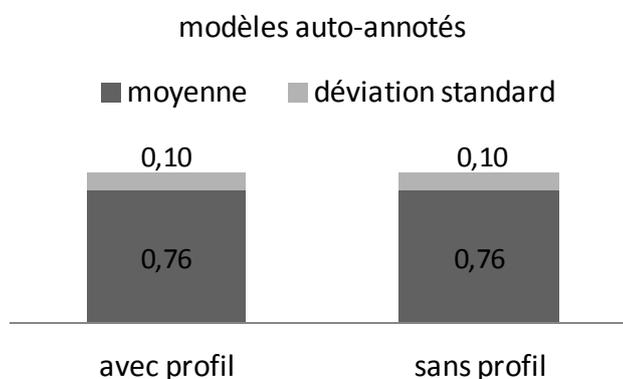


Figure 56 : Reconnaissance du stress en fusion pondéré d'informations audio et posturales avec et sans informations de profil à partir des auto-annotations sur le corpus CP2 (18 sujets en prise de parole en public). Score de f-mesure moyen et déviation standard sur 3 sous-ensembles de sujets.

Détection Automatique du Stress

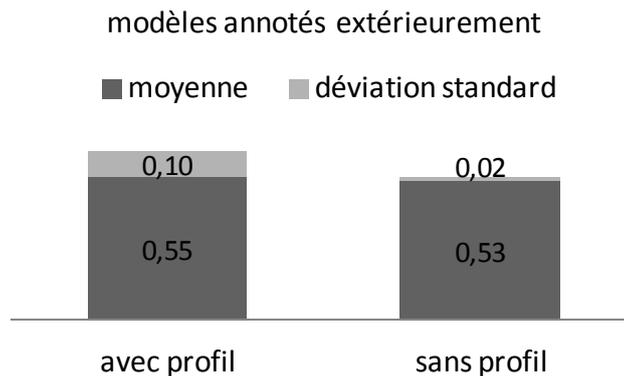


Figure 57 : Reconnaissance du stress en fusion pondéré d'informations audio et posturales avec et sans informations de profil à partir des auto-annotations sur le corpus CP2 (18 sujets en prise de parole en public). Score de f-mesure moyen et déviation standard sur 3 sous-ensembles de sujets.

L'ajout des informations de profil, qui ont montré une influence sur les performances des modèles de reconnaissance du stress suivant les individus, n'influent pas sur les performances des modèles auto-annotés. Les performances des modèles annotés extérieurement montrent une variabilité des performances d'un individu à l'autre plus importante lorsque le modèle tient compte des informations de profil, pour une performance moyenne comparable à celle du modèle n'utilisant pas ces informations.

En étudiant la classification des paramètres suivant le gain qu'ils apportent à la reconnaissance du stress, les paramètres de profil sont systématiquement dans les 10 meilleurs, soulignant l'importance des facteurs individuels dans l'expression du stress. L'ajout des informations de profil n'est pas suffisant pour que l'algorithme SVM puisse séparer plus efficacement l'espace des paramètres audio et posturaux afin de distinguer le stress des sujets.

Ce résultat suggère une variabilité de l'expression du stress lié à d'autres facteurs. Nous avons vu dans la comparaison des performances des modèles monomodaux et posturaux que d'un sujet à l'autre, la modalité la plus performante variait, et que l'écart de performance entre les deux modalités était plus ou moins important. Cette diversité expressive est un élément influant sur la reconnaissance du stress.

En observant les performances des modèles par sujet, nous constatons que pour certains le stress est particulièrement mal identifié, quelque soit la modalité considérée pour mesure des indices du stress. Pour ces individus, il est possible que le stress s'exprime par d'autres canaux, par exemple les expressions faciales – non étudiées sur ce corpus – le regard ou le vocabulaire utilisé pour répondre aux questions des examinateurs ou pour décrire leurs qualités et défauts. Toutes ces informations vont influencer l'évaluation du stress par les observateurs, mais ne transparaîtrons pas dans les paramètres audio ou posturaux.

Détection Automatique du Stress

Ces individus peuvent également montrer des signes de stress atypiques par rapport aux autres sujets du corpus. Au cours de l'expérimentation, un sujet fortement stressé a éclaté en sanglots. Ce signe est particulièrement explicite pour les observateurs, mais n'est pas représentatif du reste des données et les performances des modèles sur ce sujet en sont affectées, les pleurs altérant la voix.

9.4 Implications pour la conception de logiciels de remédiation

Dans l'objectif de concevoir un système capable de reconnaître de façon automatique le stress des utilisateurs afin d'adapter les exercices d'un outil de remédiation, il est nécessaire de tenir compte de la diversité des expériences et expressions de stress.

Cette étude présente une approche pluridisciplinaire de la reconnaissance du stress, utilisant à la fois des mesures de la psychologie pour caractériser les individus, et des informations extraites de l'analyse de signaux multimodaux capturés lors de mises en situations stressantes.

Dans les expériences que nous avons menées, les performances des modèles de reconnaissance du stress sont influencées à la fois par des facteurs contextuels suivant la tâche utilisée pour éliciter du stress chez les sujets, et par des facteurs individuels ayant un impact sur l'expérience et l'expression du stress : pathologie anxieuse, stratégies de coping pour réguler le stress, et traits de personnalités.

Ces informations sont modérément reconnues de façon automatique à partir de l'analyse des signaux vocaux, faciaux ou posturaux. Nous avons utilisé les résultats des échelles de mesures psychologiques pour les recueillir et les avons ajoutées aux paramètres extraits des signaux pour entraîner des modèles de reconnaissance du stress tenant compte à la fois de l'expression ponctuelle d'un individu et de ses spécificités de profil sur le long terme. Ces informations sont valorisées pour la reconnaissance du stress par la sélection de paramètres, mais ne sont pas suffisantes pour améliorer les performances interindividuelles.

La comparaison des performances des modèles suivant les sujets et suivant la modalité (audio, faciale ou posturale) montrent une variabilité expressive d'un individu à l'autre, valorisant l'utilisation d'indices issus de différents canaux pour la reconnaissance du stress. Les performances faibles des modèles sur certains sujets quelque soit la modalité observée indiquent également que le stress peut s'exprimer par d'autres modalités que celles que nous avons étudié.

L'étude de cas de stress atypiques fortement expressifs (e.g. pleurs) montre également que la performance des modèles est liée à la représentativité des données d'apprentissage. L'utilisation de données collectées auprès d'un grand nombre de sujets permet de couvrir une plus large gamme de comportements expressifs et rendre les modèles plus robustes à la variabilité des comportements.

Synthèse de la Détection Automatique du Stress

Nous avons utilisé les données collectées à travers les différentes expérimentations de cette étude afin d'entraîner des modèles de reconnaissance automatique du stress.

Nous avons d'abord entraîné ces modèles à partir d'une seule modalité (audio, faciale ou posturale selon les corpus), en variant la fréquence d'extraction des paramètres, et le nombre de paramètres utilisés pour la modalité audio. Nous avons évalué la reconnaissance du stress auto-évalué et du stress annoté par des observateurs externes.

Les résultats n'indiquent pas une configuration systématiquement plus performante à travers les différentes expérimentations. Les modèles entraînés à partir de peu d'indices (sélection de 100 à 25 paramètres audio, 15 paramètres posturaux) obtiennent des performances comparables ou supérieures aux modèles utilisant les 384 paramètres audio du set du Challenge Interspeech 2009 pour la reconnaissance des émotions. Ces résultats sont avantageux pour la conception d'un système de reconnaissance du stress pour des logiciels de remédiation à destination du grand public, sur des machines aux capacités de calcul limitées

Nous avons ensuite entraîné ces modèles à partir de deux modalités suivant différentes stratégies de fusion :

- une approche « précoce » utilisant les paramètres issus de deux modalités pour la constitution d'un nouveau modèle

- des approches « tardives » combinant les résultats de classification des modèles monomodaux pour évaluer la catégorie d'un segment donné. Nous avons comparé trois combinaisons : un vote majoritaire simple, un vote pondéré par les performances des modèles monomodaux, et une cascade de spécialiste utilisant le meilleur modèle pour chaque classe (stress ou non-stress).

Les meilleurs résultats sont atteints en fusion par votes pondérés, surpassant les performances des modèles monomodaux. Ces résultats soulignent l'expression multimodale du stress et l'intérêt de combiner des informations provenant plusieurs sources pour sa reconnaissance.

Nous avons évalué les performances des modèles monomodaux et multimodaux suivants différents critères susceptibles d'influencer l'expérience et l'expression du stress : public phobique ou non-pathologique, stratégies de coping fonctionnelles ou dysfonctionnelles, et traits de personnalité. Les performances des modèles varient fortement suivant le profil des sujets, illustrant l'aspect fortement individuel du stress.

Nous avons conçu des modèles de reconnaissance du stress à partir des paramètres audio, faciaux et posturaux incluant ces informations de profil. Les résultats ne modifient pas les performances pour les sujets non pathologiques mais influent sur les performances pour les sujets phobiques, soulignant l'intérêt de disposer d'informations de profil pour la conception d'outils thérapeutiques visant un public souffrant de troubles anxieux.

Conclusion

"Stress is not necessarily undesirable. It all depends on how you take it."

– Hans Selye 'The Nature of Stress', 1985 [283]

Contexte et rappel des objectifs de recherche

Le stress est une réaction naturelle à une menace réelle ou perçue, dont les impacts peuvent être néfastes sur la capacité d'un individu à atteindre ses objectifs, qu'il s'agisse de buts personnels ou professionnels. Afin de contrer ces effets néfastes, il est possible d'apprendre des stratégies de coping, des méthodes qui permettent d'appréhender une situation stressante et d'y faire face plutôt que de la fuir. L'apprentissage de ces stratégies passe par des exercices de remédiation, qui peuvent prendre des formes variées et consistent à réévaluer des comportements négatifs dans une situation donnée, pour les remplacer par des comportements positifs.

Les travaux présentés dans cette thèse ont été financés par le projet FEDER E-Thérapie, étude visant à valider la faisabilité de logiciels à but thérapeutique, à destination d'un public sociophobe. L'objectif ambitieux de ce projet est la création d'un outil de remédiation proposant des exercices de Thérapie Cognitive Comportementale *in virtuo* avec des agents conversationnels animés (ACA), paramétrables par les thérapeutes pour s'adapter aux besoins de leurs patients. Pour les sujets phobiques, cet outil permettrait la répétition d'exercices anxiogènes dans un cadre contrôlé, tout en bénéficiant d'une mesure de leur progression par un *biofeedback* de leur état de stress. Un tel système nécessite la reconnaissance de signes expressifs révélant le stress des sujets, à partir de capteurs peu invasifs.

Notre étude a porté sur l'évaluation de la reconnaissance du stress à partir d'indices vocaux et faciaux chez des personnes phobiques et non-phobiques en situation d'interaction anxiogène. Nous avons collaboré avec les médecins psychiatres de l'Hôpital de la Pitié Salpêtrière pour concevoir des scénarios d'interactions avec des humains et avec des animations simplifiées sans ACA afin de collecter des données et étudier la perception humaine et la reconnaissance automatique du stress.

Conclusion

Nous avons également collaboré au projet ANR Compare, dont l'objectif est l'étude des comportements affectifs dans des interactions avec des humains et avec des ACA. Nous avons participé à la collecte de données de stress dans les interactions avec des humains, sous la forme d'un entretien d'embauche simulé. Dans ce projet nous avons enregistré des indices vocaux et posturaux afin d'étudier le stress de sujets non pathologiques en prise de parole en public. Enfin dans le cadre de l'Action Incitative « Nao Fait de l'Humour » nous avons étudié les rires dans des interactions humain-machine suivant des tâches plus ou moins stressantes. Ces projets nous ont également donné l'opportunité d'étudier l'influence du contexte, de la personnalité, et des stratégies de *coping* sur le stress des sujets.

Nos travaux abordent l'étude du stress de façon pluridisciplinaire, dans l'objectif de concevoir des modèles de reconnaissance du stress pour des outils de remédiation sous la forme de *serious games*, à destinations de publics phobiques ou non-pathologiques, à partir d'informations collectées par des capteurs peu invasifs. Dans ce but, nous avons étudié la reconnaissance du stress dans différents contextes interactionnels (avec des humains, ou avec des machines) en tenant compte des caractéristiques individuelles des sujets affectant leur expérience du stress (phobie, traits de personnalité) mais également leur expression du stress. Nous avons collecté des données audio, faciales et posturales et cherché à identifier des signes perceptibles du stress suivant les contextes et les individus, mais également des indices robustes à ces variations.

Résumé des contributions

Afin d'étudier les différences interindividuelles et inter-contextuelles sur l'expérience et l'expression du stress, nous avons contribué à l'élaboration de trois protocoles expérimentaux ciblant des individus phobiques et des individus non pathologiques, dans différentes tâches d'interaction anxiogènes. Ces protocoles incluent la collecte des caractéristiques individuelles des participants, à la fois par des échelles de mesure psychométriques (mesure de la phobie sociale, mesure des traits de personnalité, mesure des stratégies de coping), des informations socioculturelles (âge, sexe, milieu culturel), et des questionnaires d'évaluation de l'expérience permettant aux sujets d'indiquer le stress ressenti au cours de l'expérimentation. Nous avons collaboré à la conception d'outil de deux interfaces « Magicien d'Oz » pour la mise en œuvre de ces protocoles afin d'enregistrer des sujets en interaction avec une machine, et élaboré des outils d'extraction d'indices faciaux et posturaux pour analyser les données enregistrées. Les différents protocoles et outils ont permis la collecte de 5 corpus auprès de 164 personnes pour plus de 20 heures d'enregistrements audio, vidéo et posturaux.

Conclusion

L'auto-évaluation du stress est subjective, elle relève des processus cognitifs des individus qui ne sont pas directement perceptibles par des signes extérieurs. Afin d'obtenir une évaluation plus objective, nous avons établi un protocole d'annotation du stress en nous appuyant sur les signes proposés en psychologie, et fait évaluer le stress par des annotateurs externes sur les corpus collectés. Nous avons confronté cette seconde annotation à celle réalisée par les sujets des expérimentations, révélant une différence d'évaluation entre stress ressenti et stress perçu, particulièrement chez les sujets phobiques.

Des données collectées et annotées, nous avons extrait des paramètres caractérisant la voix, l'expression faciale et la posture des sujets au cours des différentes tâches anxiogènes proposées par les protocoles expérimentaux. Nous avons choisi des paramètres issus de la recherche dans le domaine de *l'affective computing*, utilisés pour la reconnaissance des états affectifs et du stress chez les humains par des machines.

Nous avons utilisé ces indices pour entraîner et évaluer des modèles de reconnaissance du stress, en choisissant l'algorithme SVM pour sa capacité à catégoriser des données dans un espace de paramètres de grandes dimensions. Nous avons comparé les performances de modèles monomodaux conçus à partir de l'une des modalités capturées (audio, faciale ou posturale), puis de modèles fusionnant les informations provenant de plusieurs modalités, pour la reconnaissance du stress auto-évalué par les sujets, et du stress annoté par des observateurs externes à partir du protocole établi. De ces comparaisons, nous tirons plusieurs observations :

- Le stress auto-évalué est moins bien reconnu que le stress évalué à partir de signes perceptifs mesurables dans la voix, la posture, ou les expressions faciales, soulignant le caractère subjectif de l'expérience du stress.
- Les modèles entraînés à partir d'une seule modalité reconnaissent le stress avec des scores supérieurs au hasard, et présentent une variabilité d'un individu à l'autre. Cette hétérogénéité est liée aux différences expressives interindividuelles : certains sujets expriment plus visiblement leur état de stress par leur voix, d'autres par leur posture ou par leurs expressions faciales. Cette variabilité expressive explique également que dans nos travaux nous n'ayons pas identifié une modalité plus adaptée qu'une autre pour la reconnaissance du stress.
- Les modèles entraînés en fusion précoce – c'est-à-dire en combinant les paramètres audio et faciaux, ou audio et posturaux selon les corpus utilisés – sélectionnent une combinaison d'informations audio et visuelles pour distinguer le stress des sujets, valorisant l'utilisation de multiples modalités pour la reconnaissance du stress afin de tenir compte des variabilités d'expressivité des individus.

Conclusion

- Les modèles entraînés en fusion tardive – c'est-à-dire en combinant les résultats des modèles monomodaux – obtiennent les performances les plus élevées en pondérant les décisions des modèles monomodaux par les performances de ces modèles. Comparativement aux modèles n'utilisant les informations que d'une seule modalité, les résultats obtenus par cette fusion pondérée présentent une variabilité interindividuelle réduite et sont comparables ou légèrement supérieurs. Ces résultats valorisent également l'utilisation d'informations multimodales pour identifier le stress chez des individus aux caractéristiques expressives variables dans différents canaux (audio et visuel).
- Les performances des modèles suivant les caractéristiques individuelles influentes sur le stress des sujets sont variables. La phobie sociale, l'extraversion, le névrotisme, et l'alexithymie impactent le ressenti et l'expressivité des sujets et les performances des modèles sur les auto-évaluations en sont affectées; la phobie sociale et le trait anxieux induisent des signes de stress perceptibles par des observateurs externes et influencent les performances des modèles sur leurs évaluations.

Perspectives et Discussion

Les résultats de nos travaux soulignent l'hétérogénéité du stress, processus cognitif influencé par des facteurs à la fois contextuels et individuels, notamment par les troubles anxieux tels que la phobie sociale. A la variabilité des expériences s'ajoute la variabilité des manifestations de ce ressenti : suivant les individus, le stress est exprimé de manière plus ou moins ouverte, et se traduit par des signes perceptibles dans différents aspects de leur comportement. Cette variabilité est visible dans les résultats de l'annotation du stress présenté dans le Chapitre 5 de cette étude, qui montre la subjectivité de l'auto-évaluation du stress – particulièrement sous l'influence d'un trouble phobique – et la difficulté de l'annotation perceptive consensuelle du stress. L'annotation en *crowdsourcing* mettant à contribution un plus grand nombre d'observateurs non experts peut être une voie d'amélioration du processus d'annotation ; il faut cependant anticiper les implications éthiques d'un tel procédé, particulièrement sur des données à caractère fortement personnel dans le cas de personnes souffrant d'un trouble anxieux. L'utilisation de ces données nécessite d'assurer leur confidentialité, en particulier pour des outils d'annotation par applications web impliquant l'échange d'informations entre deux machines distantes. Ce procédé implique également l'élaboration d'un protocole de segmentation et d'annotation des données adapté pour de multiples observateurs.

Conclusion

Nous avons étudié le stress chez des sujets sociophobes et chez des sujets non pathologiques à travers les variations de leur voix, de leur posture et de leurs expressions faciales, pour évaluer la faisabilité d'un système de reconnaissance du stress pour des outils de remédiation. La diversité des comportements expressifs humains implique que leur reconnaissance et leur interprétation par une machine seront plus homogènes si celle-ci a été exposée à une variété d'exemples représentatifs. Les résultats présentés dans la Section III de cette étude tendent vers une reconnaissance du stress à partir d'indices multimodaux et en tenant compte des caractéristiques psychologiques et pathologiques des sujets pour une reconnaissance plus robuste du stress. Ces résultats pourront être renforcés par la poursuite de cette étude auprès de sujets supplémentaires pour tenir compte d'autres facteurs d'influence sur le stress, comme d'autres pathologies ou d'autres situations anxiogènes. Il serait également intéressant d'étudier d'autres modalités porteuses d'informations sur l'état de stress des sujets, comme les indices physiologiques, le regard, le rougissement, ou encore le contenu linguistique des interactions. Enfin l'utilisation d'autres procédés de fusion permettant de tenir compte de l'expressivité variable d'un individu suivant les modalités observées, et de la possible asynchronie des informations dans ces modalités pourrait améliorer la reconnaissance multimodale du stress.

Il nous semble important de conclure cette étude par un regard sur les applications ciblées par les projets E-Thérapies et Comparses auxquels nous avons participé. Ces projets visent à la conception de *serious games* comme outils de remédiation dans le traitement des troubles anxieux, mais également pour l'entraînement à des situations anxiogènes comme un entretien d'embauche. L'impact négatif du stress sur les objectifs tant personnels que professionnels des individus ne parvenant pas à maîtriser leur anxiété motive la diffusion d'outils pour l'apprentissage de stratégies de coping. En particulier, des systèmes ciblant des populations pathologiques comme les personnes atteintes de trouble anxieux représentent un complément aux exercices de thérapie classiques : les interactions avec une machine sont perçues comme moins menaçantes que des exercices *in vivo* pour se confronter à des situations anxiogènes. Si l'expertise du thérapeute reste nécessaire pour le traitement des phobies, les outils informatiques sont un complément intéressant dans la panoplie d'exercices du professionnel et se sont avérés tout aussi efficaces que des exercices *in vivo*. Il est cependant nécessaire d'anticiper les implications éthiques de tels systèmes : les données nécessaires à l'élaboration de modèles performants sont à caractère personnel et relativement sensibles et leur utilisation nécessite d'assurer leur confidentialité.

Bibliographie

- [1] ABRILIAN, S., DEVILLERS, L., MARTIN, J. C. (2006) Annotation of emotions in real-life video interviews: Variability between coders. In : 5th Int. Conf. on Language Resources and Evaluation (LREC 2006), Genoa, Italy. 2006
- [2] AMAYA, A., (2012) Facteurs de stress dans la prise de parole en public. Mémoire de Master 2 professionnel Sciences de l'Information et la Communication, Université Paris-Sorbonne IV. Septembre 2012
- [3] AMERICAN PSYCHIATRIC ASSOCIATION, et al. The Diagnostic and Statistical Manual of Mental Disorders: DSM 5. bookpointUS, 2013.
- [4] ANDERSON, P. L., ZIMAND, E., HODGES, L. F., et al. (2005) Cognitive behavioral therapy for public-speaking anxiety using virtual reality for exposure. *Depression and anxiety*, 2005, vol. 22, no 3, p. 156-158.
- [5] ANTONY, M. M., SWINSON, R. P. (2000) Phobic disorders and panic in adults: A guide to assessment and treatment. American Psychological Association, 2000.
- [6] ARNOLD, M B, (1960). *Emotion and Personality* (2 Vols.). New York: Columbia University Press
- [7] AMIR, N., WEISS, A., HADAD, R. (2009) Is there a dominant channel in perception of emotions?. In : *Affective Computing and Intelligent Interaction and Workshops*, 2009. ACII 2009. 3rd International Conference on. IEEE, 2009. p. 1-6.
- [8] AMIRKHAN, J. H. (1990). A factor analytically derived measure of coping: The Coping Strategy Indicator. *Journal of Personality and Social Psychology*, 59, 1066-1074.
- [9] ANDERSON, P. L., ZIMAND, E., HODGES, L. F., et al. (2005) Cognitive behavioral therapy for public-speaking anxiety using virtual reality for exposure. *Depression and anxiety*, 2005, vol. 22, no 3, p. 156-158.
- [10] ANESHENSEL, C. S. (1992) Social stress: Theory and research. *Annual review of sociology*, 1992, p. 15-38
- [11] ASHTON, M. C.; LEE, K.; PERUGINI, M.; SZAROTA, P.; DE VRIES, et.al. (2004) A Six-Factor Structure of Personality-Descriptive Adjectives: Solutions From Psycholexical Studies in Seven Languages. *Journal of Personality and Social Psychology* 86 (2): 356-366
- [12] BAGBY, R. M., PARKER, J. DA., TAYLOR, G. J. (1994) The twenty-item Toronto Alexithymia Scale–I. Item selection and cross-validation of the factor structure. *Journal of psychosomatic research*, 1994, vol. 38, no 1, p. 23-32.
- [13] BAGBY, R. M., TAYLOR, G. J., PARKER, J. DA. (1994) The twenty-item Toronto Alexithymia Scale–II. Convergent, discriminant, and concurrent validity. *Journal of psychosomatic research*, 1994, vol. 38, no 1, p. 33-40.
- [14] BANDURA, A. (1977) Social learning theory.
- [15] BANSE, R., SCHERER, K. R. (1996) Acoustic profiles in vocal emotion expression. *Journal of personality and social psychology*, 1996, vol. 70, no 3, p. 614.
- [16] BÄNZIGER T, (2004) *Communication Vocale des Emotions - Perception de l'expression vocale et attributions émotionnelles*, Thèse de doctorat 2004

Bibliographie

- [17] BÄNZIGER, T., MORTILLARO, M., SCHERER, K. R. (2012) Introducing the Geneva Multimodal expression corpus for experimental research on emotion perception. *Emotion*, 2012, vol. 12, no 5, p. 1161.
- [18] BÄNZIGER, T., SCHERER, K. R. (2007) Using actor portrayals to systematically study multimodal emotion expression: The GEMEP corpus. In : *Affective computing and intelligent interaction*. Springer Berlin Heidelberg, 2007. p. 476-487.
- [19] BARRAS, C., GEOFFROIS, E., WU, Z., *et al.* (1998) Transcriber: a free tool for segmenting, labeling and transcribing speech. In : *First international conference on language resources and evaluation (LREC)*. 1998. p. 1373-1376.
- [20] BARLOW, D. H. (2004) *Anxiety and its disorders: The nature and treatment of anxiety and panic*. Guilford press, 2004.
- [21] BARTLETT, M.S., LITTLEWORT, G., FRANK, M., *et al.* (2006) Fully automatic facial action recognition in spontaneous behavior. In : *Automatic Face and Gesture Recognition, 2006. FGR 2006. 7th International Conference on*. IEEE, 2006. p. 223-230.
- [22] BATLINER, A., FISCHER, K., HUBER, R., *et al.* (2000) Desperately seeking emotions or: Actors, wizards, and human beings. In : *ISCA Tutorial and Research Workshop (ITRW) on Speech and Emotion*. 2000
- [23] BATLINER, A., STEIDL, S., SCHULLER, B., *et al.* (2011) Whodunnit-searching for the most important feature types signalling emotion-related user states in speech. *Computer Speech & Language*, 2011, vol. 25, no 1, p. 4-28.
- [24] BILLINGS, A. G., MOOS, R. H. (1984). Coping, stress, and resources among adults with unipolar depression. *Journal of Personality and Social Psychology*, 46, 877-891.
- [25] BIRKETT, M. A. (2011) The Trier Social Stress Test protocol for inducing psychological stress. *Journal of visualized experiments: JoVE*, 2011, no 56.
- [26] BISHOP, C. M., *et al.* (2006) *Pattern recognition and machine learning*. New York : springer, 2006.
- [27] BOERSMA, P. (2002) Praat, a system for doing phonetics by computer. *Glot international*, 2002, vol. 5, no 9/10, p. 341-345.
- [28] BOONE, R. T., CUNNINGHAM, J. G. (1998) Children's decoding of emotion in expressive body movement: the development of cue attunement. *Developmental psychology*, 1998, vol. 34, no 5, p. 1007.
- [29] BOUCHARD, S, ST JACQUES, J, COTE, S, *et al.* (2003) Exemples de l'utilisation de la réalité virtuelle dans le traitement des phobies. *Revue francophone de clinique comportementale et cognitive*, 2003, vol. 8, no 4, p. 5-12.
- [30] BRADSKI, G. (2000) The opencv library. *Doctor Dobbs Journal*, 2000, vol. 25, no 11, p. 120-126.
- [31] BRANCO, A., FEKETE, S. MW, RUGOLO, L. MSS, *et al.* (2007) The newborn pain cry: Descriptive acoustic spectrographic analysis. *International journal of pediatric otorhinolaryngology*, 2007, vol. 71, no 4, p. 539-546
- [32] BRANNON, L., FEIST, J. (2009) Defining, Measuring and Managing Stress (ch5 pp 93-131), in *Health Psychology: An Introduction to Behavior and Health*, 7th edition (2009)
- [33] BRENDL, M., ZACCARELLI, R., DEVILLERS, L. (2010) Building a System for Emotions Detection from Speech to Control an Affective Avatar. In : *LREC*. 2010

Bibliographie

- [34] BROWN, B. L., STRONG, W. J., RENCHER, A. C. (1975) Acoustic determinants of perceptions of personality from speech. *International journal of the Sociology of Language*, 1975, vol. 1975, no 6, p. 11-32.
- [35] BRUNSWIK, E., (1956) *Perception and the representative design of psychological experiments* (2nd ed.). Berkeley, CA: University of California Press
- [36] BUHRMESTER, M., KWANG, T., GOSLING, S. D. (2011) Amazon's Mechanical Turk a new source of inexpensive, yet high-quality, data?. *Perspectives on Psychological Science*, 2011, vol. 6, no 1, p. 3-5.
- [37] BURKE, H. M., DAVIS, M. C., OTTE, C., et al. (2005) Depression and cortisol responses to psychological stress: a meta-analysis. *Psychoneuroendocrinology*, 2005, vol. 30, no 9, p. 846-856.
- [38] CALLEJAS, Z., LÓPEZ-CÓZAR, R. (2008) Influence of contextual information in emotion annotation for spoken dialogue systems. *Speech Communication*, 2008, vol. 50, no 5, p. 416-433
- [39] CALLEJAS, Z., LÓPEZ-CÓZAR, R. (2009) Improving acceptability assessment for the labelling of affective speech corpora. In : *INTERSPEECH*. 2009. p. 2863-2866.
- [40] CAMURRI, A., LAGERLÖF, I., VOLPE, G. (2003) Recognizing emotion from dance movement: comparison of spectator recognition and automated techniques. *International journal of human-computer studies*, 2003, vol. 59, no 1, p. 213-225.
- [41] CAMURRI, A., MAZZARINO, B., & VOLPE, G. (2004). "Analysis of expressive gesture: The eyesweb expressive gesture processing library." In *Gesture-based communication in human-computer interaction* (pp. 460-467). Springer Berlin Heidelberg.
- [42] CANNIZZARO, M. S., COHEN, H., RAPPARD, F., et al. (2005) Bradyphrenia and bradykinesia both contribute to altered speech in schizophrenia: a quantitative acoustic study. *Cognitive and behavioral neurology*, 2005, vol. 18, no 4, p. 206-210.,
- [43] CANNON, W (1932) *Wisdom of the Body*
- [44] CARIDAKIS, G., WAGNER, J., RAOUZAIIOU, A., et al. (2013) A cross-cultural, multimodal, affective corpus for gesture expressivity analysis. *Journal on Multimodal User Interfaces*, 2013, vol. 7, no 1-2, p. 121-134.
- [45] CARLSON, N R. (2012) *Physiology of Behavior* 11th Edition. Pearson, 2012
- [46] CARVER C.S., (1997) You want to measure coping but your protocol's too long: consider the brief COPE, in *Int. J. Behav. Med.* vol 4, issue 1, pp. 92-100, 1997.
- [47] CARVER, C. S., SCHEIER, M. F., WEINTRAUB, J. K. (1989). Assessing coping strategies: A theoretically based approach. *Journal of Personality and Social Psychology*, 56 (2), 267-283.
- [48] CASALE, S., RUSSO, A., SCEBBA, G., *et al.* (2008) Speech emotion classification using machine learning algorithms. In : *Semantic Computing, 2008 IEEE International Conference on*. IEEE, 2008. p. 158-165.
- [49] CASSELL J, (1998) A framework for gesture generation and interpretation. In R. Cipolla and A. Pentland, editors, *Computer Vision in Human-Machine Interaction*, New York, 1998. Cambridge University Press

Bibliographie

- [50] CASSELL J, PELACHAUD C, BADLER N, STEEDMAN M, ACHORN B, BECKET T, DOUVILLE B, PREVOST S, STONE M. (1994) Animated conversation: Rule-based generation of facial expression gesture and spoken intonation for multiple conversational agents. In Andrew Glassner, editor, Proceedings of SIGGRAPH '94, Computer Graphics Proceedings, Annual Conference Series, pages 413-420. ACM SIGGRAPH, ACM Press, July 1994.
- [51] CASTELLANO, G., KESSOUS, L., CARIDAKIS, G. (2008) Emotion recognition through multiple modalities: face, body gesture, speech. In : Affect and emotion in human-computer interaction. Springer Berlin Heidelberg, 2008. p. 92-103.
- [52] CASTELLANO, G., VILLALBA, S. D., & CAMURRI, A. (2007). Recognising human emotions from body movement and gesture dynamics. In Affective computing and intelligent interaction (pp. 71-82). Springer Berlin Heidelberg
- [53] CATTELL, R. B., EBER, H. W. (1972) The sixteen personality factor questionnaire (16PF). Institute for Personality and Ability Testing, Champaign, Illinois, USA, 1972
- [54] CATTELL, H., MEAD, A. (2008). The sixteen personality factor questionnaire (16PF). In G. Boyle, G. Matthews, & D. Saklofske (Eds.), The SAGE handbook of personality theory and assessment: Volume 2 – Personality measurement and testing. (pp. 135-160). London: SAGE Publications Ltd
- [55] CHASTAGNOL, C. (2013) Reconnaissance automatique des dimensions affectives dans l'interaction orale homme-machine pour des personnes dépendantes. 2013. Thèse de doctorat. Université Paris Sud-Paris XI
- [56] CHASTAGNOL, C., CLAVEL, C., COURGEON, M., et al. (2014) Designing an emotion detection system for a socially intelligent human-robot interaction. In : Natural Interaction with Robots, Knowbots and Smartphones. Springer New York, 2014. p. 199-211
- [57] CHASTAGNOL, C., DEVILLERS, L. (2012) Personality traits detection using a parallelized modified SFFS algorithm. computing, 2012, vol. 15, p. 16.
- [58] CHAVHAN, Y., DHORE, M. L., YESAWARE, P. (2010) Speech emotion recognition using support vector machine. International Journal of Computer Applications, 2010, vol. 1, no 20, p. 6-9.
- [59] CHETOUANI, M., MAHDHAOUI, A., RINGEVAL, F. (2009) Time-scale feature extractions for emotional speech characterization. Cognitive Computation, 2009, vol. 1, no 2, p. 194-201.
- [60] CICHETTI, D. V., FEINSTEIN, A. R. (1990) High agreement but low kappa: II. Resolving the paradoxes. Journal of clinical epidemiology, 1990, vol. 43, no 6, p. 551-558
- [61] CLARK, D. M. (2005) A cognitive perspective on social phobia. *The essential handbook of social anxiety for clinicians*, 2005, p. 193-218.
- [62] CLAVEL, C. (2007) Analyse et reconnaissance des manifestations acoustiques des émotions de type peur en situations anormales. 2007. Thèse de doctorat. Télécom ParisTech.
- [63] COAN, J. A., ALLEN, J. JB. (2007) Handbook of emotion elicitation and assessment. Oxford university press, 2007.
- [64] COHEN, J. (1968) Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. Psychological bulletin, 1968, vol. 70, no 4, p. 213.
- [65] COHEN, S., KESSLER, R. C., GORDON, L. (1997) *Measuring stress: A guide for health and social scientists*. Underwood (ed.) Oxford University Press, 1997.

Bibliographie

- [66] CONNOR, K., M., DAVIDSON, J. RT, CHURCHILL, L. E., et al. (2000) Psychometric properties of the Social Phobia Inventory (SPIN) New self-rating scale. *The British Journal of Psychiatry*, 2000, vol. 176, no 4, p. 379-386.
- [67] COSTA, P. T., MCCRAE, R. R. (2008) The revised neo personality inventory (neo-pi-r). *The SAGE handbook of personality theory and assessment*, 2008, vol. 2, p. 179-198
- [68] CÔTÉ, S., BOUCHARD, S. (2008) Virtual reality exposure for phobias: A critical review. *Journal of CyberTherapy & Rehabilitation*, 2008, vol. 1, no 1, p. 75-91.
- [69] COURGEON, M. (2011) Marc: modèles informatiques des émotions et de leurs expressions faciales pour l'interaction Homme-machine affective temps réel. 2011. Thèse de doctorat. Université Paris Sud-Paris XI.
- [70] COWIE, R., DOUGLAS-COWIE, E., TSAPATSOU LIS, N., VOTSIS, G., KOLLIAS, S., FELLE NZ, W., and TAYLOR, J. (2001). Emotion recognition in human-computer interaction. *IEEE Signal Processing Magazine*, 18(1), 32-80 2001
- [71] D'MELLO S, KORY, J. (2012) Consistent but modest: a meta-analysis on unimodal and multimodal affect detection accuracies from 30 studies. In *Proceedings of the 14th ACM international conference on Multimodal interaction (ICMI '12)*. ACM, New York, NY, USA, 31-38.
- [72] DARWIN C (1872), *The Expression of the Emotions in Man and Animals*, John Murray, London, , p. 273 n. 17.
- [73] DAVIDSON, R. J., MARSHALL, J. R., TOMARKEN, A. J., et al. (2000) While a phobic waits: regional brain electrical and autonomic activity in social phobics during anticipation of public speaking. *Biological psychiatry*, 2000, vol. 47, no 2, p. 85-95.
- [74] DE MEIJER, M. (1989) The contribution of general features of body movement to the attribution of emotions. *Journal of Nonverbal behavior*, 1989, vol. 13, no 4, p. 247-268
- [75] DEDOVIC, K., RENWICK, R., MAHANI, N.K. et al. (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*, 2005, vol. 30, no 5, p. 319.
- [76] DEMULIER, V., BEVACQUA, E., FOCONE, F., et al. (2014) A Database of Full Body Virtual Interactions Annotated with Expressivity Scores. *LREC 2014: 3505-3510*
- [77] DEVILLERS, L., COWIE, R., MARTIN, J. C., et al. (2006) Real life emotions in French and English TV video clips: an integrated annotation protocol combining continuous and discrete approaches. In : *5th international conference on Language Resources and Evaluation (LREC 2006)*, Genoa, Italy. 2006
- [78] DEVILLERS, L., VIDRASCU, L. (2006) Real-life emotions detection with lexical and paralinguistic cues on human-human call center dialogs. In : *Interspeech*. 2006.
- [79] DEVILLERS L, VIDRASCU L, LAMEL. (2005) Challenges in real-life emotion annotation and machine learning based detection. *Neural Networks*, 2005, 18: 407-422
- [80] DHALL, A, et al. (2012) Collecting large, richly annotated facial-expression databases from movies. 2012.
- [81] DIFEDE, J-A, CUKOR, J, JAYASINGHE, N, et al. (2007) Virtual reality exposure therapy for the treatment of posttraumatic stress disorder following September 11, 2001. *Journal of Clinical Psychiatry*, 2007, vol. 68, no 11, p. 1639-1647.
- [82] DISE-LEWIS, J. E. (1988). *The Life Events and Coping Inventory: An assessment of stress in children*. *Psychosomatic Medicine*, 50, 484-499. 1988

Bibliographie

- [83] DOBSON, K. S. (2009). Handbook of cognitive-behavioral therapies. Ed Guilford Press, 2009.
- [84] DOHRENWEND-SNELL, B. (1985) Social status and responsibility for stressful life events. *Issues in mental health nursing*, 1985, vol. 7, no 1-4, p. 105-127
- [85] DOUGLAS-COWIE, E., CAMPBELL, N., COWIE, R., et al. (2003) Emotional speech: Towards a new generation of databases. *Speech communication*, 2003, vol. 40, no 1, p. 33-60.
- [86] DOUGLAS-COWIE, E., COWIE, R., SNEDDON, I, et al. (2007) The HUMANE database: addressing the collection and annotation of naturalistic and induced emotional data. In : *Affective computing and intelligent interaction*. Springer Berlin Heidelberg, 2007. p. 488-500
- [87] EDWARDS, J., JACKSON, H. J., PATTISON, P. E. (2002) Emotion recognition via facial expression and affective prosody in schizophrenia: a methodological review. *Clinical psychology review*, 2002, vol. 22, no 6, p. 789-832.
- [88] EKMAN P, FRIESEN W. (1978) *The Facial Action Coding System: A Technique For The Measurement of Facial Movement*. Consulting Psychologists Press, Inc., San Francisco, CA, 1978
- [89] EKMAN, P. (1992) Are there basic emotions?. *Psychol Rev*, 1992 99(3):550-553
- [90] EKMAN P, FRIESEN W.V. (1976), Measuring facial movement. *Environmental psychology and nonverbal behavior*, vol. 1, no 1, p. 56-75.
- [91] EL AYADI, M., KAMEL, M. S., KARRAY, F. (2011) Survey on speech emotion recognition: Features, classification schemes, and databases. *Pattern Recognition*, 2011, vol. 44, no 3, p. 572-587.
- [92] ENDLER, N. S., & PARKER, J. D. A. (1990). *Coping Inventory for Stressful Situations (CISS): Manual*. Toronto: Multi Health Systems.
- [93] EUROFOUND, enquête européenne sur les conditions de travail, 2005
- [94] EYBEN, F., WÖLLMER, M., GRAVES, A., et al. (2010) On-line emotion recognition in a 3-D activation-valence-time continuum using acoustic and linguistic cues. *Journal on Multimodal User Interfaces*, 2010, vol. 3, no 1-2, p. 7-19
- [95] EYBEN, F., WOLLMER, M., SCHULLER, B. (2009) OpenEAR—introducing the Munich open-source emotion and affect recognition toolkit. In : *Affective Computing and Intelligent Interaction and Workshops*, 2009. ACII 2009. 3rd International Conference on. IEEE, 2009. p. 1-6.
- [96] EYBEN F, WÖLLMER M, SCHULLER B (2010), "openSMILE - The Munich Versatile and Fast Open-Source Audio Feature Extractor", In *Proc. ACM Multimedia (MM)*, ACM, Florence, Italy, ACM, ISBN 978-1-60558-933-6, pp. 1459-1462, October 2010.
- [97] EYSENCK, Hans Jurgen. *Eysenck personality inventory*. San Diego : Educational and Industrial Testing Service, 1968.
- [98] FAN, R-E, CHEN, P-H, LIN, C-J. (2005) Working set selection using second order information for training support vector machines. *The Journal of Machine Learning Research*, 2005, vol. 6, p. 1889-1918.
- [99] FAUR, C., CLAVEL, C., PESTY, S., et al. (2013) PERSEED: a Self-based Model of Personality for Virtual Agents Inspired by Socio-cognitive Theories. In : *Affective Computing and Intelligent Interaction (ACII)*, 2013 Humaine Association Conference on. IEEE, 2013. p. 467-472

Bibliographie

- [100] FEIFEL, H., & STRACK, S. (1989) Coping with conflict situations: Middle-aged and elderly men. *Psychology and Aging*, 4(1), 26-33.
- [101] FEINSTEIN, A. R., CICCETTI, D. V. (1990) High agreement but low kappa: I. The problems of two paradoxes. *Journal of clinical epidemiology*, 1990, vol. 43, no 6, p. 543-549
- [102] FERNANDEZ, R., PICARD, R. W. (2003) Modeling drivers' speech under stress. *Speech Communication*, 2003, vol. 40, no 1, p. 145-159.
- [103] FINK, B., NEAVE, N., MANNING, J. T., et al. (2005) Facial symmetry and the 'big-five' personality factors. *Personality and Individual Differences*, 2005, vol. 39, no 3, p. 523-529.
- [104] FOLKMAN, S., LAZARUS, R. S. (1988). *Manual for the Ways of Coping Questionnaire*. Palo Alto, CA: Consulting Psychologists Press.
- [105] FOLKMAN, S, LAZARUS, R S. (1988a****) The relationship between coping and emotion: Implications for theory and research. *Social Science & Medicine*, 1988, vol. 26, no 3, p. 309-317
- [106] FOLKMAN, S, LAZARUS, R S., GRUEN, R. J., et al. (1986) Appraisal, coping, health status, and psychological symptoms. *Journal of personality and social psychology*, 1986, vol. 50, no 3, p. 571.
- [107] GARCIA-PALACIOS, A., BOTELLA, C., HOFFMAN, H., et al. (2007) Comparing acceptance and refusal rates of virtual reality exposure vs. in vivo exposure by patients with specific phobias. *Cyberpsychology & behavior*, 2007, vol. 10, no 5, p. 722-724
- [108] GILBOA-SCHECHTMAN, E., FOA, E. B., AMIR, N. (1999) Attentional biases for facial expressions in social phobia: The face-in-the-crowd paradigm. *Cognition & Emotion*, 1999, vol. 13, no 3, p. 305-318.
- [109] GILROY, S. W., CAVAZZA, M. O., VERVONDEL, V. (2011) Evaluating multimodal affective fusion using physiological signals. In : *Proceedings of the 16th international conference on Intelligent user interfaces*. ACM, 2011. p. 53-62
- [110] GOLDBERG, L. R. (1990) An alternative" description of personality": the big-five factor structure. *Journal of personality and social psychology*, 1990, vol. 59, no 6, p. 1216.
- [111] GORMAN, J.M. (2000) *Anxiety Disorders Kaplan & Sadock's Comprehensive Textbook of Psychiatry, Seventh Edition* Baltimore: Williams & Wilkins
- [112] GRATCH J., MARSELLA, S. (2004) *Evaluating a General Model of Emotional Appraisal and Coping* AAAI Spring Symposium on Architectures for Modeling Emotion: Cross-disciplinary Foundations, Palo Alto, CA 2004
- [113] GROSS, J. J., LEVENSON, R. W. (1995) Emotion elicitation using films. *Cognition & Emotion*, 1995, vol. 9, no 1, p. 87-108.
- [114] GUNES, H., PICCARDI, M. (2007) Bi-modal emotion recognition from expressive face and body gestures. *Journal of Network and Computer Applications*, 2007, vol. 30, no 4, p. 1334-1345
- [115] HAGENAARS, M. A., & VAN MINNEN, A. (2005). The effect of fear on paralinguistic aspects of speech in patients with panic disorder with agoraphobia. *Journal of Anxiety Disorders*, 19, 521-537
- [116] HALL, M., FRANK, E., HOLMES, G., et al. (2009) The WEKA data mining software: an update. *ACM SIGKDD explorations newsletter*, 2009, vol. 11, no 1, p. 10-18.

Bibliographie

- [117] HAMILTON, M. (1960) A rating scale for depression. *Journal of neurology, neurosurgery, and psychiatry*, 1960, vol. 23, no 1, p. 56.
- [118] HANSEN, J. HL, BOU-GHAZALE, S. E., SARIKAYA, R., et al. (1997) Getting started with SUSAS: a speech under simulated and actual stress database. In : *EUROSPEECH*. 1997. p. 1743-46.
- [119] HARRIGAN, J. A., WILSON, K., ROSENTHAL, R. (2004). Detecting state and trait anxiety from auditory and visual cues: A meta-analysis. *Personality and Social Psychology Bulletin*, 30, 56-66
- [120] HARMON, D. K., MASUDA, M., HOLMES, T. H. (1970) The social readjustment rating scale: a cross-cultural study of Western Europeans and Americans. *Journal of Psychosomatic Research*, 1970, vol. 14, no 4, p. 391-400
- [121] HATHAWAY, S. R., McKINLEY, J. C. (1990) Minnesota multiphasic personality inventory (MMPI). *Gender roles: a handbook of tests and measures*, 1990, vol. 41, no 4, p. 142.
- [122] HE, L. (2010) Stress and emotion recognition in natural speech in the work and family environments. 2010. Thèse de doctorat. RMIT University.
- [123] HEALEY, J. A., PICARD, R. W. (2005) Detecting stress during real-world driving tasks using physiological sensors. *Intelligent Transportation Systems, IEEE Transactions on*, 2005, vol. 6, no 2, p. 156-166
- [124] HEIDER, F., SIMMEL, M. (1944) An experimental study of apparent behavior. *The American Journal of Psychology*, 1944, p. 243-259.
- [125] HEIMBERG, R. G., HOLT, C. S., SCHNEIER, F. R., et al. (1993) The issue of subtypes in the diagnosis of social phobia. *Journal of Anxiety Disorders*, 1993, vol. 7, no 3, p. 249-269.
- [126] HEIMBERG, R. G., MUELLER, G. P., HOLT, C. S., et al. (1993) Assessment of anxiety in social interaction and being observed by others: The Social Interaction Anxiety Scale and the Social Phobia Scale. *Behavior Therapy*, 1993, vol. 23, no 1, p. 53-73.
- [127] HENDIN H.M., CHEEK J.M., (1997) "Assessing Hypersensitive Narcissism: A Re-examination of Murray's Narcissism Scale", in *Journal of Research in Personality*, vol 31, issue 4, 588-599, 1997.
- [128] HEYLEN, D., BEVACQUA, E., PELACHAUD, C., et al. (2011) Generating listening behaviour. In : *Emotion-Oriented Systems*. Springer Berlin Heidelberg, 2011. p. 321-347.
- [129] HEYLEN, Dirk, POEL, Mannes, NIJHOLT, Anton, et al. Generation of facial expressions from emotion using a fuzzy rule based system. In : *AI 2001: Advances in Artificial Intelligence*. Springer Berlin Heidelberg, 2001. p. 83-94
- [130] HOFMANN, S. G., ASNAANI, A., VONK, I. JJ, et al. (2012) The efficacy of cognitive behavioral therapy: a review of meta-analyses. *Cognitive therapy and research*, 2012, vol. 36, no 5, p. 427-440.
- [131] HOFMANN, S. G., BARLOW, D. H. (2002) Social phobia (social anxiety disorder). *Anxiety and its disorders: The nature and treatment of anxiety and panic*, 2002, vol. 2, p. 454-476.
- [132] HOFMANN, S. G., DIBARTOLO, P. M. (2000) An instrument to assess self-statements during public speaking: Scale development and preliminary psychometric properties. *Behavior Therapy*, 2000, vol. 31, no 3, p. 499-515.

Bibliographie

- [133] HOFSTEE, W. K., DE RAAD, B., GOLDBERG, L. R. (1992) Integration of the big five and circumplex approaches to trait structure. *Journal of Personality and Social Psychology*, 1992, vol. 63, no 1, p. 146
- [134] HOLMES, R., OATES, J., PHYLAND, D., et al. (2000) Voice characteristics in the progression of Parkinson's disease. *International Journal of Language & Communication Disorders*, 2000, vol. 35, no 3, p. 407-418.
- [135] HOLMES, D., RAHE, R. (1967). The Social Readjustment Rating Scale. *Journal of Psychosomatic Research*, 11:213-218
- [136] HOYT, C. L., BLASCOVICH, J., SWINTH, K. R. (2003) Social inhibition in immersive virtual environments. *Presence: Teleoperators and Virtual Environments*, 2003, vol. 12, no 2, p. 183-195
- [137] HSU, C-W, CHANG, C-C, LIN, C-J, et al. (2003) A practical guide to support vector classification. 2003
- [138] INRS, rapport de la commission des affaires sociales sur le mal-être au travail, 2009
- [139] IVANOV, A., CHEN, X. (2012) Modulation Spectrum Analysis for Speaker Personality Trait Recognition. In : *INTERSPEECH*. 2012
- [140] IZARD C. E. (1977) *Human emotions*. Springer, 1977.
- [141] IZARD, C. E., LIBERO, D. Z., PUTNAM, P., et al. (1993) Stability of emotion experiences and their relations to traits of personality. *Journal of personality and social psychology*, 1993, vol. 64, no 5, p. 847
- [142] JACKSON, C., NISSENSON, K., CLOITRE, M., et al. (2009) Cognitive-behavioral therapy. *Treating complex traumatic stress disorders: An evidence-based guide*, 2009, p. 243-263.
- [143] JAIMES, A., SEBE, N. (2005) Multimodal human computer interaction: A survey. In : *Computer vision in human-computer interaction*. Springer Berlin Heidelberg, 2005. p. 1-15.
- [144] JIANG, B., VALSTAR, M.F., PANTIC, M. (2011) Action unit detection using sparse appearance descriptors in space-time video volumes. In : *Automatic Face & Gesture Recognition and Workshops (FG 2011)*, 2011 IEEE International Conference on. IEEE, 2011. p. 314-321.
- [145] JOHN, O. P., DONAHUE, E. M., KENTLE, R. L. (1991) *The big five inventory: Versions 4a and 54*, institute of personality and social research. University of California, Berkeley, CA, 1991
- [146] JOORMANN, J., GOTLIB, I. H. (2006) Is this happiness I see? Biases in the identification of emotional facial expressions in depression and social phobia. *Journal of abnormal psychology*, 2006, vol. 115, no 4, p. 705.
- [147] KAPOOR, A., PICARD, R. W. (2005) Multimodal affect recognition in learning environments. In : *Proceedings of the 13th annual ACM international conference on Multimedia*. ACM, 2005. p. 677-682.
- [148] KAPUR, A., KAPUR, A., VIRJI-BABUL, N., et al. (2005) Gesture-based affective computing on motion capture data. In : *Affective Computing and Intelligent Interaction*. Springer Berlin Heidelberg, 2005. p. 1-7.
- [149] KASL, S. V., MAHL, G. F. (1965). The relationship of disturbances and hesitations in spontaneous speech to anxiety. *Journal of Personality and Social Psychology*, 1, 425-433

Bibliographie

- [150] KEINAN, G., FRIEDLAND, N., KAHNEMAN, D., et al. (1999) The effect of stress on the suppression of erroneous competing responses. *Anxiety, stress, and coping*, 1999, vol. 12, no 4, p. 455.
- [151] KEMPER, Theodore D. Social constructionist and positivist approaches to the sociology of emotions. *American Journal of Sociology*, 1981, p. 336-362.
- [152] KESSLER, R. C., CHIU, W. T., DEMLER, O., et al. (2005) Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of general psychiatry*, 2005, vol. 62, no 6, p. 617-627.
- [153] KESSOUS, L., CASTELLANO, G., CARIDAKIS, G. (2010) Multimodal emotion recognition in speech-based interaction using facial expression, body gesture and acoustic analysis. *Journal on Multimodal User Interfaces*, 2010, vol. 3, no 1-2, p. 33-48.
- [154] KIRSCHBAUM, PIRKE, HELLHAMMER, (1993) The Trier Social Stress Test - A Tool for Investigating Psychobiological Stress Response in a Laboratory Setting, 1993
- [155] KIRSCHBAUM, C., WÜST, S., et HELLHAMMER, D. (1992) Consistent sex differences in cortisol responses to psychological stress. *Psychosomatic medicine*, 1992, vol. 54, no 6, p. 648-657.
- [156] KLEINSMITH, A., BIANCHI-BERTHOUBE, N. (2013) Affective body expression perception and recognition: A survey. *Affective Computing, IEEE Transactions on*, 2013, vol. 4, no 1, p. 15-33.
- [157] KLEINSMITH, A., DE SILVA, P. R., BIANCHI-BERTHOUBE, N. (2006) Cross-cultural differences in recognizing affect from body posture. *Interacting with Computers*, 2006, vol. 18, no 6, p. 1371-1389.
- [158] KLINGBEIL, D. A. (2008) A Review of the Five Factor Personality Inventory-Children. *Assessment for Effective Intervention*, 2008
- [159] KLINGER, E., BOUCHARD, S., LÉGERON, P., et al. (2005) Virtual reality therapy versus cognitive behavior therapy for social phobia: A preliminary controlled study. *Cyberpsychology & behavior*, 2005, vol. 8, no 1, p. 76-88
- [160] KOMAROFF, A. L., MASUDA, M., HOLMES, T. H. (1968) The social readjustment rating scale: A comparative study of Negro, Mexican and white Americans. *Journal of Psychosomatic Research*, 1968, vol. 12, no 2, p. 121-128.
- [161] KOOB, G. F. (2008). A role for brain stress systems in addiction. *Neuron*, 59(1), 11-34.
- [162] KROHNE, H. W. (2001). Stress and coping theories. *The international encyclopedia of the social and behavioral sciences*, 22, 15163-15170.
- [163] KROHNE, H. W., SCHUMACHER, A., & EGLOFF, B. (1992). Das Angstbewältigungs-Inventar (ABI) (The Mainz Coping Inventory (MCI)) (Mainzer Berichte zur Persönlichkeitsforschung). Mainz: Johannes-Gutenberg-Universität, Psychologisches Institut.
- [164] KRUMHUBER E. & SCHERER, K. (2011) Affect bursts: Dynamic patterns of facial expressions, *journal of Emotion*. vol. 11, No 4 (pp 742), 2011
- [165] KUDIELKA, Brigitte M., HELLHAMMER, Dirk H., KIRSCHBAUM, Clemens, et al. Ten years of research with the Trier Social Stress Test—revisited. *Social neuroscience: Integrating biological and psychological explanations of social behavior*, 2007, p. 56-83.
- [166] LABAN, R., ULLMANN, L. (1971) The mastery of movement.

Bibliographie

- [167] LAMBREY, S., JOUVENT, R., ALLILAIRE, J.-F., et al. (2010) Les thérapies utilisant la réalité virtuelle dans les troubles phobiques. In : *Annales Médico-psychologiques, revue psychiatrique*. Elsevier Masson, 2010. p. 44-46.
- [168] LANDIS, J. R., KOCH, G. G. (1977) The measurement of observer agreement for categorical data. *biometrics*, 1977, p. 159-174.
- [169] LANG, P. J. (1985). The cognitive psychophysiology of emotion: Fear and anxiety. In A. H. Tuma & J. D. Maser (Eds.), *Anxiety and the anxiety disorders* (pp. 131-170). Hillsdale, NJ: Lawrence Erlbaum.
- [170] LARSEN, R. J., KETELAAR, T. (1991) Personality and susceptibility to positive and negative emotional states. *Journal of personality and social psychology*, 1991, vol. 61, no 1, p. 132
- [171] LAUKKA, P., JUSLIN, P. N., BRESIN, R. (2005). A dimensional approach to vocal expression of emotion. *Cognition and Emotion*, 19, 633-653
- [172] LAUKKA, P., LINNMAN, C., ÅHS, F., et al. (2008) In a nervous voice: Acoustic analysis and perception of anxiety in social phobics' speech. *Journal of Nonverbal Behavior*, 2008, vol. 32, no 4, p. 195-214
- [173] LAUKKA, P., NEIBERG, D., FORSELL, M., et al. (2011) Expression of affect in spontaneous speech: Acoustic correlates and automatic detection of irritation and resignation. *Computer Speech & Language*, 2011, vol. 25, no 1, p. 84-104.
- [174] LAZARUS, R.S., (1966) *Psychological Stress and the Coping Process*. New York: McGraw-Hill
- [175] LAZARUS, R.S., (1991) *Emotion and adaptation*. New York: Oxford University Press.
- [176] LAZARUS, R.S., (1993) Coping theory and research: past, present, and future. *Psychosomatic medicine*, 55(3), 234-247
- [177] LAZARUS, R.S., FOLKMAN S. (1984) *Stress, Appraisal, and Coping*.
- [178] LAZARUS, R.S., LAUNIER, R. (1978) Stress-related transactions between person and environment. In L A Pervin and M Lewis (Eds), 1978. *Perspectives in Interactional Psychology* (pp. 287-327). New York,: Plenum
- [179] LAZARUS, R.S., SPEISMAN, J. C., MORDKOFF, A. M., et al. (1962) A laboratory study of psychological stress produced by a motion picture film. *Psychological Monographs: General and Applied*, 1962, vol. 76, no 34, p. 1.
- [180] LEFTER, I., ROTHKRANTZ, L. JM, VAN LEEUWEN, D. A., et al. (2011) Automatic stress detection in emergency (telephone) calls. *International Journal of Intelligent Defence Support Systems*, 2011, vol. 4, no 2, p. 148-168.
- [181] LEPINE JP, GASQUET I, KOVÉSS V et al. (2005) Prévalence et co-morbidité des troubles psychiatriques dans la population générale française : résultats de l'étude épidémiologique ESMeD/MHEDEA 2000/ (ESEMED). *Encéphale*, 2005, 31 : 182-194.
- [182] LEWIN, M. R., MCNEIL, D. W., & LIPSON, J. M. (1996). Enduring without avoiding: Pauses and verbal dysfluencies in public speaking fear. *Journal of Psychopathology and Behavioral Assessment*, 18, 387-402
- [183] LIEBOWITZ MR. (1987) Social Phobia. *Mod Probl Pharmacopsychiatry*, 1987, 22:141-173.
- [184] LINGENFELSER, F., WAGNER, J., ANDRÉ, E. (2011) A systematic discussion of fusion techniques for multi-modal affect recognition tasks. In : *Proceedings of the 13th international conference on multimodal interfaces*. ACM, 2011. p. 19-26.

Bibliographie

- [185] LIU, G., LEI, Y., HANSEN, J. HL. (2012) Robust feature front-end for speaker identification. In : Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on. IEEE, 2012. p. 4233-4236.
- [186] LOAS G., PARKER J. D. A., OTMANI O., VERRIER A. and FREMAUX D., (1997) Confirmatory Factor Analysis Of The French Translation Of The 20-Item Toronto Alexithymia Scale, in *Perceptual and Motor Skills*, vol. 85, pp. 1018-1018, 1997
- [187] LU, H., FRAUENDORFER, D., RABBI, M., et al. (2012) StressSense: Detecting stress in unconstrained acoustic environments using smartphones. In : *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*. ACM, 2012. p. 351-360
- [188] LUCEY, P., COHN, J. F., KANADE, T., et al. (2010) The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression. In : *Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2010 IEEE Computer Society Conference on. IEEE, 2010. p. 94-101
- [189] LUCEY, S., MATTHEWS, I., HU, C., et al. (2006) AAM derived face representations for robust facial action recognition. 2006.
- [190] LYNAM, D. R., WIDIGER, T. A. (2001) Using the five-factor model to represent the DSM-IV personality disorders: An expert consensus approach. *Journal of abnormal psychology*, 2001, vol. 110, no 3, p. 401.
- [191] MALBOS, E., BOYER, L., et LANÇON, C. (2013) L'utilisation de la réalité virtuelle dans le traitement des troubles mentaux. *La Presse Médicale*, 2013, vol. 42, no 11, p. 1442-1452.
- [192] MANCIA, G., GRASSI, G., POMIDOSSI, G., et al. (1983) Effects of blood-pressure measurement by the doctor on patient's blood pressure and heart rate. *The Lancet*, 1983, vol. 322, no 8352, p. 695-698
- [193] MARKS, I. M., MATHEWS, A. M. (1979) Brief standard self-rating for phobic patients. *Behaviour research and therapy*, 1979, vol. 17, no 3, p. 263-267.
- [194] MARSELLA, S., GRATCH, J. (2003) Modeling coping behavior in virtual humans: don't worry, be happy. In : *Proceedings of the second international joint conference on Autonomous agents and multiagent systems*. ACM, 2003. p. 313-320
- [195] MASUDA, M., HOLMES, T. H. (1967) The social readjustment rating scale: a cross-cultural study of Japanese and Americans. *Journal of Psychosomatic Research*, 1967, vol. 11, no 2, p. 227-237.
- [196] McCRAE, R. R. (1984). Situational determinants of coping responses: Loss, threat, and challenge. *Journal of Personality and Social Psychology*, 46, 919-928
- [197] McCRAE, R. R., COSTA, P. T., BUSCH, C. M. (1986) Evaluating comprehensiveness in personality systems: The California Q-Set and the five-factor model. *Journal of Personality*, 1986, vol. 54, no 2, p. 430-446
- [198] McDUFF, D., KALIOUBY, R. E., PICARD, R. W. (2012) Crowdsourcing facial responses to online videos. *Affective Computing, IEEE Transactions on*, 2012, vol. 3, no 4, p. 456-468
- [199] McDUFF, D., EL KALIOUBY, R., SENECHAL, T., et al. (2013) Affectiva-MIT Facial Expression Dataset (AM-FED): Naturalistic and Spontaneous Facial Expressions Collected "In-the-Wild". In : *Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2013 IEEE Conference on. IEEE, 2013. p. 881-888
- [200] McGHEE, P. E., (1996) *Humor, healing, and the amuse system: Humor as survival training*. Dubuque, Iowa: Kendall/Hunt Publishing, 1996

Bibliographie

- [201] McKEOWN, G., VALSTAR, M., COWIE, R., et al. (2012) The semaine database: Annotated multimodal records of emotionally colored conversations between a person and a limited agent. *Affective Computing, IEEE Transactions on*, 2012, vol. 3, no 1, p. 5-17
- [202] McKINNEY, M. E., MINER, M. H., RÜDDEL, H., et al. (1985) The Standardized Mental Stress Test Protocol: Test-Retest Reliability and Comparison with Ambulatory Blood Pressure Monitoring. *Psychophysiology*, 1985, vol. 22, no 4, p. 453-463.
- [203] McLAY, R. N., WOOD, D. P., WEBB-MURPHY, J. A., et al. (2011) A randomized, controlled trial of virtual reality-graded exposure therapy for post-traumatic stress disorder in active duty service members with combat-related post-traumatic stress disorder. *Cyberpsychology, behavior, and social networking*, 2011, vol. 14, no 4, p. 223-229
- [204] McNALLY, R. J., ENGLISH, G. E., LIPKE, H. J. (1993) Assessment of intrusive cognition in PTSD: Use of the modified Stroop paradigm. *Journal of Traumatic Stress*, 1993, vol. 6, no 1, p. 33-41.
- [205] McNEILL D. (1995) *Hand and Mind: What Gestures Reveal About Thought*. University of Chicago Press, Chicago, 2nd edition, 1995
- [206] MICHEL, P., EL KALIOUBY, R. (2003) Real time facial expression recognition in video using support vector machines. In : *Proceedings of the 5th international conference on Multimodal interfaces*. ACM, 2003. p. 258-264
- [207] MIKHAILOVA, E. S., VLADIMIROVA, T. V., IZNAK, A. F., et al. (1996) Abnormal recognition of facial expression of emotions in depressed patients with major depression disorder and schizotypal personality disorder. *Biological psychiatry*, 1996, vol. 40, no 8, p. 697-705.
- [208] MILLER, S. M. (1987) Monitoring and blunting: Validation of a questionnaire to assess styles of information seeking under threat. *Journal of Personality and Social Psychology*, 52 (2), 345-353.
- [209] MILLER, L. H., SMITH, A. D. (1994) *The stress solution: an action plan to manage the stress in your life*. sl. 1994.
- [210] MIROWSKY, John et ROSS, Catherine E. Social pattern of distress. *Annual review of sociology*, 1986, p. 23-45
- [211] MOHAMMADI, G., VINCIARELLI, A., et MORTILLARO, M. (2010) The voice of personality: Mapping nonverbal vocal behavior into trait attributions. In : *Proceedings of the 2nd international workshop on Social signal processing*. ACM, 2010. p. 17-20
- [212] MONTACIÉ, C., CARATY, M-J. (2012) Pitch and Intonation Contribution to Speakers' Traits Classification. In : *INTERSPEECH. 2012*
- [213] MORRIS, R. (2011) Crowdsourcing workshop : the emergence of affective crowdsourcing. In *CHI Workshop on Crowdsourcing and Human Computation*, (Vancouver, Canada), 7-12 May 2011
- [214] MORRIS, A. C., JASSIM, S., SELLAHEWA, H., ALLANO, L., EHLERS, J., WU, D., KOREMAN J, GARCIA-SALICETTI S, LY-VAN B, DORIZZI, B. (2006). Multimodal person authentication on a smartphone under realistic conditions. In *Proceedings of SPIE Vol. 6250*, pp. 120-131, april 2006
- [215] MUNDT, J. C., SNYDER, P. J., CANNIZZARO, M. S., et al. (2007) Voice acoustic measures of depression severity and treatment response collected via interactive voice response (IVR) technology. *Journal of neurolinguistics*, 2007, vol. 20, no 1, p. 50-64
- [216] MYERS-BRIGGS, I. (1962) *The Myers-Briggs Type Indicator: Manual* 1962.

Bibliographie

- [217] NEGRAO, A. B., DEUSTER, P. A., GOLD, P. W., et al. (2000) Individual reactivity and physiology of the stress response. *Biomedicine & pharmacotherapy*, 2000, vol. 54, no 3, p. 122-128
- [218] OCHS, M., PELACHAUD, C., SADEK, D. (2008) An empathic virtual dialog agent to improve human-machine interaction. In : *Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems 2008*. p. 89-96
- [219] ÖHMAN, A. (2008) Fear and anxiety. *Handbook of Emotions*, 2008, p. 709.
- [220] OLSON, K. R., & WEBER, D. A. (2004). Relations between big five traits and fundamental motives. *Psychological Reports*, 95(3), 795-802.
- [221] OPENNI CONSORTIUM, OpenNI, the standard framework for 3D sensing. <http://openni.org/>. [Accessed: 2013-04-15]
- [222] OPRIS, David, PINTEA, Sebastian, GARCÍA-PALACIOS, Azucena, et al. Virtual reality exposure therapy in anxiety disorders: a quantitative meta-analysis. *Depression and anxiety*, 2012, vol. 29, no 2, p. 85-93
- [223] ÖST, L. G. (1997) Rapid treatment of specific phobias. *Phobias: A handbook of theory, research, and treatment*, 1997, p. 227-247.
- [224] OUDEYER P-Y, (2003) The production and recognition of emotions in speech: features and algorithms. *International Journal of Human-Computer Studies*, 2003, vol. 59, no 1, p. 157-183.
- [225] PANTIC M , ROTHKRANTZ L.J.M. (2000) Automatic analysis of facial expressions: The state of the art. *Pattern Analysis and Machine Intelligence*, *IEEE Transactions on*, vol. 22, no 12, p. 1424-1445.
- [226] PARSONS, T. D., RIZZO, A. A. (2008) Affective outcomes of virtual reality exposure therapy for anxiety and specific phobias: A meta-analysis. *Journal of behavior therapy and experimental psychiatry*, 2008, vol. 39, no 3, p. 250-261.
- [227] PATEL, S., SCHERER, K R., BJÖRKNER, E., et al. (2011) Mapping emotions into acoustic space: the role of voice production. *Biological psychology*, 2011, vol. 87, no 1, p. 93-98.
- [228] PATTERSON, J. M., McCUBBIN, H. I. (1987). Adolescent coping style and behaviors: conceptualization and measurement. *Journal of Adolescence*, 10, 163-186.
- [229] PEACOCK E., WONG P.,(1990) The stress appraisal measure (SAM): A multidimensional approach to cognitive appraisal. *Stress Medicine*, vol. 6, pp. 227-236, 1990.
- [230] PEARLIN, L. I. (1989) The sociological study of stress. *Journal of health and social behavior*, 1989, p. 241-256
- [231] PELACHAUD, C. (2009) Modelling multimodal expression of emotion in a virtual agent. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 2009, vol. 364, no 1535, p. 3539-3548
- [232] PELISSOLO A., LOBJOIE C., MONTEFIORE D. (2010) Validation du questionnaire d'éreutophobie de la Salpêtrière, *La Revue canadienne de psychiatrie*, vol 55, no 9, septembre 2010
- [233] PERKINS, A. M., INCHLEY-MORT, S. L., PICKERING, A. D., et al. (2012) A facial expression for anxiety. *Journal of personality and social psychology*, 2012, vol. 102, no 5, p. 910.
- [234] PERREZ, M., & REICHERTS, M. (1992). *Stress, coping, and health*. Seattle/Toronto: Hogrefe & Huber Publishers

Bibliographie

- [235] PERTAUB, D-P, SLATER, M, BARKER, C. (2002) An experiment on public speaking anxiety in response to three different types of virtual audience. *Presence: Teleoperators and virtual environments*, 2002, vol. 11, no 1, p. 68-78.
- [236] PHILIPPOT, P. (1993) Inducing and assessing differentiated emotion-feeling states in the laboratory. *Cognition & Emotion*, 1993, vol. 7, no 2, p. 171-193.
- [237] PIANA, S., STAGLIANÒ, A., ODONE, F., VERRI, A., & CAMURRI, A. (2014). Real-time Automatic Emotion Recognition from Body Gestures. arXiv preprint arXiv:1402.5047
- [238] PIANESI, F., MANA, N., CAPPELLETTI, A., et al. (2008) Multimodal recognition of personality traits in social interactions. In : *Proceedings of the 10th international conference on Multimodal interfaces*. ACM, 2008. p. 53-60
- [239] PICARD, Rosalind Wright. *Affective computing*. 1995
- [240] POH, M-Z, MCDUFF, DI J., PICARD, R. W. (2011) Advancements in noncontact, multiparameter physiological measurements using a webcam. *Biomedical Engineering, IEEE Transactions on*, 2011, vol. 58, no 1, p. 7-11
- [241] POPE, B., BLASS, T., SIEGMAN, A. W., & RAHER, J. (1970). Anxiety and depression in speech. *Journal of Consulting and Clinical Psychology*, 35, 128-133
- [242] POWERS, M. B., EMMELKAMP, P. M.G. (2008) Virtual reality exposure therapy for anxiety disorders: A meta-analysis. *Journal of anxiety disorders*, 2008, vol. 22, no 3, p. 561-569.
- [243] RAMMSTEDT, B., JOHN, O. P. (2007) Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of research in Personality*, 2007, vol. 41, no 1, p. 203-212
- [244] RAPEE, R. M., HEIMBERG, R. G. (1997) A cognitive-behavioral model of anxiety in social phobia. *Behaviour research and therapy*, 1997, vol. 35, no 8, p. 741-756.
- [245] RATHUS SA. (1973) A 30-item schedule for assessing assertive behavior. *Behavior Therapy*, 1973, 4(3): 398-406
- [246] RAY, Rebecca D. *Emotion elicitation using films. Handbook of emotion elicitation and assessment*, 2007, p. 9.
- [247] REVELLE, W, SCHERER, K R. (2009) *Personality and emotion. Oxford companion to emotion and the affective sciences*, 2009, p. 304-306.
- [248] REZZOUG, N., GORCE, P., HELOIR, A., et al. (2006) Virtual humanoids endowed with expressive communication gestures: the HuGEx project. In : *Systems, Man and Cybernetics, 2006. SMC'06. IEEE International Conference on*. IEEE, 2006. p. 4445-4450
- [249] ROSSIGNOL, M., PHILIPPOT, P., DOUILLIEZ, C., et al. (2005) The perception of fearful and happy facial expression is modulated by anxiety: an event-related potential study. *Neuroscience letters*, 2005, vol. 377, no 2, p. 115-120.
- [250] RUIZ R, ABSIL E, HARMEGNIES B, LEGROS V, POCH D. (1996) Time- and spectrum-related variabilities in stressed speech under laboratory and real conditions. *Speech Communication*, 20(1-2):111 - 129, 1996
- [251] RUSSELL, J. A. (1980) A circumplex model of affect. *Journal of personality and social psychology*, 1980, vol. 39, no 6, p. 1161.
- [252] SANO, A., PICARD, R. W. (2013) Stress recognition using wearable sensors and mobile phones. In : *Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference on*. IEEE, 2013. p. 671-676

Bibliographie

- [253] SARAGIH J. M., LUCEY S., COHN J. F. (2009), Face Alignment through Subspace Constrained Mean-Shifts. International Conference of Computer Vision (ICCV), September, 2009.
- [254] SCHAEFER, A., NILS, F., SANCHEZ, X., et al. (2010) Assessing the effectiveness of a large database of emotion-eliciting films: A new tool for emotion researchers. *Cognition and Emotion*, 2010, vol. 24, no 7, p. 1153-1172.
- [255] SCHEIER MF, WEINTRAUB JK, CARVER CS (1986) Coping with stress: Divergent strategies of optimists and pessimists. *J Pers Soc Psychol* 51:1257-1264, 1986
- [256] SCHERER, K R. (1978) Personality inference from voice quality: The loud voice of extroversion. *European Journal of Social Psychology*, 1978, vol. 8, no 4, p. 467-487
- [257] SCHERER, K. R. (1981) Vocal indicators of stress. In J. Darby (Ed.), *Speech evaluation in psychiatry* (pp. 189-220). New York: Grune & Stratton, 1981.
- [258] SCHERER K. (1984) On the nature and function of emotion: A component process approach," in *Approaches to emotion*, K. R. Scherer and P. Ekman, Eds., 1984, pp. 293-317
- [259] SCHERER, K. R. (1994) Affect Bursts, in *Emotions: Essays on Emotion Theory* (S. H.M. van Goozen, N. E. van de Poll, & J. A. Sergeant, eds.), p. 161-193. Hillsdale, NJ: Lawrence Erlbaum, 1994
- [260] SCHERER, K R. (2003) Vocal communication of emotion: A review of research paradigms. *Speech communication*, 2003, vol. 40, no 1, p. 227-256.
- [261] SCHERER K, GRANDJEAN D, JOHNSTONE T, KLASMEYER G, BÄNZIGER T, (2002) Acoustic correlates of task load and stress. In *Proc. ICSLP2002*, pages 2017-2020
- [262] SCHERER, S., HOFMANN, H., LAMPMANN, M., et al. (2008) Emotion Recognition from Speech: Stress Experiment. In : *LREC*. 2008.
- [263] SCHMIDT, Thomas, DUNCAN, Susan, EHMER, Oliver, et al. An exchange format for multimodal annotations. In : *Multimodal corpora*. Springer Berlin Heidelberg, 2009. p. 207-221
- [264] SCHNEIER, F. R., HECKELMAN, L. R., GARFINKEL, R., et al. (1994) Functional impairment in social phobia. *Journal of Clinical Psychiatry*, 1994.
- [265] SCHNEIDER T. R., RENCH T. A., LYONS J. B and RIFFLE R. R, "The influence of neuroticism, extraversion and openness on stress responses," in *Stress and Health*. vol 28, issue 2, pp. 102-110, 2012.
- [266] SCHRÖDER, M. (2003) Experimental study of affect bursts, *Speech Communication - Special session on speech and emotion*, vol. 40, No 1-2, 2003
- [267] SCHRÖDER, M., DEVILLERS, L., KARPOUZIS, K., et al. (2007) What should a generic emotion markup language be able to represent?. In : *Affective Computing and Intelligent Interaction*. Springer Berlin Heidelberg, 2007. p. 440-451
- [268] SCHULLER, B., BATLINER, A., STEIDL, S., et al. (2011) Recognising realistic emotions and affect in speech: State of the art and lessons learnt from the first challenge. *Speech Communication*, 2011, vol. 53, no 9, p. 1062-1087.
- [269] SCHULLER, B., MÜLLER, R., LANG, M. K., & RIGOLL, G. (2005) Speaker independent emotion recognition by early fusion of acoustic and linguistic features within ensembles. In *INTERSPEECH 2005* (pp. 805-808)
- [270] SCHULLER, B., STEIDL, S., BATLINER, A. (2009) The INTERSPEECH 2009 emotion challenge. In : *INTERSPEECH*. 2009. p. 312-315.

Bibliographie

- [271] SCHULLER, B., STEIDL, S., BATLINER, A., et al. (2010) The INTERSPEECH 2010 paralinguistic challenge. In : INTERSPEECH. 2010. p. 2794-2797.
- [272] SCHULLER, B., STEIDL, S., BATLINER, A., et al. (2011) The INTERSPEECH 2011 Speaker State Challenge. In : INTERSPEECH. 2011. p. 3201-3204.
- [273] SCHULLER, B., STEIDL, S., BATLINER, A., et al. (2012) The INTERSPEECH 2012 Speaker Trait Challenge. In : INTERSPEECH. 2012.
- [274] SCHULLER, B., STEIDL, S., BATLINER, A., et al. (2013) The INTERSPEECH 2013 computational paralinguistics challenge: social signals, conflict, emotion, autism. 2013.
- [275] SCHULLER, B., STEIDL, S., BATLINER, A., et al. (2014) The INTERSPEECH 2014 Computational Paralinguistics Challenge: Cognitive & Physical Load. In : Proceedings of Interspeech. 2014.
- [276] SCHULLER, B., VALSTAR, M., EYBEN., F, et al. (2011) Avec 2011 - the first international audio/visual emotion challenge. In : Affective Computing and Intelligent Interaction. Springer Berlin Heidelberg, 2011. p. 415-424.
- [277] SCHULLER, B., VALSTAR, M., EYBEN., F, et al. (2012) Avec 2012 - the continuous audio/visual emotion challenge. In : Proceedings of the 14th ACM international conference on Multimodal interaction. ACM, 2012. p. 449-456
- [278] SCHULLER, B., ZACCARELLI, R., ROLLET, N., DEVILLERS, L. (2010) CINEMO-A French Spoken Language Resource for Complex Emotions: Facts and Baselines. In : LREC. 2010
- [279] SCHWARZER R.,KNOLL N. (2003) Positive Coping: Mastering Demands and Searching for Meaning, In Lopez, S. J. & Snyder, C. R. (Eds.), Handbook of Positive Psychological Assessment Washington, DC: American Psychological Association, 2003
- [280] SCHWARZER, R. & SCHWARZER, C. (1996). A critical survey of coping instruments. In M. Zeidner & N. S. Endler (Eds.), Handbook of coping: Theory, research and applications (pp. 107-132). New York: Wiley
- [281] SEHILI, M. A., LECOUTEUX, B., VACHER, M., et al. (2012) Sound environment analysis in smart home. In : Ambient Intelligence. Springer Berlin Heidelberg, 2012. p. 208-223
- [282] SELYE, Hans. The stress of life. 1956.
- [283] SELYE, Hans. The nature of stress. Basal Facts, 1985, vol. 7, no 1, p. 3-11.
- [284] SHASHIDHAR G. KOOLAGUDI, K. SREENIVASA B. (2012) Emotion recognition from speech: a review pp 99-117 International Journal of Speech Technology, Vol 15 No2, 2012
- [285] SIM, J., WRIGHT, C. C. (2005) The kappa statistic in reliability studies: use, interpretation, and sample size requirements. Physical therapy, 2005, vol. 85, no 3, p. 257-268
- [286] SIMON-THOMAS, E. R., KELTNER, D. J., SAUTER, D., et al. (2009) The voice conveys specific emotions: evidence from vocal burst displays. Emotion, 2009, vol. 9, no 6, p. 838.
- [287] SKINNER, E A, EDGE, K., ALTMAN, J., et al. (2003) Searching for the structure of coping: a review and critique of category systems for classifying ways of coping. Psychological bulletin, 2003, vol. 129, no 2, p. 216.
- [288] SMITH, G. A. (1977). Voice analysis for the measurement of anxiety. British Journal of Medical Psychology, 50, 367-373

Bibliographie

- [289] SNOW, R., O'CONNOR, B., JURAFSKY, D., et al. (2008) Cheap and fast---but is it good?: evaluating non-expert annotations for natural language tasks. In : Proceedings of the conference on empirical methods in natural language processing. Association for Computational Linguistics, 2008. p. 254-263
- [290] SPIELBERGER C. D., GORSUCH R. L., LUSHENE R., VAGG P. R., JACOBS G. A. (1983) Manual for the State-Trait Anxiety Inventory, in Palo Alto, CA: Consulting Psychologists Press, 1983.
- [291] SPITZER, R. L., KROENKE, K., WILLIAMS, J. BW, et al. (2006) A brief measure for assessing generalized anxiety disorder: the GAD-7. Archives of internal medicine, 2006, vol. 166, no 10, p. 1092-1097
- [292] STEIDL S, LEVIT M, BATLINER A, NOTH E, NIEMANN H, (2005) Of all things the measure is man: Automatic classification of emotions and inter-labeler consistency International Conference on Acoustics, Speech and Signal Processing, pp. 317-320 by
- [293] STONE, A. A., & NEALE, J. M. (1984). New measure of daily coping: Development and preliminary results. Journal of Personality and Social Psychology, 46, 892-906.
- [294] STROOP, J. R. (1935) Studies of interference in serial verbal reactions. Journal of experimental psychology, 1935, vol. 18, no 6, p. 643.
- [295] SUMTER, S. R., BOKHORST, C. L., MIERS, A. C., et al. (2010) Age and puberty differences in stress responses during a public speaking task: Do adolescents grow more sensitive to social evaluation?. Psychoneuroendocrinology, 2010, vol. 35, no 10, p. 1510-1516.
- [296] SUN, R., MOORE II, E. (2011) Investigating glottal parameters and teager energy operators in emotion recognition. In : Affective Computing and Intelligent Interaction. Springer Berlin Heidelberg, 2011. p. 425-434
- [297] SYLVERS, P., LILIENFELD, S. O., & LAPRAIRIE, J. L. (2011). Differences between trait fear and trait anxiety: Implications for psychopathology. Clinical psychology review, 31(1), 122-137
- [298] SWARTZ, L., ELK, R., TEGGIN, A. F., et al. (1983) Life events in Xhosas in Cape Town. Journal of psychosomatic research, 1983, vol. 27, no 3, p. 223-231
- [299] TAHON, M. (2013) Analyse acoustique de la voix émotionnelle de locuteurs lors d'une interaction humain-robot. 2013. Thèse de doctorat. Université Paris-Sorbonne.
- [300] TAHON, M., DELABORDE, A., DEVILLERS, L. (2011) Real-life Emotion Detection from Speech in Human-Robot Interaction: Experiments across diverse Corpora with Child and Adult Voices. Interspeech2011, Sept 2011.
- [301] TAN, E. S. (2013) Emotion and the structure of narrative film: Film as an emotion machine. Routledge, 2013.
- [302] TANG J, ALELYANI S, LIU H (2014) Feature Selection for Classification: A Review, in Data Classification: Algorithms and Applications. Editor: Charu Aggarwal, CRC Press In Chapman & Hall/CRC Data Mining and Knowledge Discovery Series
- [303] TARASOV, A., DELANY, S.J., CULLEN, C. (2010) Using crowdsourcing for labelling emotional speech assets. W3C workshop on Emotion ML, Paris, France. 2010
- [304] TEAGER H. (1980) Some Observations on Oral Air Flow During Phonation. Acoustics, Speech and Signal Processing, IEEE Transactions on, 28(5):599-601, 1980
- [305] TRULL, T. J., WIDIGER, T. A., BURR, R. (2001) A Structured Interview for the Assessment of the Five-Factor Model of Personality: Facet-Level Relations to the Axis II Personality Disorders. Journal of Personality, 2001, vol. 69, no 2, p. 175-198

Bibliographie

- [306] TURKER, B.B., MARZBAN, S., ERZIN, E., et al. (2014) Affect burst recognition using multi-modal cues. In : Signal Processing and Communications Applications Conference (SIU), 2014 22nd. IEEE, 2014. p. 1608-1611.
- [307] TZANETAKIS, G., COOK, P. (2002) Musical genre classification of audio signals. *Speech and Audio Processing*, IEEE transactions on, 2002, vol. 10, no 5, p. 293-302
- [308] ULTES, S., SCHMITT, A., MINKER, W. (2011) Attention, Sobriety Checkpoint! Can Humans Determine by Means of Voice, if Someone is Drunk... and can Automatic Classifiers Compete?. In : INTERSPEECH. 2011. p. 3221-3224
- [309] VALSTAR, M., JIANG, B., MEHU, M., et al. (2011) The first facial expression recognition and analysis challenge. In : Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on. IEEE, 2011. p. 921-926.
- [310] VALSTAR, M. F., PANTIC, M. (2007) Combined support vector machines and hidden markov models for modeling facial action temporal dynamics. In : Human-Computer Interaction. Springer Berlin Heidelberg, 2007. p. 118-127.
- [311] VALSTAR, M., SCHULLER, B., SMITH, K, et al. (2013) AVEC 2013 - the continuous audio/visual emotion and depression recognition challenge. In : Proceedings of the 3rd ACM international workshop on Audio/visual emotion challenge. ACM, 2013. p. 3-10.
- [312] VALSTAR, M., SCHULLER, B., SMITH, K, et al. (2014) AVEC 2014 - 3D Dimensional Affect and Depression Recognition Challenge, In : proc. 4th ACM international workshop on Audio/visual emotion challenge , preprint, 2014
- [313] VERVERIDIS D., KOTROPOULOS C. (2003) A state of the art review on emotional speech databases, Proceedings of the first Richmedia Conference, pages 109-119.
- [314] VERVERIDIS D., KOTROPOULOS C., (2004) Automatic speech classification to five emotional states based on gender information. In: Proc. European Signal Processing Conf. (EUSIPCO '04), 2004. 1: p. 341-344.
- [315] VOGT, T. and ANDRÉ, E. (2005). Comparing feature sets for acted and spontaneous speech in view of automatic emotion recognition. In Proc. Multimedia and Expo (ICME05), pages 474-477, Amsterdam.
- [316] VOGT, T., ANDRÉ, E., BEE, N. (2008) EmoVoice—A framework for online recognition of emotions from voice. In : Perception in multimodal dialogue systems. Springer Berlin Heidelberg, 2008. p. 188-199.
- [317] VOLLRATH M., TORGERSEN S., ALRUES R. (1995) Personality as Long-Term Predictor of Coping, in *Person. individ. Diff: Vol. 18. No. 1*. pp. 117-125, 1995
- [318] VON DAWANS, B., KIRSCHBAUM, C., HEINRICHS, M. (2011) The Trier Social Stress Test for Groups (TSST-G): A new research tool for controlled simultaneous social stress exposure in a group format. *Psychoneuroendocrinology*, 2011, vol. 36, no 4, p. 514-522.
- [319] WAGNER, J., LINGENFELSER, F., ANDRE, E., KIM, J. (2011) Exploring fusion methods for multimodal emotion recognition with missing data. *Affective Computing*, IEEE Transactions on, 2011, vol. 2, no 4, p. 206-218.
- [320] WALLBOTT, H. G. (1998) Bodily expression of emotion. *European journal of social psychology*, 1998, vol. 28, no 6, p. 879-896.
- [321] WATANABE, S., MIKI, K., et KAKIGI, R. (2005) Mechanisms of face perception in humans: A magneto-and electro-encephalographic study. *Neuropathology*, 2005, vol. 25, no 1, p. 8-20.

Bibliographie

- [322] WATSON, D., CLARK, L. A. (1997). Extraversion and its positive emotional core. In R. Hogan, J. Johnson, & S. Briggs (Eds.), *Handbook of personality psychology* (p. 767-793). San Diego, CA: Academic Press.
- [323] WATSON D., CLARK L. A., TELLEGEN A. (1988) "Development and validation of brief measures of positive and negative affect: The PANAS scales," in *J. Pers. Social Psychol.*, vol. 54, pp. 1063-1070, 1988
- [324] WATSON, D., FRIEND, R. (1969) Measurement of social-evaluative anxiety. *Journal of consulting and clinical psychology*, 1969, vol. 33, no 4, p. 448.
- [325] WESTENBERG, P. M., BOKHORST, C. L., MIERS, A. C., et al. (2009) A prepared speech in front of a pre-recorded audience: Subjective, physiological, and neuroendocrine responses to the Leiden Public Speaking Task. *Biological Psychology*, 2009, vol. 82, no 2, p. 116-124.
- [326] WIEDERHOLD, B. K., GEVIRTZ, R., WIEDERHOLD, M. D. (1998) Fear of flying: A case report using virtual reality therapy with physiological monitoring. *CyberPsychology & Behavior*, 1998, vol. 1, no 2, p. 97-103.
- [327] WIEDERHOLD, B. K. WIEDERHOLD, M. D. (2003) Three-year follow-up for virtual reality exposure for fear of flying. *Cyberpsychology & Behavior*, 2003, vol. 6, no 4, p. 441-445
- [328] WILLIAMS, J. M. G., MATHEWS, A., MACLEOD, C. (1996) The emotional Stroop task and psychopathology. *Psychological bulletin*, 1996, vol. 120, no 1, p. 3.
- [329] WILLIAMS, J. M. G., NULTY, D. D. (1986) Construct accessibility, depression and the emotional stroop task: Transient mood or stable structure?. *Personality and Individual Differences*, 1986, vol. 7, no 4, p. 485-491.
- [330] YADAK, J. (2013) Explorations de marqueurs émotionnels vocaux dans la phobie sociale, Mémoire de Master Recherche de Sciences et Technologie Mention Biologie Intégrative et Physiologie, UPMC, septembre 2013
- [331] ZENG, Z., PANTIC, M., ROISMAN, G. I., et al. (2009) A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 2009, vol. 31, no 1, p. 39-58.
- [332] ZHAO, L, BADLER, N. I. (2005) Acquiring and validating motion qualities from live limb gestures. *Graphical Models*, 2005 vol. 67, no 1, p. 1-16.
- [333] ZHOU, G., HANSEN, J., KAISER, J. (1998) Classification of speech under stress based on features derived from the nonlinear Teager energy operator. In: *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, Vol. I, pp. 549-552
- [334] ZHOU, G., HANSEN, J.HL, KAISER, J. F. (2001) Nonlinear feature based classification of speech under stress. *Speech and Audio Processing, IEEE Transactions on*, 2001, vol. 9, no 3, p. 201-216
- [335] ZIGMOND, A. S., SNAITH R. P. (1983) The hospital anxiety and depression scale. *Acta psychiatrica scandinavica* 1983 67.6: 361-370.
- [336] ZUO X, FUNG P (2011), A Cross Gender and Cross Lingual Study of Stress Recognition in Speech Without Linguistic Features, In *The 17th International Congress of Phonetic Sciences (ICPhS XVII)*, Hong Kong, August 2011
- [337] ZUO X, LI T, FUNG P (2012) A Multilingual Database of Natural Stress Emotion, *LREC* 2012,

Publications

Communication avec Actes

[I] SOURY, M. et DEVILLERS, L. *Collecte de données pour la détection du stress dans les interactions sociales*. In : WACAI 2012 Workshop Affect, Compagnon Artificiel, Interaction. p. 175.

[II] GIRAUD, T., SOURY, M., HUA, J. et al. *Multimodal Expressions of Stress during a Public Speaking Task: Collection, Annotation and Global Analyses*. In : Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference on. IEEE, 2013. p. 417-422.

[III] SOURY, M. et DEVILLERS, L. *Stress detection from audio on multiple window analysis size in a public speaking task*. In : Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference on. IEEE, 2013. p. 529-533.

[IV] SOURY, M., GOSSART, C., ADDA-DECKER, M., et al. *A tool to elicit and collect multicultural and multimodal laughter*. In : INTERSPEECH. 2013. p. 773-774.

[V] DEVILLERS, L., SOURY, M. *A social interaction system for studying humor with the Robot NAO*. In : Proceedings of the 15th ACM on International conference on multimodal interaction. ACM, 2013. p. 313-314.

[VI] SOURY, M. et DEVILLERS, L. *Smile and Laughter in Human-Machine Interaction: a study of engagement*. In: Proceedings of the Ninth International Conference on Language Resources and Evaluation, LREC-2014

Communications sans Actes

[VII] SOURY, M. et DEVILLERS, L. *Nao makes me laugh: the impact of humor in human-robot interactions*, Workshop "Toward Social Humanoid Robots: What makes interactions human-like?", IROS 2013

Présentations sans comité de relecture

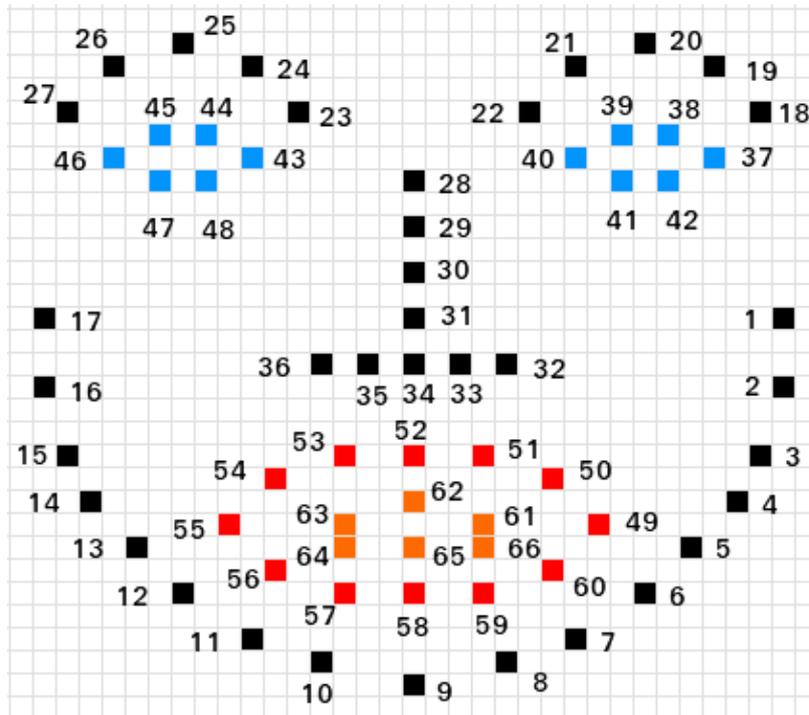
[VIII] *Détection du stress dans la voix, tâche de prise de parole en public dans le cadre du projet COMPARSE*, groupe de travail « Affects, Compagnons Artificiels et Interactions » (ACAI) du GDR 13, 06 juin 2013, UPMC Paris

[IX] *Détection multimodale du Stress pour la conception de logiciels de remédiation et d'entraînement*, Séminaire du groupe « Traitement du Langage Parlé » (TLP), 11 février 2014, LIMSI Orsay

Annexes

Calculs des distances pour l'extraction d'indices faciaux.....	201
Calcul de la Quantité de Mouvement.....	202
Questionnaire de retour d'expérience E-Thérapie.....	203
Questionnaire de retour d'expérience Compare.....	205
Questionnaire de retour d'expérience Interspeech.....	206
Questionnaire de Personnalité OCEAN (Expériences E-Thérapies, Compare et Interspeech).....	207
Liebowitz Social Anxiety Scale - LSAS (Expérience E-Thérapies)	209
Sense of Humor Scale - SHS (Expérience "Nao fait de l'Humour")	210
Echelle d'affirmation de soit de Rathus (Expérience E-Thérapies - corpus ETCC HH)	211
Hamilton Depression Rating Scale - HAM-D (Expérience E-Thérapies - corpus ETCC HH).....	213
Questionnaire de tendance au Rougissement - BQR (Expérience E-Thérapies - corpus ETCC HH).....	216
Script des juges (Expérience Compare)	218

Calculs des distances pour l'extraction d'indices faciaux



Arche des sourcils

$$\frac{\frac{18y + 22y}{2} - 20y + \frac{23y + 27y}{2} - 25y}{2}$$

Hauteur

Écart entre les sourcils

$$\frac{(22x - 23x) + (21x - 24x)}{2}$$

Largeur

Écart entre les sourcils et les yeux

$$\frac{(38y - 19y) + (39y - 21y) + (44y - 24y) + (45y - 26y)}{4}$$

Hauteur

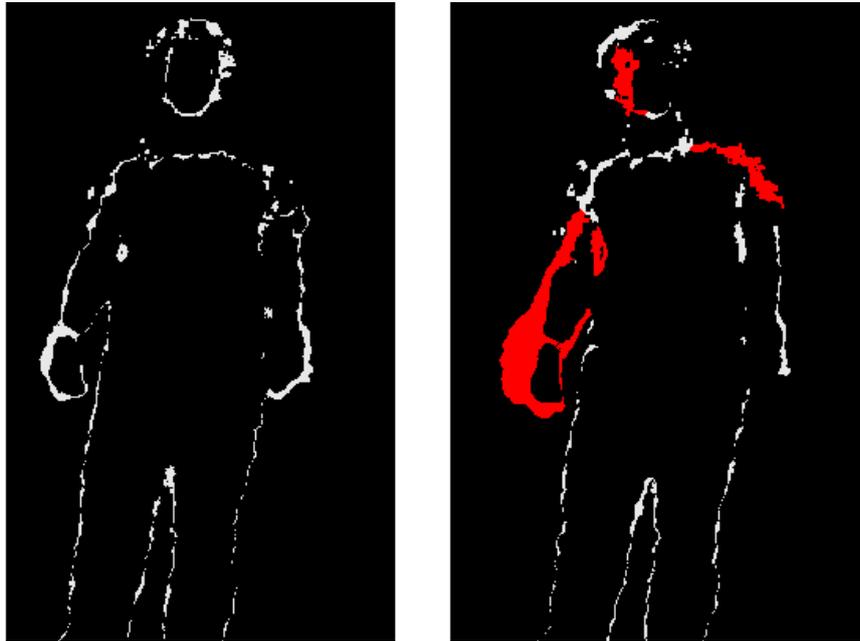
Ouverture de la bouche

$$\frac{(65y - 62y) + (58y - 52y) + (59y - 51y) + (57y - 53y)}{4}$$

Hauteur

Rapport entre l'épaisseur de la lèvre supérieure et l'épaisseur de la lèvre inférieure

$$\frac{\left(\frac{49y + 55y}{2}\right) - 52y}{\frac{58y - (49y + 55y)}{2}}$$

Calcul de la Quantité de Mouvement

```

Pour chaque fenêtre f de n images
  Matrice de mouvement  $m_f$  = image 1 de f
  Aire mouvement  $Am_f$  = 0
  Aire silhouette  $As_f$  = 0
  Pour les images i 2 à n-1 de f
    Pour chaque pixel  $p_{x,y}$  de la matrice i
      Si  $p_{x,y}(i) \neq$  arrière-plan
         $m_{x,y}$  = humain
  Pour l'image n de f
    Pour chaque pixel  $p_{x,y}$  de la matrice n
      Si  $p_{x,y}(n) \neq$  arrière-plan
         $m_{x,y}$  = arrière-plan
         $As_f = As_f + 1$ 
  Pour chaque pixel  $p_{x,y}$  de la matrice  $m_f$ 
    Si  $p_{x,y}m ==$  humain
       $Am_f = Am_f + 1$ 
   $QoM_f = Am_f/As_f$ 

```

Questionnaire de retour d'expérience E-Thérapie

Scénario 1 «Convaincre l'auditoire d'adhérer à la cause écologique» :

Crédibilité du scénario:

1. Est-ce que la situation vous a semblé réaliste dans l'animation?
2. Est-ce que les modifications de couleur symbolisaient bien votre état?
3. Est-ce que l'utilisation de formes géométriques vous a permis de vous projeter dans la situation?
4. Est-ce que vous préféreriez l'utilisation de personnages réalistes?
5. Est-ce que vous avez l'impression que les interlocuteurs vous écoutaient?
6. Est-ce que vous étiez stressé pendant le scénario 1? (évaluez votre stress de 1-pas du tout à 10-terriblement)
7. Si vous avez repensé à une situation que vous avez déjà vécu pour jouer ce scénario, veuillez la décrire (question ouverte)

Evaluation de l'application:

7. Est-ce que le délai pour atteindre le but du scénario était trop long?
8. Est-ce que vous avez eu du mal à atteindre le but du scénario?
9. Est-ce que votre émotion a bien été reconnue?
- 10a. Est-ce que vous trouvez cette animation utile comme entraînement à la prise de parole en public?
- 10b. Si non, quel exercice vous semblerait efficace? (question ouverte)

Scénario 2 «Inviter à dîner quelqu'un qui vous plaît beaucoup pour la première fois» :

Crédibilité du scénario:

1. Est-ce que la situation vous a semblé réaliste dans l'animation?
2. Est-ce que les modifications de couleur symbolisaient bien votre état?
3. Est-ce que l'utilisation de formes géométriques vous a permis de vous projeter dans la situation?
4. Est-ce que vous préféreriez l'utilisation de personnages réalistes?
5. Est-ce que vous avez l'impression que les interlocuteurs vous écoutaient?
6. Est-ce que vous étiez stressé pendant le scénario 2?
7. Si vous avez repensé à une situation que vous avez déjà vécu pour jouer ce scénario, veuillez la décrire (question ouverte)

Evaluation de l'application:

7. Est-ce que le délai pour atteindre le but du scénario était trop long?
8. Est-ce que vous avez eu du mal à atteindre le but du scénario?
9. Est-ce que votre émotion a bien été reconnue?
- 10a. Est-ce que vous trouvez cette animation utile comme entraînement à la gestion des émotions?
- 10b. Si non, quel exercice vous semblerait efficace? (question ouverte)

Scénario 3 «Refuser les sollicitations d'un vendeur insistant» :

Crédibilité du scénario:

1. Est-ce que la situation vous a semblé réaliste dans l'animation?
2. Est-ce que les modifications de couleur symbolisaient bien votre état?
3. Est-ce que l'utilisation de formes géométriques vous a permis de vous projeter dans la situation?
4. Est-ce que vous préféreriez l'utilisation de personnages réalistes?
5. Est-ce que vous avez l'impression que les interlocuteurs vous écoutaient?
6. Est-ce que vous étiez stressé pendant le scénario 3?
7. Si vous avez repensé à une situation que vous avez déjà vécu pour jouer ce scénario, veuillez la décrire (question ouverte)

Evaluation de l'application:

7. Est-ce que le délai pour atteindre le but du scénario était trop long?
8. Est-ce que vous avez eu du mal à atteindre le but du scénario?
9. Est-ce que votre émotion a bien été reconnue?
- 10a. Est-ce que vous trouvez cette animation utile comme entraînement à la gestion des émotions?
- 10b. Si non, quel exercice vous semblerait efficace? (question ouverte)

Application en général :

1. Quel est votre niveau de stress (e) pendant l'animation ?
2. Quel est votre niveau de stress (e) après l'animation ?
3. Vous sentez-vous stressé(e) après cette expérience? Oui./ Non.

Commentaires :

Pensez-vous à d'autres situations anxieuses dans lesquelles un entraînement serait intéressant ?

Seriez-vous d'accord pour refaire partie d'une autre expérience du même type ?

Questionnaire de retour d'expérience Comparsa

1 - pas du tout d'accord – 5 - tout à fait d'accord.

Évaluez votre niveau de stress

Globalement, vous vous sentiez stressé(e) :

Si vous avez répondu « D'accord » ou « Tout à fait d'accord » à la question précédente :

Les conditions d'expérimentation vous stressaient :

L'attitude, la présence des juges vous stressait :

Vous avez su gérer votre stress :

Évaluez l'attitude des juges à votre égard

Juge féminin

4) L'attitude de ce juge à votre égard était agréable :

5) L'attitude de ce juge à votre égard vous a stimulé à mieux faire :

6) De façon générale, l'attitude de ce juge correspond à ce que vous attendez d'un juge :

7) Le juge était crédible dans son rôle :

Juge masculin

8) L'attitude de ce juge à votre égard était agréable :

9) L'attitude de ce juge à votre égard vous a stimulé à mieux faire :

10) De façon générale, l'attitude de ce juge correspond à ce que vous attendez d'un juge :

11) Le juge était crédible dans son rôle :

Questionnaire de retour d'expérience Interspeech

A propos du Robot..

Avez-vous facilement compris le robot?

Oui Non

Comment évalueriez-vous l'attitude du robot durant cette expérience?

Veillez l'évaluer selon les cinq traits suivants sur une échelle de 1 à 5 (1 correspondant surtout à l'adjectif de gauche - 5 correspondant surtout à l'adjectif de droite)

je ne l'ai pas aimée	1	2	3	4	5	je l'ai aimée
non-amicale	1	2	3	4	5	amicale
désobligeante		1	2	3	4	5 attentionnée
déplaisante	1	2	3	4	5	plaisante
horrible	1	2	3	4	5	agréable

A propos de l'expérience..

Est-ce que cette expérience vous a mis mal à l'aise? *(1 pas du tout - 5 tout à fait)*

1 2 3 4 5

Quelle partie avez-vous préféré ?

les vidéos amusantes les fourches-langue les blagues du robot

Avez-vous apprécié cette expérience?

Oui Non

Questionnaire de Personnalité OCEAN (Expériences E-Thérapies, Comparsé et Interspeech)

Je me vois comme une personne... 1 : *pas du tout d'accord* ; 5 : *tout à fait d'accord*

- 1) loquace, volubile, qui aime parler
- 2) qui tend à remarquer les défauts des autres
- 3) qui fait un travail consciencieux
- 4) déprimée, mélancolique
- 5) originale, qui a des idées innovantes
- 6) réservée
- 7) serviable et généreuse avec les autres
- 8) parfois négligente, insouciante
- 9) détendue, qui sait gérer son stress
- 10) curieuse à l'égard de plein de choses
- 11) pleine d'énergie
- 12) qui entre en conflit avec les autres
- 13) sur qui son employeur peut compter
- 14) parfois tendue
- 15) ingénieuse, avec une réflexion élaborée
- 16) qui génère beaucoup d'enthousiasme autour d'elle
- 17) indulgente, qui pardonne facilement
- 18) assez désorganisée
- 19) qui s'inquiète beaucoup
- 20) à l'imagination débordante
- 21) tranquille, calme
- 22) qui fait confiance facilement
- 23) plutôt paresseuse
- 24) émotionnellement stable, qui ne se laisse pas contrarier
- 25) inventive, créative
- 26) pleine d'assurance, capable de s'affirmer
- 27) parfois froide et distante
- 28) qui persévère jusqu'à l'accomplissement de sa tâche
- 29) d'humeur changeante
- 30) qui valorise les expériences artistiques et esthétiques
- 31) parfois timide, inhibée
- 32) attentionnée, aimable avec les autres
- 33) qui fait les choses de façon efficaces

Annexes

- 34) qui reste calme dans les situations tendues
- 35) qui préfère le travail routinier
- 36) ouverte et sociable
- 37) parfois grossière envers les autres
- 38) qui planifie ses actions et s'y tient
- 39) facilement nerveuse
- 40) qui aime réfléchir et jouer avec les idées
- 41) qui a peu de centres d'intérêts artistiques
- 42) qui aime collaborer avec les autres
- 43) facilement distraite
- 44) qui a des goûts pointus en musique, en art ou en littérature

Liebowitz Social Anxiety Scale - LSAS (Expérience E-Thérapies)

Echelles

Peur/Anxiété :	Évitement:
0 pas du tout	0 jamais
1 un peu	1 quelques fois
2 assez	2 souvent
3 beaucoup	3 presque toujours

Comment vous évaluez vous dans les situations suivantes ? :

	Ça m'angoisse	Je l'évite
1) Téléphoner en public		
2) Participer au sein d'un petit groupe		
3) Manger dans un lieu public		
4) Boire avec d'autres personnes dans un lieu public		
5) Parler à une figure d'autorité		
6) Jouer quelque chose ou parler devant une audience		
7) Aller à une fête		
8) Travailler pendant qu'on vous observe		
9) Écrire pendant qu'on vous observe		
10) Téléphoner à quelqu'un que vous ne connaissez pas bien		
11) Parler avec des gens que vous ne connaissez pas bien		
12) Rencontrer des inconnus		
13) Uriner dans des WC publics		
14) Entrer dans une salle où d'autres personnes sont déjà assises		
15) Être le centre de l'attention		
16) Prendre la parole en réunion		
17) Passer un examen		
18) Exprimer votre désaccord à des gens que vous ne connaissez pas bien		
19) Regarder dans les yeux des personnes que vous ne connaissez pas bien		
20) Faire un rapport à un groupe		
21) Draguer quelqu'un		
22) Échanger ou vous faire rembourser un article dans une boutique		
23) Organiser une fête		
24) Refuser l'offre d'un vendeur insistant		

Sense of Humor Scale - SHS (Expérience "Nao fait de l'Humour")

Évaluez les affirmations suivantes sur une échelle de 1 à 7 (1 vous n'êtes pas du tout d'accord - 7 vous êtes tout à fait d'accord)

- 1) Il m'est facile de rire lorsque je suis le dindon de la farce.
- 2) Je trouve souvent des choses drôles dans le quotidien à la maison.
- 3) Je cherche généralement des sitcoms ou des choses amusantes lorsque je regarde la TV.
- 4) J'utilise souvent mon sens de l'humour pour gérer mon stress.
- 5) Quand je vais au cinéma, je préfère généralement une bonne comédie.
- 6) J'utilise souvent l'humour au travail pour réduire le stress et rester efficace.
- 7) Mon sens de l'humour m'abandonne rarement lorsque je suis stressé(e).
- 8) Je raconte souvent des blagues.
- 9) Je partage souvent avec les autres les incidents amusants que j'observe, ou qui m'arrivent.
- 10) Un de mes parents (ou les deux) riait(en)t souvent lorsque j'étais enfant.
- 11) Je partage souvent avec les autres le côté amusant de mes impairs/incidents embarrassants.
- 12) Je trouve souvent des choses drôles dans le quotidien au travail.
- 13) Je crée souvent mes propres jeux de mots.
- 14) Je fais souvent des remarques pleines d'esprit.
- 15) Me moquer de mes imperfections physiques ne me pose pas de problème.
- 16) Je trouve souvent des choses drôles dans le quotidien en dehors du travail ou de la maison.
- 17) Quand je prends un magazine, en général je regarde d'abord les bandes dessinées dedans.
- 18) Je suis à l'aise pour rire, même quand les autres ne rient pas.
- 19) J'ai un rire plus profond que la plupart des gens.
- 20) C'est important pour moi qu'il y ait beaucoup d'humour dans ma vie.
- 21) Je raconte souvent des histoires amusantes.
- 22) Je rie franchement plusieurs fois par jour.
- 23) Mon sens de l'humour est mon outil le plus efficace pour lutter contre le stress.
- 24) Je me moque souvent de mes propres impairs/incidents embarrassants.

Echelle d'affirmation de soit de Rathus (Expérience E-Thérapies - corpus ETCC HH)

Evaluez vous sur les affirmations suivantes, de - 3 (vraiment non caractéristique) à + 3 (très caractéristique)

1. La plupart des gens me semblent être plus agressifs et défendre mieux leurs droits que moi.
2. Il m'est arrivé d'hésiter par timidité au moment de donner ou d'accepter un rendez-vous.
3. Quand la nourriture dans un restaurant ne me satisfait pas, je m'en plains au serveur ou à la serveuse.
4. Je fais attention à ne pas heurter le sentiment des autres, même lorsque l'on m'a blessé(e).
5. Si un vendeur s'est donné beaucoup de mal à me montrer une marchandise qui ne me convient pas parfaitement, j'ai un moment difficile lorsque je dois dire "non".
6. Lorsqu'on me demande de faire quelque chose, j'insiste pour en savoir la raison.
7. Il y a des moments où je cherche une bonne et vigoureuse discussion.
8. Je me bats pour arriver aussi bien que les autres dans ma profession.
9. A vrai dire, les gens tirent souvent profit de moi.
10. J'ai du plaisir à entreprendre des conversations avec de nouvelles connaissances ou des étrangers.
11. Souvent je ne sais rien dire à des personnes séduisantes du sexe opposé.
12. J'hésiterais à téléphoner à un grand établissement de commerce ou à une administration.
13. Je préférerais poser ma candidature pour un autre travail par lettre plutôt que par un entretien personnel.
14. Je trouve embarrassant de renvoyer une marchandise.
15. Si un parent proche et respecté est en train de m'ennuyer, j'étoufferai mes sentiments plutôt que d'exprimer cette gêne.
16. Il m'est arrivé de ne pas poser des questions par peur de paraître stupide.
17. Pendant une discussion serrée, j'ai souvent peur d'être bouleversé(e) au point de tremble de tout mon corps.
18. Si un conférencier réputé et respecté dit quelque chose que je pense inexact, j'aimerais que le public entende aussi mon point de vue.
19. J'évite de discuter les prix avec les représentants et les vendeurs.
20. Lorsque j'ai fait quelque chose d'important et de valable, je m'arrange pour le faire savoir des autres.
21. Je suis ouvert et franc au sujet de mes sentiments.
22. Si quelqu'un a répandu des histoires fausses et de mauvais gout à mon sujet, je le vois aussi vite que possible pour une explication.
23. J'ai souvent un moment difficile lorsque je dois dire "non".
24. J'ai tendance à contenir des émotions plutôt que de faire une scène.

Annexes

25. Je me plains lorsque le service est mal fait dans un restaurant ou ailleurs.
26. Souvent je ne sais pas quoi dire lorsque l'on me fait un compliment.
27. Au cinéma, au théâtre, ou lors d'une conférence, si un couple près de moi est en train de parler à voix haute, je lui demande se taire ou d'aller autre part.
28. Quiconque cherche à passer devant moi dans une queue risque une bonne explication.
29. Je suis rapide dans l'expression de mes opinions.
30. Il y a des moments où je ne sais pas quoi dire.

Hamilton Depression Rating Scale – HAM-D (Expérience E-Thérapies - corpus ETCC HH)

Cette échelle est conçue pour être utilisée par un professionnel formé lors d'une consultation à partir de ses observations et des réponses de la personne qui consulte.

1) Humeur dépressive (tristesse, sentiment d'être sans espoir, impuissant, auto-dépréciation)

0 Absent

1 Ces états affectifs ne sont signalés que si l'on interroge le sujet

2 Ces états affectifs sont signalés verbalement spontanément

3 Le sujet communique ces états affectifs non verbalement (expression facile, attitude, voix, pleurs)

4 Le sujet ne communique pratiquement que ses états affectifs dans ses communications spontanées verbales et non verbales

2) Sentiments de culpabilité

0 Absent

1 S'adresse des reproches à lui-même, a l'impression qu'il a causé un préjudice à des gens

2 Idées de culpabilité ou ruminations sur des erreurs passées ou des actions condamnables

3 La maladie actuelle est une punition. Idées délirantes de culpabilité

4 Entend des voix qui l'accusent ou le dénoncent et/ou a des hallucinations visuelles menaçantes

3) Suicide

0 Absent

1 A l'impression que la vie ne vaut pas la peine d'être vécue

2 Souhaite être mort ou équivalent : toute pensée de mort possible dirigée contre lui-même

3 Idées ou gestes de suicide

4 Tentatives de suicide

4) Insomnie du début de nuit

0 Absent

1 Se plaint de difficultés éventuelles à s'endormir

2 Se plaint d'avoir chaque soir des difficultés à s'endormir

5) Insomnie du milieu de nuit

0 Pas de difficulté

1 Le malade se plaint d'être agité ou troublé pendant la nuit

2 Il se réveille pendant la nuit

6) Insomnie du matin

0 Pas de difficulté

1 Se réveille de très bonne heure le matin mais se rendort

2 Incapable de se rendormir s'il se lève

7) Travail et activités

0 Pas de difficulté

1 Pensées et sentiments d'incapacité, fatigue ou faiblesse se rapportant à des activités professionnelles ou de détente

2 Perte d'intérêt pour les activités professionnelles ou de détente, ou décrite directement par le malade ou indirectement par son apathie, son indécision et ses hésitations

3 Diminution du temps d'activité ou diminution de la productivité

4 A arrêté son travail en raison de sa maladie actuelle

8) Ralentissement (lenteur de la pensée et du langage, baisse de la faculté de concentration, baisse de l'activité motrice)

0 Langage et pensées normaux

1 Léger ralentissement à l'entretien

2 Ralentissement manifeste à l'entretien

3 Entretien difficile

4 Stupeur

9) Agitation

0 Aucune

1 Crispations, secousses musculaires

2 Joue avec ses mains, ses cheveux, etc

3 Bouge, ne peut rester assis tranquille

4 Se tord les mains, ronge ses ongles, arrache ses cheveux, se mord les lèvres

10) Anxiété psychique

0 Aucun trouble

1 Tension subjective et irritabilité

2 Se fait du souci à propos de problèmes mineurs

3 Attitude inquiète, apparente dans l'expression faciale et le langage

4 Peurs exprimées sans que l'on pose de questions

11) Anxiété somatique (bouche sèche, troubles digestifs, palpitations, céphalées, pollakiurie, hyperventilation, transpiration, soupirs)

0 Absente

1 Discrète

2 Moyenne

3 Grave

4 Frappant le sujet d'incapacité fonctionnelle

12) Symptômes somatiques gastro-intestinaux

0 Aucun

- 1 Perte d'appétit mais mange sans y être poussé. Sentiment de lourdeur abdominale
- 2 A des difficultés à manger en l'absence d'incitations. Demande ou besoins de laxatifs, de médicaments intestinaux

13) Symptômes somatiques généraux

0 Aucun

1 Lourdeur dans les membres, dans le dos ou la tête. Douleurs dans le dos, céphalées, douleurs musculaires, perte d'énergie et fatigabilité

2 Si n'importe quel symptôme est net

14) Symptômes génitaux (perte de libido, troubles menstruels)

0 Absents

1 Légers

2 Graves

15) Hypochondrie

0 Absente

1 Attention concentrée sur son propre corps

2 Préoccupations sur sa santé

3 Plaintes fréquentes, demandes d'aide

4 Idées délirantes hypochondriaques

16) Perte de poids

A : selon les dires du malade

0 Pas de perte de poids

1 Perte de poids probable liée à la maladie actuelle

2 Perte de poids certaine

B : appréciée par pesées

0 Moins de 500 g de perte de poids par semaine

1 Plus de 500 g de perte de poids par semaine

2 Plus de 1 kg de perte de poids par semaine

17) Prise de conscience

0 Reconnaît qu'il est déprimé et malade

1 Reconnaît qu'il est malade mais l'attribue à la nourriture, au climat, au surmenage, à un virus, à un besoin de repos, etc

2 Nie qu'il est malade

Questionnaire de tendance au Rougissement - BQR (Expérience E-Thérapies - corpus ETCC HH)

Cochez la ligne qui correspond à la réponse qui convient le mieux pour décrire ce que vous avez ressenti au cours de la dernière semaine.

1. A quelle fréquence avez-vous ressenti des débuts de rougissement ?

- 0 Jamais au cours de la dernière semaine
- 1 Rarement (moins d'une fois par jour)
- 2 Environ une fois par jour
- 3 Plusieurs fois par jour
- 4 Plus de 10 fois par jour

2. A quelle fréquence avez-vous ressenti des crises de rougissements longues et pénibles ?

- 0 Jamais au cours de la dernière semaine
- 1 Rarement (moins d'une fois par jour)
- 2 Environ une fois par jour
- 3 Plusieurs fois par jour
- 4 Plus de 10 fois par jour

3. A quelle fréquence avez-vous été perturbé ou anxieux par l'idée de rougir ?

- 0 Jamais au cours de la dernière semaine
- 1 Rarement (pas tous les jours)
- 2 Tous les jours, mais moins de la moitié de la journée
- 3 Tous les jours, plus de la moitié de la journée
- 4 Tous les jours, toute la journée ou presque en continu

4. Quelle est, globalement, l'intensité des crises de rougissement que vous avez présentées ?

- 0 Très légère (ou aucune crise)
- 1 Légère
- 2 Moyennement intense
- 3 Intense
- 4 Extrêmement intense

5. A quel point vous sentez-vous actuellement impuissant(e) ou dépassé(e) par votre problème de rougissement ?

- 0 Pas dépassé(e) ni impuissant(e), ou pas de problème du tout
- 1 Légèrement dépassé(e) ou impuissant(e)
- 2 Moyennement dépassé(e) ou impuissant(e)
- 3 Grandement dépassé(e) ou impuissant(e)
- 4 Extrêmement ou complètement dépassé(e) ou impuissant(e)

6. Quel est, en ce moment, le niveau du retentissement global de votre problème de rougissement dans votre vie quotidienne (travail ou possibilité de travailler, vie sociale, vie familiale, etc.) ?

0 Aucun retentissement ou pas de problème du tout

1 Retentissement léger

2 Retentissement moyen

3 Retentissement important

4 Retentissement extrême

*Script des juges (Expérience Comparse)**Attitudes Sociales Négatives → Déstabiliser*

Attitude sociale du juge	Tag associés	Remarques
Mise en valeur de soi-même exagérée	Narcissique, dominant, arrogant	Exagérer sa propre valeur
Exemple d'énonciations	Phrases de relance	
<p>« Regardez-moi, ce n'est pas en faisant ce genre de présentation que j'ai atteint ce niveau. »</p> <p>« Vous me faites perdre mon temps avec ce genre de présentation. »</p> <p>« Écoutez, je suis une personne à responsabilités, et j'ai un emploi du temps chargé. Tâchons d'en finir vite. »</p>	<p>« Bien que ça ne m'intéresse pas spécialement, est-ce que vous pouvez détailler un peu plus [tel ou tel point abordé par le participant] ? »</p> <p>« Moi, je serais [point faible évoqué par le participant], je n'aurais jamais atteint mon niveau professionnel. Vous comptez y remédier ? »</p>	

Attitude sociale du juge	Tag associés	Remarques
Critique	Méprisant, belliqueux, condescendant	Mettre en doute les capacités du locuteur
Exemple d'énonciations	Phrases de relance	
<p>« Là, je suis absolument sûr que vous ne répondez pas aux critères attendus. »</p> <p>« Quel charabia... Votre présentation n'a aucun intérêt. »</p> <p>« Comment pensez-vous que je peux vous juger avec une présentation aussi médiocre. »</p>	<p>« Vous comprenez vraiment pourquoi vous êtes ici ? »</p> <p>« Est-ce que vous avez quelque chose de plus intéressant à mettre en avant ? »</p> <p>« Reparlons de vos points faibles, puisque c'est ce qui prédomine. Vous pensez que vous allez résoudre ce problème de [point faible abordé par le locuteur], ou bien vous allez laisser ça en l'état ? »</p>	

Attitude sociale du juge	Tag associés	Remarques
Soupçonneux	Sceptique, suspicieux	Mettre en doute l'honnêteté du locuteur
Exemple d'énonciations		Phrases de relance
<p>« Je ne vous trouve pas franc. »</p> <p>« Non mais vous mentez, c'est flagrant. »</p> <p>« J'é mets un gros doute quant à vos compétences. »</p>		<p>« Est-ce que vous pourriez me dire, maintenant, ce que sont <i>véritablement</i> vos points faibles ? »</p> <p>« Est-ce que vous êtes <i>vraiment</i> [point fort abordé par le participant] ?! »</p>

Attitude sociale du juge	Tag associés	Remarques
Effacé	Embarrassé, doute de soi	Ne répond pas aux qualités de décision attendues d'un juge
Exemple d'énonciations		Phrases de relance
<p>« Je n'en sais rien, je n'ai pas d'avis... »</p> <p>« Je ne suis pas qualifié pour traiter ce genre de cas moi... »</p> <p>« Je ne sais pas quoi penser de votre présentation. Il faudrait demander à quelqu'un qui s'y connaît. »</p>		<p>« Euh, qu'est-ce qu'on doit faire maintenant ? On vous pose des questions, c'est ça ? »</p> <p>« Je ne sais pas, on peut reparler de vos points forts non ? C'était quoi déjà ? »</p>

Attitudes Sociales Positives → Stimuler

Attitude sociale du juge	Tag associés	Remarques
Meneur	Conseiller, assuré, intrusif	Le juge insiste pour prodiguer de bons conseils
Exemple d'énonciations		Phrases de relance
<p>« C'était pas trop mal, mais vous devriez insister un peu plus sur vos points forts. »</p> <p>« Je serais vous, j'insisterais plus sur les points faibles. Là c'était un peu bâclé. »</p>		<p>« On va reprendre jusqu'à ce que vous le fassiez correctement. Reparlez-moi de vos points faibles. »</p> <p>« Il faut que vous y arriviez correctement. Reparlez-nous de vos points forts. »</p>

Attitude sociale du juge	Tag associés	Remarques
Protecteur	Encourageant, surprotecteur, exprime de la pitié, paternaliste	Animé de bonnes intentions, le juge aide le participant
Exemple d'énonciations		Phrases de relance
<p>« Mais si voyons rassurez-vous ! C'est déjà un très bon début ! »</p> <p>« Vous pouvez faire encore mieux, la prochaine sera la bonne, allez ! »</p> <p>« Je pense que vous pouvez progresser, je ne peux que vous encourager. »</p> <p>« Il faut vous entraîner, vous pouvez faire mieux. »</p>		<p>« Parlons à nouveau de vos points forts, vous allez le faire mieux cette fois-ci, vous verrez. »</p> <p>« Je sais que c'est pas facile, mais reparlez un peu de [point faible évoqué par le participant]. Comment comptez-vous faire pour vous améliorer ?»</p>

Attitude sociale du juge	Tag associés	Remarques
Coopératif	Agréable, confiant, satisfait	Le juge n'en demande pas plus que ce le participant a à donner
Exemple d'énonciations		Phrases de relance
<p>« C'est très bien comme ça. »</p> <p>« Votre présentation était bien. »</p> <p>« C'était très agréable de vous écouter. »</p>		<p>« Est-ce que vous avez quelque chose à ajouter ? »</p> <p>« C'est bien d'avoir cité [point fort évoqué par le participant]. Voulez-vous m'en dire plus ? »</p> <p>« Vos points faibles étaient bien présentés. Pourriez-vous détailler un peu [point faible évoqué par le participant] ? »</p>

Attitude sociale du juge	Tag associés	Remarques
Docile	Obéissant, d'accord	Le juge est influencé par le locuteur, il dit « oui » à tout, il peut même exprimer un changement d'avis
Exemple d'énonciations		Phrases de relance
<p>« Ah oui, vous avez absolument raison en fait. »</p> <p>« Je pense tout à fait comme vous, oui oui. »</p> <p>« Votre présentation était vraiment parfaite. »</p>		<p>« Vous vous dites [point fort évoqué par le participant], vous avez tout à fait raison. Est-ce que vous seriez d'accord pour nous en dire plus ? »</p> <p>« Je suis vraiment d'accord avec vous, [point faible évoqué par le participant] doit être amélioré. Comment comptez-vous y prendre ? »</p>