

HEC MONTRÉAL

**Création d'un index d'expérience émotionnelle pour l'industrie de
l'assurance**

par

Melissa Turp-Yonezawa

**Sciences de la gestion
(Spécialisation Intelligence d'affaires)**

*Mémoire présenté en vue de l'obtention
du grade de maîtrise ès sciences en gestion
(M. Sc.)*

Août 2021

© Melissa Turp-Yonezawa, 2021

CERTIFICAT D'APPROBATION ÉTHIQUE

La présente atteste que le projet de recherche décrit ci-dessous a fait l'objet d'une évaluation en matière d'éthique de la recherche avec des êtres humains et qu'il satisfait aux exigences de notre politique en cette matière.

Projet # : 2018-2904

Titre du projet de recherche : Recherche et développement collaborative en UX

Chercheur principal :

Marc Fredette,
Professeur titulaire, Département de sciences de la décision, HEC Montréal

Cochercheurs :

Pierre-Majorique Léger; François Courtemanche; Shang-Lin Chen

Date d'approbation du projet : 11 janvier 2018

Date d'entrée en vigueur du certificat : 11 janvier 2018

Date d'échéance du certificat : 01 janvier 2019

Maurice Lemelin
Président du CER de HEC Montréal

ATTESTATION D'APPROBATION ÉTHIQUE COMPLÉTÉE

La présente atteste que le projet de recherche décrit ci-dessous a fait l'objet des approbations en matière d'éthique de la recherche avec des êtres humains nécessaires selon les exigences de HEC Montréal.

La période de validité du certificat d'approbation éthique émis pour ce projet est maintenant terminée. Si vous devez reprendre contact avec les participants ou reprendre une collecte de données pour ce projet, la certification éthique doit être réactivée préalablement. Vous devez alors prendre contact avec le secrétariat du CER de HEC Montréal.

Nom de l'étudiant : Melissa Turp-Yonezawa

Titre du projet supervisé/mémoire/thèse :
Création d'un index d'expérience émotionnelle pour l'industrie de l'assurance

Titre du projet sur le certificat :
Recherche et développement collaborative en UX

Projet # : 2018-2904

Chercheur principal / directeur de recherche :
Marc Fredette

Cochercheurs :
Pierre-Majorique Léger; François Courtemanche; Shang-Lin Chen; Élise Boissonneault; Florian Coustures;
Melissa Turp-Yonezawa; Amay Cheam; Pierre-Jude Souverain; Omar Chikhar

Date d'approbation initiale du projet : 11 janvier 2018

Date de fermeture de l'approbation éthique pour l'étudiant(e) : 12 août 2021

Maurice Lemelin
Président
CER de HEC Montréal

Résumé

L'index d'expérience émotionnelle permet de convertir les mesures de la valence, qui est le type d'émotion vécue, et l'activation, qui est l'intensité de l'émotion vécue, en une seule métrique. Cet indice place une émotion mesurée sur une échelle de 0 à 100, où 100 est une émotion positive pure. Cet outil a suscité le désir d'explorer la mesure de l'émotion dans des domaines particuliers. Un domaine d'intérêt est celui de l'assurance. Plusieurs pistes d'exploration ont été considérées dans le but de nuancer l'indice pour cette industrie. La piste principalement étudiée est la possibilité d'utiliser cet outil sans une phase de calibration, qui était une partie intégrante de la version originale de l'indice. Les résultats obtenus démontrent l'importance de la phase de calibration dans le fonctionnement de l'indice et qu'une version de l'indice sans cette étape ne pourra pas se fier sur les principes démontrés par la première version.

En analysant les données utilisées pour développer la version originale de l'indice, dans le but de mieux comprendre le rôle de la phase de calibration pour l'indice, certaines questions ont émergées en lien avec la qualité des données. Ce mémoire présente les résultats de ces analyses, qui illustrent des différences entre les méthodes physiologiques (FaceReader, activité électrodermale, EDA phasique et variabilité de la fréquence cardiaque) pour mesurer l'émotion, utilisées par l'indice, et les méthodes autodéclarées, qui sont utilisées par la banque d'images IAPS, une partie intégrante de la phase de calibration de l'indice.

Mots clés : Valence, Activation, Activité Électrodermale, EDA Phasique, Variabilité de la fréquence cardiaque, SAM Scale, IAPS, Identification des émotions

Abstract

The emotional experience index converts measurement of valence, the type of experienced emotion, and arousal, the intensity of the experienced emotion, into a single metric. This index grades a measured emotion on a scale of 0 to 100, where 100 is a purely positive emotion. The existence of this tool generated interest in exploring its utility within multiple industries, particularly from the field of insurance. Different avenues were considered with the goal of better contextualizing the index for this field. The main path explored is the possibility of using this tool without a calibration phase, an integral part of the original version of the index. The results obtained indicate the importance of this calibration phase to the proper functioning of the index and that a version of the index without this phase cannot be developed by taking the principles of the original version of the index for granted.

When analyzing the data used to develop the original version of the index, to better understand the role of the calibration phase for the index, several questions arose regarding the quality of the data obtained. This thesis presents the results of these analyses that illustrate differences between the physiological methods (FaceReader, Electrodermal activity, Phasic EDA, and Heart rate variability) for measuring emotion, that are used by the index, and self-reported methods that are used by the IAPS image database, a core component of the calibration phase of the index

Keywords: Valence, Arousal, Electrodermal Activity, Phasic EDA, Heart Rate Variability, SAM Scale, IAPS, Identification of emotions

Table des matières

Résumé.....	iii
Abstract	v
Table des matières.....	vii
Liste des tableaux.....	xi
Liste des figures	xiii
Liste des abréviations.....	xvii
Avant-propos.....	xix
Remerciements.....	xxi
Chapitre 1 Introduction	1
Chapitre 2 Revue de la littérature.....	5
Résumé	5
2.1 La théorie circumplex de l'émotion	5
2.2 Outils mesurant l'émotion	7
2.2.1 Outils mesurant l'émotion de manière autodéclarée	7
2.2.1.1 Self-Assessment Manikin	7
2.2.1.2 Curseur affectif	8
2.2.2 Outils mesurant l'émotion par des moyens physiologiques	10
2.2.2.1 Analyse faciale.....	11
2.2.2.2 Activité électrodermale.....	13
2.2.2.3 Variabilité du rythme cardiaque	14
2.3 Outils pour générer des émotions spécifiques	15
2.3.1 International Affective Picture System.....	15
2.3.2 Geneva Affective Picture Database	15
2.4 Conclusion.....	17
Chapitre 3 Données initiales et méthodologie	19
Résumé	19
3.1 Description de l'index d'expérience émotionnelle.....	19
3.1.1 Présentation de l'indice.....	19
3.1.2 Calcul de l'indice.....	20

3.1.3 Présentation de l'étude originale de l'indice	23
3.2 Présentation des données initiales, accès aux données et manipulations effectuées	26
3.3 Présentation des analyses effectuées	28
3.3.1 Pistes de recherche originales de l'étude	28
3.3.2 Comparaison entre l'indice original et des méthodes ne se fiant pas à la calibration	29
3.3.3 Analyse des signaux physiologiques et autodéclarés	31
3.3.4 Analyse des variances des émotions générées lors de la phase de calibration	33
3.4 Conclusion.....	35
Chapitre 4 Analyse.....	37
Résumé.....	37
4.1 Comparaison entre l'indice original et des méthodes ne se fiant pas à la calibration	37
4.1.1 Analyse des données de l'indice agrégées par seconde, par stimulus et par participant	38
4.1.2 Analyse des données de l'indice agrégées par stimulus et par participant.....	41
4.2 Analyse des signaux physiologiques et autodéclarés	44
4.2.1 Analyse de la corrélation entre les signaux physiologiques et autodéclarés ...	45
4.2.2 Analyse des statistiques descriptives de la distance entre les signaux du point théorique	49
4.3 Analyse des variances des émotions générées lors de la phase de calibration.....	60
4.3.1 Analyse des variances pour les signaux mesurant la valence.....	61
4.3.1.1 Analyse des variances pour la valence mesurée par l'analyse faciale.....	63
4.3.1.2 Analyse des variances pour la valence mesurées par le curseur affectif	64
4.3.2 Analyse des variances pour les signaux mesurant l'activation.....	65
4.3.2.1 Analyse des variances pour l'activation mesurées par l'EDA phasique et tonique	68
4.3.2.2 Analyse des variances pour l'activation mesurées par l'EDA phasique	69
4.3.2.3 Analyse des variances pour l'activation mesurées par le HRV	70
4.3.2.4 Analyse des variances pour l'activation mesurées par le curseur affectif....	71

4.3.3 Analyse des variances pour les signaux physiologiques de durée suffisante .	72
4.3.3.1 Analyse des variances pour la valence mesurée par l'analyse faciale	75
4.3.3.2 Analyse des variances pour l'activation mesurée par l'EDA phasique et tonique	76
4.3.3.3 Analyse des variances pour l'activation mesurée par l'EDA phasique	77
4.3.3.3 Analyse des variances pour l'activation mesurée par l'HRV	78
4.3.4 Conclusion de l'analyse	79
4.4 Conclusion	83
Chapitre 5 Discussion	85
Résumé	85
5.1 Interprétation des résultats	85
5.1.1 Interprétation de l'analyse de la corrélation entre des méthodes de calcul de l'indice sans la phase de calibration et la méthode originale	85
5.1.2 Interprétation de l'analyse sur la relation entre les signaux physiologiques et autodéclarés pour mesurer la valence et l'activation	87
5.1.3 Interprétation de l'analyse des variances des émotions générées lors de la phase de calibration	88
5.2 Limites	91
5.3 Conclusion	92
Chapitre 6 Conclusion	93
6.1 Piste de recherche future	93
6.2 Contribution industrielle	94
Bibliographie	97
Annexes	i
Annexe 1 : Images IAPS et GAPED utilisées lors de la phase de calibration	i
Annexe 2 : Images de calibration dans leur contexte de présentation lors de l'étude (fil d'actualité Facebook)	xiii
Annexe 3 : Moyennes des valences et activations pour les images de calibration selon la méthode de mesure de l'émotion et distance euclidienne avec la valeur théorique	xvii
Annexe 4 : Résultats des tests deux à deux de l'ANOVA	xx

Liste des tableaux

Tableau 4.1 — Statistiques descriptives des indices avec les données agrégées à la seconde.....	39
Tableau 4.2 — Résultats du test du coefficient de corrélation entre les méthodes candidates pour calculer l'indice et la méthode originale avec les données agrégées à la seconde.....	40
Tableau 4.3 — Statistiques descriptives des indices avec les données agrégées au stimulus	42
Tableau 4.4 — Résultats du test du coefficient de corrélation entre les méthodes candidates pour calculer l'indice et la méthode originale pour les données agrégées au stimulus	43
Tableau 4.5 — Statistiques descriptives des signaux mesurant la valence.....	46
Tableau 4.6 — Statistiques descriptives des signaux mesurant l'activation.....	47
Tableau 4.7 — Résultats du test du coefficient de corrélation entre les signaux mesurant la valence.....	48
Tableau 4.8 — Résultats du test du coefficient de corrélation entre les signaux mesurant l'activation.....	49
Tableau 4.9 — Statistiques descriptives des distances entre les méthodes pour mesurer l'émotion et les valeurs théoriques.....	51
Tableau 4.10 — Moyennes observées pour chaque image de calibration pour les méthodes mesurant la valence	61
Tableau 4.11 — Moyennes observées pour chaque image de calibration pour les méthodes mesurant l'activation	66
Tableau 4.12 — Moyennes observées pour chaque image de calibration pour les méthodes physiologiques avec un regard total d'au moins une seconde	73

Liste des figures

Figure 2.1 — Schéma illustrant la théorie circumplex de l'émotion	6
Figure 2.2 — Illustration de l'échelle SAM pour la valence (échelle du haut) et l'activation (échelle du bas)	8
Figure 2.3 — Curseur affectif mesurant l'activation (échelle du haut) et la valence (échelle du bas)	9
Figure 2.4 — Images des banques IAPS et GAPED générant des émotions selon la théorie circumplex de l'émotion	16
Figure 3.1 — Exemple de la relation entre les composantes du calcul de l'indice.....	21
Figure 3.2 — Exemples d'émotions mesurées dans une étude et leur valeur de l'indice (faible à gauche; forte à droite)	22
Figure 3.3 — Exemples de variations de l'indice en fonction des points p0 et p100 (indice faible à gauche; indice fort à droite)	23
Figure 3.4 — Exemple d'une page de la phase de calibration.....	25
Figure 4.1 — Illustration des points p0 et p100 pour les trois méthodes de calcul de l'indice étudiées	38
Figure 4.2 — p0 et p100 calculés selon les différentes méthodes de calcul de l'indice pour un participant (utilisant les données agrégées à la seconde)	41
Figure 4.3 — p0 et p100 calculés selon les différentes méthodes de calcul de l'indice pour un participant (utilisant les données agrégées au stimulus)	44
Figure 4.4 — Image de calibration 4_3	53
Figure 4.5 — Graphique illustrant l'émotion (valence et activation) moyenne mesurée pour chaque méthode en comparaison avec la valeur théorique pour l'image de calibration 4_3.....	54
Figure 4.6 — Graphique illustrant l'émotion (valence et activation) mesurée pour chaque participant par méthode pour l'image de calibration 4_3	55
Figure 4.7 — Image de calibration 1_2	55
Figure 4.8 — Graphique illustrant l'émotion (valence et activation) moyenne mesurée pour chaque méthode en comparaison avec la valeur théorique pour l'image de calibration 1_2.....	56

Figure 4.9 — Graphique illustrant l'émotion (valence et activation) mesurée pour chaque participant par méthode pour l'image de calibration 1_2	57
Figure 4.10 — Image de calibration 2_1	57
Figure 4.11 — Graphique illustrant l'émotion (valence et activation) moyenne mesurée pour chaque méthode en comparaison avec la valeur théorique pour l'image de calibration 2_1	58
Figure 4.12 — Graphique illustrant l'émotion (valence et activation) mesurée pour chaque participant par méthode pour l'image de calibration 2_1	59
Figure 4.13 — Diagrammes en boîte de la valence mesurée physiologiquement pour chaque image de calibration (classés de manière croissante selon la moyenne)	63
Figure 4.14 — Diagrammes en boîte de la valence mesurée par le curseur affectif pour chaque image de calibration (classés de manière croissante selon la moyenne)	65
Figure 4.15 — Diagrammes en boîte de l'activation mesurée par l'EDA phasique et tonique pour chaque image de calibration (classés de manière croissante selon la moyenne).....	68
Figure 4.16 — Diagrammes en boîte de l'activation mesurée par l'EDA phasique pour chaque image de calibration (classés de manière croissante selon la moyenne)	69
Figure 4.17 — Diagrammes en boîte de l'activation mesurée par l'EDA phasique pour chaque image de calibration (classés de manière croissante selon la moyenne; omettant les valeurs extrêmes).....	70
Figure 4.18 — Diagrammes en boîte de l'activation mesurée par le HRV pour chaque image de calibration (classés de manière croissante selon la moyenne; omettant une valeur extrême)	71
Figure 4.19 — Diagrammes en boîte de l'activation mesurée par le curseur affectif pour chaque image de calibration (classés de manière croissante selon la moyenne)	72
Figure 4.20 — Diagrammes en boîte de l'analyse faciale pour chaque image de calibration en gardant que les observations où le participant a regardé pour au moins une seconde (classés de manière croissante selon la moyenne)	75
Figure 4.21 — Diagrammes en boîte de l'EDA phasique et tonique pour chaque image de calibration en gardant que les observations où le participant a regardé pour au moins une seconde (classés de manière croissante selon la moyenne)	76

Figure 4.22 — Diagrammes en boîte de l’EDA phasique pour chaque image de calibration en gardant que les observations où le participant a regardé pour au moins une seconde (classés de manière croissante selon la moyenne; omettant les valeurs extrêmes).....	77
Figure 4.23 — Diagrammes en boîte de l’HRV pour chaque image de calibration en gardant que les observations où le participant a regardé pour au moins une seconde (classés de manière croissante selon la moyenne; omettant une valeur extrême).....	78
Figure 4.24 — Diagrammes en boîte des méthodes mesurant la valence.....	79
Figure 4.25 — Diagrammes en boîte des méthodes mesurant l’activation (omettant les valeurs extrêmes)	80
Figure 4.26 — Graphique des valences et activations moyennes par image de calibration pour la technique du curseur affectif.....	81
Figure 4.27 — Graphique des valences et activations moyennes par image de calibration pour la technique de l’analyse faciale avec l’EDA phasique et tonique	82
Figure 4.28 — Graphique des valences et activations moyennes par image de calibration pour la technique de l’analyse faciale avec l’EDA phasique.....	82
Figure 4.29 — Graphique des valences et activations moyennes par image de calibration pour la technique de l’analyse faciale avec l’HRV	83
Figure 5.1 — Nuage de points des méthodes de calcul de l’indice original et avec les points extrêmes (données agrégées au stimulus et au participant).....	86
Figure A1.1 – Image de calibration 1_1.....	i
Figure A1.2 – Image de calibration 1_2.....	ii
Figure A1.3 – Image de calibration 1_3.....	iii
Figure A1.4 – Image de calibration 2_1.....	iv
Figure A1.5 – Image de calibration 2_2.....	v
Figure A1.6 – Image de calibration 2_3.....	vi
Figure A1.7 – Image de calibration 3_1.....	vii
Figure A1.8 – Image de calibration 3_2.....	viii
Figure A1.9 – Image de calibration 3_3.....	ix
Figure A1.10 – Image de calibration 4_1	x
Figure A1.11 – Image de calibration 4_2	xi
Figure A1.12 – Image de calibration 4_3	xii

Liste des abréviations

EDA = Activité Électrodermale

Indice = Index d'expérience émotionnelle

HRV = Variabilité de la fréquence cardiaque

SAM = Self-Assessment Manikin

Avant-propos

L'autorisation de rédiger ce mémoire a été obtenue par la direction du programme de M.Sc. de HEC Montréal.

Remerciements

En premier lieu, je tiens à remercier la fabuleuse équipe du Tech3Lab pour son soutien et son aide tout au long de ce projet de recherche. J'aimerais remercier en particulier David Briegne, Élise Labonte-LeMoyne, Audrey Valiquette, Vanessa Georges et François Courtemanche.

De plus, je voudrais remercier tous les membres de l'équipe statistique du Tech3Lab avec qui j'ai eu le plaisir de travailler pendant mon mémoire : Shang Lin Chen, Jared Boasen, Amay Cheam, Hakan Hekingil, Marin Opari, Élise Boissonneault, Florian Coustures et Zi-Su Zhao. Merci pour votre aide et vos conseils pendant ma recherche et ma rédaction!

Je veux également remercier mon directeur, Marc Fredette, qui m'a offert conseil, encouragement et soutien tout au long de mon mémoire. Sans son aide, ce travail ne serait jamais arrivé à conclusion.

Je tiens à remercier tous mes collègues, amis et famille qui m'ont donné du soutien, reconnaissance et amour pendant ma recherche et ma rédaction. Vous êtes trop nombreux pour remercier un à un dans cette section, mais sans votre soutien, je n'aurais pas eu la chance de finir ce travail.

Finalement, je tiens à remercier ma mère, Hélène. Sans ton aide et ton encouragement pendant les moments les plus difficiles, je n'aurais certainement pas eu la chance de continuer ce travail. Tu m'as permis de me lancer dans cette aventure et d'arriver à la fin! Les mots ne pourront jamais exprimer ma gratitude envers toi.

P.S. Merci à Métisse et Gaia, mes deux compagnonnes canines. Leur affection et leur soutien m'ont gardée énergisée pendant les longues journées de recherche et de travail.

Chapitre 1

Introduction

Dans le domaine de l'assurance, créer une application en ligne pour compléter les tâches clés comme les soumissions et les réclamations est un élément important dans la stratégie des entreprises de ce domaine. Afin de pouvoir identifier les points forts et faibles de leurs applications, il est nécessaire de faire l'évaluation de la performance de l'application. La performance dans ce contexte est étroitement liée avec les émotions générées par l'application. Par exemple, si un site génère des émotions positives fortes, on considérera que le site performe bien. À l'inverse, un site qui génère des émotions négatives fortes lors de son utilisation indique qu'il existe des problèmes avec l'application et que des optimisations sont nécessaires.

Pour notre partenaire industriel, il est donc important d'avoir un outil qui peut non seulement correctement mesurer la performance de ses applications, mais qui peut facilement transmettre cette information aux gestionnaires de ses applications.

Il est important de souligner que les émotions peuvent être mesurées de différentes manières. En expérience utilisateur (UX), deux variables sont utilisées pour mesurer l'émotion d'une personne : la valence et l'activation. La valence est une mesure qui permet de capter le type d'émotion ressentie par une personne (Russell et Mehrabian, 1977). Par exemple, une valence élevée veut donc dire qu'une émotion est positive et une valence négative implique que l'émotion ressentie est négative (Russell et Mehrabian, 1977). Dans le domaine du UX, l'utilisation de la mesure des expressions faciales est faite pour obtenir cette mesure de manière physiologique.

L'activation est une mesure qui permet de détecter le niveau d'intensité d'une émotion (Russell et Mehrabian, 1977). Donc, un niveau d'activation élevé signifie qu'une personne ressent une émotion forte (indépendamment du type d'émotion ressentie). En expérience utilisateur, l'utilisation de l'activité électrodermale est faite pour obtenir une mesure de l'activation de manière physiologique.

En combinant ces deux mesures, il est possible d'inférer l'émotion générale ressentie par une personne. Typiquement, la méthode utilisée pour faire ceci est de faire une représentation graphique en deux dimensions. Malgré la nature visuelle de cette méthode, afin de correctement interpréter les résultats, il est quand même nécessaire de connaître la signification des deux mesures et de savoir comment lire le graphique.

De plus, le partenaire industriel a remarqué que d'autres facteurs pouvaient jouer sur l'émotion ressentie lors de l'utilisation d'une application. Donc, s'il désire ajouter des mesures additionnelles pour mesurer l'émotion, il est nécessaire de trouver une nouvelle manière de mesurer la performance de leur site.

C'est pour cette raison que notre partenaire industriel était très intéressé par l'index d'expérience émotionnelle (l'indice), une mesure développée par le laboratoire de recherche Tech3Lab, le plus important laboratoire de recherche en expérience utilisateur en Amérique du Nord. Cet indice permet d'interpréter l'émotion ressentie par une personne sur une échelle de 0 à 100 où 0 représente la pire émotion ressentie (émotion fortement négative) par une personne et 100 représente la meilleure émotion ressentie (émotion fortement positive) lors d'une expérimentation.

La méthode utilisée par l'index d'expérience émotionnelle pour calculer l'émotion à un moment précis se base sur une période de calibration où un participant à une expérience doit regarder une série d'images reconnue pour générer une émotion particulière. Par exemple, une image d'un phoque blessé génère une émotion fortement négative. Pendant que le participant regarde cette série d'images, sa valence et son activation sont mesurées afin de déterminer l'émotion la plus négative (valence négative et activation positive) et l'émotion la plus positive (valence positive et activation positive). Ces deux points d'émotion sont établis comme le point 0 (la pire émotion) et le point 100 (la meilleure émotion).

Par la suite, lors de l'expérimentation principale, les valences et activations mesurées sont comparées avec le point 0 et le point 100 à l'aide de la distance euclidienne afin de déterminer la valeur de l'indice à cette émotion précise. Avec cette unique mesure, il est donc possible de rapidement détecter lors d'un test de la performance d'une application

des moments où l'émotion devient négative et les moments où des émotions positives sont détectées.

Malgré l'intérêt initial, le partenaire ne voulait pas simplement intégrer l'indice dans leur entreprise, il voulait approfondir le travail déjà fait afin de le rendre plus sensible aux besoins de l'industrie de l'assurance. Également, il voulait avoir une méthode plus portative pour déployer l'indice qui n'était pas dépendant sur une phase de calibration. C'est avec ces demandes qu'un projet ait été lancé au Tech3Lab pour raffiner et développer une nouvelle version de l'index d'expérience émotionnelle pour l'industrie de l'assurance.

Dès la première tâche de ce projet, de modifier l'indice pour ne plus dépendre d'une phase de calibration, des problèmes ont été rencontrés. Les expériences effectuées pour créer une version de l'indice qui dépend uniquement des données d'une expérience ont révélé qu'il y avait très peu de lien entre l'indice original et les nouvelles versions de l'indice. Donc, afin de pouvoir garder les acquis obtenus lors de la première étude sur l'indice, il est devenu apparent qu'il fallait garder cette phase de calibration.

Ensuite, dans le but de mieux comprendre l'importance de la phase de calibration, une exploration de cette phase et des images utilisées a été faite. Ce travail exploratoire a révélé un problème plus important avec l'étude originale que les émotions mesurées ne semblent pas être significativement différentes les unes des autres et qu'elles ne semblent pas concorder avec l'émotion que l'image est supposée générer.

Ceci a mené à un questionnement sur la fiabilité des émotions théoriques proposées par ces images de calibration compte tenu de l'âge de ces banques d'images et de la possibilité que les réactions typiques à ces images auraient changé avec le temps. Le consensus scientifique concernant ces banques d'images est que les émotions théoriques concordent avec les émotions ressenties par les participants lors d'expérience (Bradley et Lang, 2007; Dufey et al., 2010; Soares et al., 2014; Mačiukaitė et al., 2015; Gong et Wang, 2016). De plus, en utilisant les valeurs de la valence et de l'activation autodéclarées lors de la phase de calibration, ces mesures permettent de distinguer entre les émotions ressenties par chaque image et ont un lien fort avec les données émotionnelles théoriques.

Dans le contexte de ces étapes précédentes, un impact théorique concernant l'utilisation de la phase de calibration peut être fait. Considérant que les données obtenues à l'aide de mesures physiologiques concordent peu avec les valeurs théoriques, il est intéressant d'explorer l'implémentation des bases d'images dans le contexte d'une étude mesurant l'émotion physiologiquement. Par exemple, même s'il existe un lien entre les mesures physiologiques de l'émotion et les valeurs théoriques de l'émotion de ces banques d'images (Bradley et Lang, 2007), il pourrait être plus facile de recalculer les mesures théoriques avec des outils physiologiques afin d'avoir un point de référence direct entre une mesure théorique et une mesure obtenue lors d'une expérimentation.

Parallèlement à ceci, l'impact industriel de ces résultats serait que, étant donné que l'utilisation de données autodéclarées lors des phases de calibration est plus fiable, il serait donc possible de rendre le processus d'évaluation de la performance plus portatif en utilisant ces données autodéclarées, ce qui demande moins d'équipement spécialisé et pourrait même être fait à distance.

Afin d'explorer plus en détail la question de recherche, une revue de la littérature sera faite (chapitre 2) suivie d'une présentation de la méthodologie utilisée et du contexte des données initiales (chapitre 3). Par la suite, une analyse des résultats sera présentée (chapitre 4) et une discussion de ces résultats sera faite (chapitre 5) pour conclure avec une présentation des contributions théoriques et industrielles et d'une présentation des avenues de recherches futures (chapitre 6).

Chapitre 2

Revue de la littérature

Résumé

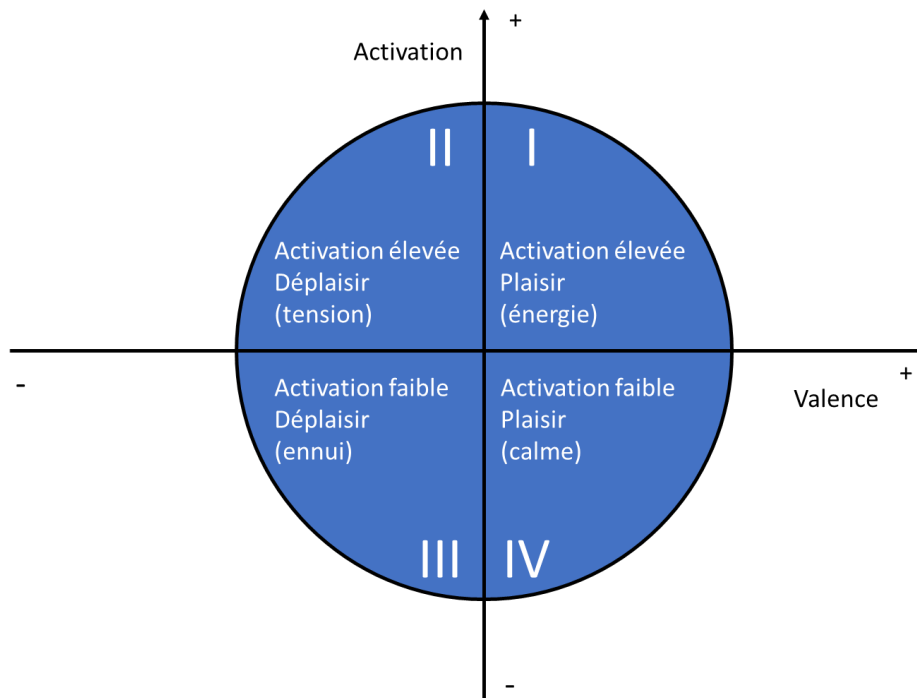
Le projet de recherche présenté dans ce mémoire se base sur plusieurs notions psychologiques qui visent à mesurer l'émotion vécue par une personne. L'utilisation de ces concepts a permis à l'étude originale de l'index d'expérience émotionnelle de développer la méthodologie utilisée lors de la phase expérimentale et, tant pour l'étude originale que ce projet, permet de faire des interprétations des résultats qui se fient sur un consensus scientifique. Ce chapitre expliquera ces concepts ainsi que les outils utilisés pour mesurer l'information nécessaire à la réalisation de ce projet de recherche. Les trois principales parties de ce chapitre sont une présentation de la théorie circumplex de l'émotion, les outils de mesure de l'émotion et les outils pour générer des émotions spécifiques.

2.1 La théorie circumplex de l'émotion

Le projet de recherche original et, par conséquent, ce projet de recherche ont besoin de mesurer les émotions vécues par des individus. La théorie utilisée pour décrire les émotions est celle du circumplex de l'émotion, qui propose de décomposer une émotion en deux dimensions (Russell, 1980). La première dimension, qui est typiquement présentée sur l'axe horizontal (figure 2.1), mesure la valence tandis que la deuxième dimension, qui est sur l'axe vertical, mesure l'activation (Russell, 1980). Concernant le premier axe, la valence permet d'identifier l'émotion sur une échelle de déplaisir (valence négative) à plaisir (valence positive) (Russell et Mehrabian, 1977). Pour sa part, l'activation caractérise l'intensité de l'émotion vécue par une personne (Russell et Mehrabian, 1977). La théorie présente également la présence d'une troisième dimension qui mesure le sentiment de contrôle émotionnel selon une échelle contrôlée-domination (Russell et Mehrabian, 1977). Selon Bradley et Lang (2007), cette dimension est moins importante que la valence et l'activation dans le contexte d'études empiriques. Bref, afin

de correctement mesurer une émotion, il est nécessaire d'identifier le type d'émotion (valence) et son intensité (activation).

Figure 2.1 — Schéma illustrant la théorie circumplex de l'émotion



Il est important de souligner que le projet de recherche original ainsi que ce projet de recherche se basent sur l'idée que le quadrant de valence positive et activation élevée (quadrant I) correspond à celui le plus désirable. Cette idée découle de l'intuition des partenaires industriels qu'une émotion intensément positive serait la meilleure émotion que leur produit ou service pourrait générer. En suivant cette logique, le quadrant IV sera la seconde meilleure région (valence positive, activation faible), le quadrant III serait la troisième meilleure région (valence négative, activation faible) et le quadrant II (valence négative et activation élevée) serait la pire région émotionnelle. Par exemple, lors d'une expérimentation, si un stimulus génère l'émotion avec la valence et l'activation la plus grande, ce serait le stimulus qui génère l'émotion la plus désirable et positive. À l'inverse, un stimulus qui génère la combinaison de la plus petite valence et la plus grande activation observée serait le stimulus qui génère l'émotion la moins désirable et la plus négative.

Bref, cette logique sera importante pour le calcul et le fonctionnement de l'index d'expérience émotionnelle, qui sera présenté en détail dans le chapitre suivant.

2.2 Outils mesurant l'émotion

En se fiant sur la théorie circumplex de l'émotion, plusieurs outils ont été développés afin de pouvoir mesurer les deux composantes de l'émotion. Dans cette section, la présentation de ces outils sera divisée en deux : dans la première moitié de la section, les outils permettant de mesurer l'activation et la valence d'une personne de manière autodéclarée seront présentés et, dans la deuxième partie de la section, les outils utilisés afin de mesurer l'émotion par le moyen de l'enregistrement d'une réponse physiologique à un stimulus seront illustrés.

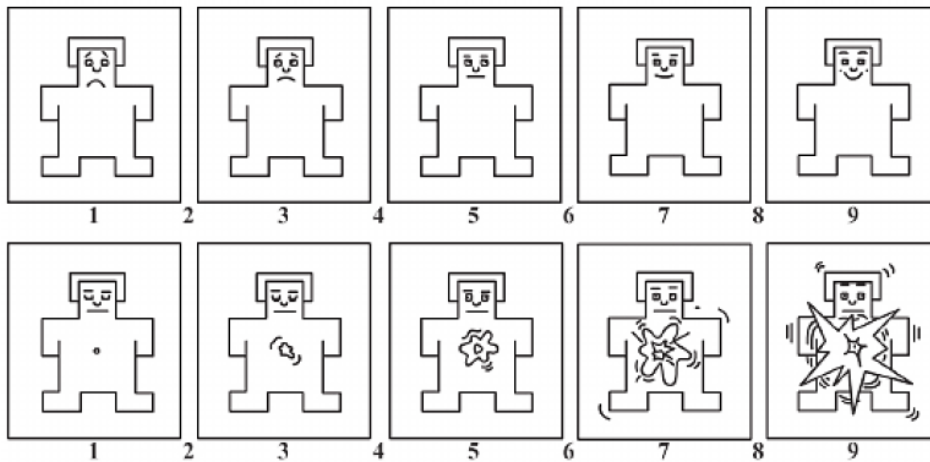
2.2.1 Outils mesurant l'émotion de manière autodéclarée

Les outils pour mesurer l'émotion de manière autodéclarée sont des moyens qui utilisent un questionnaire qui doit être rempli par un participant. Ces méthodes permettent de transmettre l'émotion vécue par une personne de manière verbale, dite subjective, ce qui inclut les réponses remises à l'écrit (Zaman et Shrimpton-Smith, 2006). Les deux méthodes autodéclarées pertinentes pour ce projet de recherche sont le Self-Assessment Manikin (SAM) et le curseur affectif.

2.2.1.1 Self-Assessment Manikin

Le SAM est un questionnaire visant à mesurer les composantes de l'émotion séparément par le moyen de pictogrammes illustrant différents niveaux d'émotion sur un personnage (Bradley et Lang, 2007). Les résultats sont notés sur une échelle allant de 1 à 9 pour les deux composantes des émotions. Pour la valence, l'échelle va de 1 (déplaisir) à 9 (plaisir) et, pour l'activation, va de 1 (calme) à 9 (activé). Ces pictogrammes sont présentés à la figure 2.2. L'échelle du haut représente la valence et l'échelle du bas représente l'activation.

**Figure 2.2 — Illustration de l'échelle SAM pour la valence (échelle du haut)
et l'activation (échelle du bas)**



La validité de l'échelle SAM a été démontrée par une forte relation avec une méthode acceptée pour mesurer l'émotion vécue par une personne (Bradley et Lang, 1994). La méthode comparée, l'échelle de différentiel sémantique, est mesurée sur 18 paires d'adjectifs opposés (par exemple, non satisfait-satisfait ou endormi-réveillé) et une analyse factorielle de valeurs obtenues par ces paires permet d'identifier la valence et l'activation (Bradley et Lang, 2007). Les résultats d'une analyse de la corrélation entre les facteurs de valence et d'activation pour les paires de l'échelle de différentiel sémantique avec les résultats de la valence et de l'activation pour l'échelle SAM démontrent une très forte corrélation ($r = 0,98$ pour le lien entre la paire sémantique déplaisir-plaisir et l'échelle de SAM pour la valence et $r = 0,90$ pour le lien entre la paire sémantique non-excité-excité et l'échelle de SAM pour l'activation), donc la performance du SAM pour mesurer l'émotion et sa validité a été démontré (Bradley et Lang, 1994). Donc, l'échelle SAM permet d'identifier la valence et l'activation vécue par une personne de manière fiable.

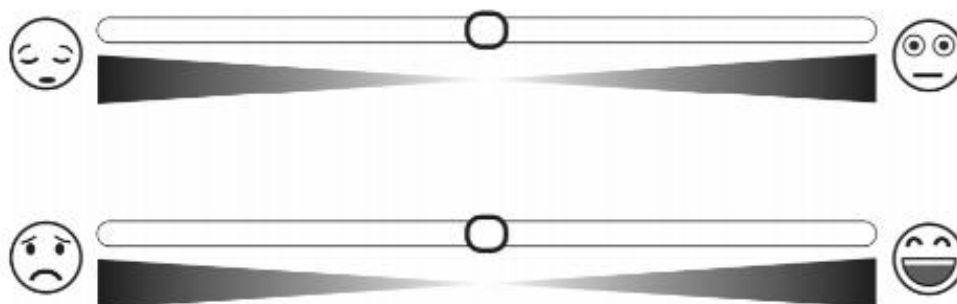
2.2.1.2 Curseur affectif

Dans le contexte actuel du Tech3Lab de HEC Montréal, les réponses des participants sont maintenant enregistrées par un sondage répondu sur un ordinateur ou une tablette. L'échelle de SAM a été conçue dans un contexte de réponses enregistrées sur papier avec

crayon (Bradley et Lang, 2007). Malgré l'existence d'une version pour affichage sur ordinateur de l'échelle SAM (Bradley et Lang, 2007), les principes fondamentaux de la structure de cette échelle sont déconnectés des principes modernes de l'interface utilisateur (Betella et Verschure, 2016). Ces chercheurs expliquent qu'il est fréquent dans leurs études utilisant le SAM, que les participants demandent d'avoir plus de détails sur le fonctionnement de ce dernier, particulièrement au niveau de la signification des pictogrammes (Betella et Verschure, 2016).

En tenant compte de ceci, un nouvel outil a été développé afin de combler aux lacunes du SAM. Cet outil est le curseur affectif. Le curseur mesure, séparément, la valence et l'activation sur une échelle allant de 0 à 100. Sur chaque côté du curseur, il y a une émoticône représentant les valeurs opposées de l'échelle (Betella et Verschure, 2016). Pour la valence, il y a le déplaisir à gauche et le plaisir à droite. Pour l'activation, il y a une émoticône endormie à gauche et, à droite, il y a une émoticône très réveillée. En dessous du curseur se trouvent deux triangles isocèles inversés (ils pointent vers le milieu du curseur) qui ont pour but d'illustrer l'intensité de l'émotion (Betella et Verschure, 2016). Finalement, dans le but d'éviter l'influence que pourrait causer l'utilisation de couleurs, le curseur est toujours présenté en noir et blanc (Betella et Verschure, 2016). La figure 2.3 présente le curseur affectif avec le curseur du haut étant celui de l'activation et celui du bas étant la valence.

**Figure 2.3 — Curseur affectif mesurant l'activation (échelle du haut)
et la valence (échelle du bas)**



Une relation entre l'échelle SAM et le curseur affectif peut être établie par le fait que la corrélation de Spearman de l'émotion mesurée (valide tant pour la valence que

l'activation) par le SAM est très forte, positive et significative avec celle du curseur, au niveau de $r_s = 0,85$ ($p\text{-value} < 0,001$) pour la valence et au niveau de $r_s = 0,86$ ($p\text{-value} < 0,001$) pour l'activation (Betella et Verschure, 2016).

Avec la pertinence de cet outil validé, il est possible de se concentrer sur les avantages identifiés pour le curseur. Le premier avantage est le fait qu'aucune instruction verbale ou écrite n'est nécessaire pour correctement utiliser le curseur (Betella et Verschure, 2016). Le fait que l'outil utilise un curseur et des émoticônes permet à un utilisateur de se fier sur leurs connaissances des interfaces utilisateur, sans avoir besoin d'instructions sur son fonctionnement (Betella et Verschure, 2016). L'autre grand avantage est la granularité de l'outil. En utilisant une échelle de 0 à 100, le curseur permet d'enregistrer une intensité d'émotion plus nuancée que l'échelle à 9 points du SAM (Betella et Verschure, 2016). En somme, le curseur affectif est un outil plus adapté à la réalité des questionnaires remplis par ordinateur tout en restant une mesure fiable de l'émotion vécue par une personne à la suite d'un stimulus.

Cette sous-section a présenté deux méthodes pour mesurer l'émotion d'une personne de manière autodéclarée : l'échelle SAM et le curseur affectif. Ces méthodes sont robustes pour mesurer l'activation et la valence d'une personne et les résultats obtenus par une méthode devraient donner des résultats compatibles avec l'autre. Bref, que ce soit le curseur affectif ou l'échelle SAM, ces techniques sont des méthodes, pour mesurer l'émotion, pertinentes dans le contexte de projets de recherche étudiant les émotions.

2.2.2 Outils mesurant l'émotion par des moyens physiologiques

Dans un contexte de vouloir minimiser les biais, une seconde catégorie de techniques et méthodes pour mesurer l'émotion, qui se fient sur des mesures non verbales, dites objectives, a été développée (Zaman et Shrimpton-Smith, 2006). Ces mesures non verbales concernent principalement l'élément physiologique (les changements dans le système nerveux autonome) d'une émotion vécue et ont comme avantages le fait de ne pas interférer avec un participant pendant l'étude et d'être des mesures continues dans le temps (Zaman et Shrimpton-Smith, 2006). De plus, elles évitent de se fier sur des informations fournies par les participants, qui pourraient être faussées afin de respecter

un certain facteur de désirabilité sociale (Larsen, Norris et Cacioppo, 2003). Cela étant dit, plusieurs études ont démontré que les méthodes non verbales mesurent les émotions de manière qui est liée aux techniques verbales (par exemple, un signal mesurant l'activation de manière physiologique devrait être lié à une mesure autodéclarée de l'activation) (Bradley et Lang, 2007). Également, avec l'exception des études où le but est d'isoler un processus cognitif, les études utilisant des mesures non verbales de l'émotion retiennent un fort niveau de validité écologique (Labonte-LeMoyne, Courtemanche, Fredette et Léger, 2018). Trois signaux physiologiques seront présentés dans cette section : l'analyse faciale, l'activité électrodermale et la variabilité du rythme cardiaque.

2.2.2.1 Analyse faciale

Une méthode utilisée pour mesurer la valence de manière physiologique est par le moyen d'une analyse faciale d'un individu, particulièrement par le moyen du logiciel FaceReader, considéré comme un outil efficace pour mesurer l'émotion (Zaman et Shrimpton-Smith, 2006). Ce logiciel permet de détecter six expressions émotionnelles (content, triste, fâché, surpris, peur et dégoût) en plus de reconnaître un état émotionnel neutre (Loijens, Krips, van Kuilenburg, den Uyl et Iva, 2015). De cette détection des émotions, il est possible d'extrapoler une mesure de la valence (Loijens et Krips, 2018). Cela étant dit, la détection des émotions vécues est la conclusion de plusieurs étapes importantes. Le premier objectif du logiciel est de reconnaître un visage dans un enregistrement vidéo (Loijens et al., 2015). Afin d'identifier un visage, le logiciel emploie l'algorithme Viola-Jones, un algorithme conçu pour reconnaître un visage dans des images (Loijens et al., 2015). Avec le visage identifié, le logiciel peut commencer une étape de modélisation du visage à l'aide de l'approche algorithmique *Active Appearance Model* (AAM) (Loijens et al., 2015). Cet algorithme permet de créer un modèle du visage qui décrit 500 points clés de la face et de la texture faciale (Loijens et al., 2015). Finalement, une identification des expressions est effectuée à l'aide d'un réseau de neurones artificiel (Loijens et al., 2015). Ce réseau a été entraîné à l'aide de 10 000 images annotées afin de pouvoir classer un visage dans les sept catégories mentionnées précédemment (les six expressions et l'état neutre) (Loijens et al., 2015).

Avec l'analyse des expressions effectuée, le logiciel peut effectuer un calcul de la valence d'un individu à un moment donné.

Notons :

Valence = La valeur de la valence

Happy = La valeur de classification accordée à l'émotion *content*

Sad = La valeur de classification accordée à l'émotion *triste*

Disgust = La valeur de classification accordée à l'émotion *dégoûté*

Angry = La valeur de classification accordée à l'émotion *fâché*

Scared = La valeur de classification accordée à l'émotion *peur*

La formule est la suivante (Loijens et Krips, 2018) :

$$Valence = Happy - MAX(Sad, Disgust, Angry, Scared) \quad (1)$$

Il est à noter que, comme l'état « surpris » peut être considéré comme une émotion positive ou négative, dépendant du contexte, cette émotion est exclue du calcul de la valence (Loijens et Krips, 2018). La capacité de FaceReader à mesurer les émotions est supérieure à la méthode traditionnelle de mesurer les émotions, la prise en note manuelle des émotions par un chercheur (Zaman et Shrimpton-Smith, 2006). Les résultats de FaceReader sont très similaires aux résultats qu'un chercheur expérimenté pourrait remarquer, mais avec l'avantage que tous les changements d'émotions peuvent être capturés, même les changements les plus nuancés (Zaman et Shrimpton-Smith, 2006). En somme, l'analyse faciale par le moyen de FaceReader permet de faire un enregistrement fiable de la valence d'une émotion vécue par une personne.

2.2.2.2 Activité électrodermale

L'activité électrodermale (EDA) demeure une des mesures physiologiques les plus couramment utilisées (Dawson, Schell et Fillion, 2007). Elle correspond aux changements de la conductivité des glandes de sudation de la peau, particulièrement les glandes eccrines sur les paumes de la main, car elles sont plus sensibles aux stimuli psychologiques qu'aux stimuli thermiques (Posada-Quintero et Chon, 2020). Ce signal demeure populaire parce qu'il peut être utilisé afin d'identifier plusieurs réactions chez un individu, particulièrement la charge cognitive, l'intérêt et l'activation (Dawson et al., 2007).

Cependant, il est important de spécifier que l'EDA n'est pas un signal unique, il est composé de deux signaux : le signal tonique et le signal phasique (Braithwaite, Watson, Jones et Rowe, 2015). La décomposition de ces signaux est possible grâce à un modèle qui décrit l'EDA comme la combinaison linéaire des deux signaux précédents et d'une partie de bruit qui représente les erreurs du modèle, les erreurs de mesure et les artefacts (Posada-Quintero et Chon, 2020). Le modèle parvient à séparer les composantes par sa construction qui repose sur la statistique bayésienne, l'optimisation convexe et le principe d'activité à faible densité (Greco, Valenza, Lanata, Scilingo et Citi, 2015).

La composante tonique de l'EDA représente les éléments qui changent lentement dans le signal (Braithwaite et al., 2015) et peut être interprétée comme la valeur moyenne de l'EDA (Posada-Quintero et Chon, 2020). Le niveau de conductivité de la peau (SCL) représente une mesure fréquemment utilisée pour mesurer la composante tonique du signal de l'EDA (Braithwaite et al., 2015). Le signal phasique correspond aux éléments de l'EDA qui changent rapidement à la suite d'une réaction à un stimulus (Posada-Quintero et Chon, 2020). Ces réactions sont associées à la réponse de la conductivité de la peau (SCR) et peuvent être classées en deux catégories : les SCRs non spécifiques (NS-SCR), qui sont des SCR qui ne peuvent pas être associées à un stimulus en particulier, et les SCR reliés à un événement (ER-SCR), donc qui peuvent être associées à un stimulus (Braithwaite et al., 2015). Afin d'identifier les ER-SCR, Braithwaite et al. (2015) proposent deux conditions à respecter lorsqu'un SCR est observé : la latence (le délai entre le stimulus et la réponse physiologique) doit être entre 1 et 3 secondes et la valeur

mesurée doit atteindre un certain seuil (typiquement 0,05 microsiemens [μ S], mais des valeurs plus petites, si mesurables, peuvent être considérées). Étant donné l'intérêt prononcé pour les SCR phasiques dans la recherche (Braithwaite et al., 2015), ce projet de recherche analysera l'activation mesurée par la combinaison des signaux toniques et phasiques et l'activation mesurée uniquement par le signal phasique. Il faut donc conclure que l'EDA est une mesure essentielle pour mesurer l'émotion physiologiquement due à sa capacité à mesurer l'activation.

2.2.2.3 Variabilité du rythme cardiaque

La dernière méthode pour mesurer l'émotion de manière physiologique, qui sera étudiée dans ce projet de recherche, est la variabilité du rythme cardiaque (HRV). Le HRV correspond à la variation dans l'intervalle de temps entre deux battements de cœur consécutifs (Malik et al., 1996). La méthode pour calculer le HRV est avec l'intervalle NN, qui correspond à l'écart entre les pics des battements de cœur observés (NN faisant référence au fait que ce sont les battements de cœur *normaux* qui sont étudiés) (Malik et al., 1996). Un lien peut être identifié entre le HRV et les émotions (Lane et al., 2009). Ce lien est établi spécifiquement entre l'HRV et l'activation (Nickel et Nachreiner, 2003). En conclusion, comme le HRV peut être identifié à la fois comme un potentiel indicateur de l'activation, cette mesure sera considérée dans le projet de recherche.

Cette sous-section a identifié trois signaux physiologiques pouvant mesurer des composantes de l'émotion. L'analyse faciale par le moyen du logiciel FaceReader permet de mesurer la valence ressentie par une personne à un moment donné d'une étude à l'aide d'un réseau de neurones pour classer le visage à partir de 500 points identifiés sur une face. L'EDA est une mesure de l'activation qui peut être analysée sous deux angles : soit en combinant les signaux toniques et phasiques ou en analysant uniquement le signal phasique. Le HRV est une mesure de l'intervalle entre deux battements du cœur qui peut être liée à la valence et à l'activation. Ces méthodes seront utilisées afin de mesurer l'émotion vécue par une personne de manière non verbale.

2.3 Outils pour générer des émotions spécifiques

Cette dernière section du chapitre présente des outils qui sont essentiels au projet de recherche parce qu'ils permettent de générer des émotions standardisées et attendues. Ces outils sont des bases de données d'images qui ont été évaluées afin qu'elles puissent être utilisées dans une étude afin de générer une certaine valeur de valence et d'activation chez un individu (Bradley et Lang, 2007). Les bases de données étudiées sont l'International Affective Picture System et le Geneva Affective Picture Database.

2.3.1 International Affective Picture System

L'International Affective Picture System (IAPS) est une base de données de plus de 1000 images illustrant des réalités de la vie humaine (Lang, Bradley et Cuthbert, 2005). Les images visent à générer des émotions qui touchent les quatre quadrants de la théorie circumplex de l'émotion (Bradley et Lang, 2007). Les valeurs de l'émotion (valence et activation) assignées à une image IAPS sont déterminées par une étude où une proportion identique d'hommes et de femmes sont exposés aux images et doivent donner leur note pour la valence et l'activation sur l'échelle SAM (Bradley et Lang, 2007). Les images sont présentées une à la fois, pour une période de 6 secondes; les participants ont ensuite 15 secondes pour octroyer une valeur de valence et d'activation avant de recevoir un avertissement pour une durée de 5 secondes que la prochaine image sera affichée (Bradley et Lang, 2007). Il est important de noter que les valeurs de valence et d'activation ont été étudiées afin de vérifier si les valeurs standardisées sont applicables pour peu importe la culture du participant (Dufey et al., 2010; Soares et al., 2015; Mačiukaitė et al., 2015; Gong et Wang, 2016) et les résultats permettent de conclure que l'IAPS est une base de données valide à travers les multiples cultures. Bref, l'IAPS est une base de données pertinente dans le contexte où générer des émotions lors d'une étude est nécessaire.

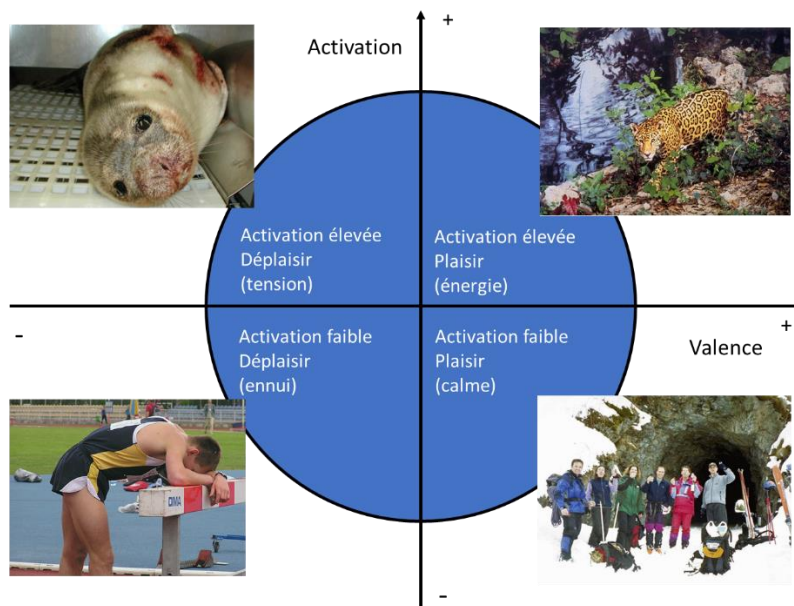
2.3.2 Geneva Affective Picture Database

Le Geneva Affective Picture Database (GAPED) est une base de données semblable à l'IAPS qui vise à combler un nombre limité d'images en lien avec certaines thématiques (Dan-Glauser et Scherer, 2011). Les thématiques visées par GAPED sont principalement à valence négative et sont les araignées, les serpents (choisies parce qu'il y a un fort

nombre d'études sur les phobies en lien avec ces animaux), les violations des droits de la personne et la maltraitance des animaux (choisies parce qu'elles sont en lien avec le bris des normes sociales/morales ou des normes légales) et les images générant des émotions neutres (Dan-Glauser et Scherer, 2011). Les valeurs de la valence et de l'activation ont été mesurées de manière verbale à l'aide d'une échelle allant de 0 à 100 qui utilisait des mots (par exemple « négatif » et « positif » pour l'échelle de la valence) pour identifier les extrêmes des axes (Dan-Glauser et Scherer, 2011). En somme, cette base de données est complémentaire à l'IAPS et peut être pertinente dans le contexte où des stimuli à valence négative devraient être utilisés.

Cette section a présenté deux bases de données d'images ayant pour but de cataloguer des stimuli qui génèrent des émotions (valence et activation) standardisées. La figure 2.4 illustre quatre images de ces banques, chacune représentant un des quadrants de la théorie circumplex de l'émotion. Ces deux outils sont complémentaires et peuvent être utilisés dans la même étude.

Figure 2.4 — Images des banques IAPS et GAPED générant des émotions selon la théorie circumplex de l'émotion



2.4 Conclusion

Ce chapitre a démontré les principes théoriques utilisés par le projet de recherche. Le pilier fondamental du projet est la théorie circumplex de l'émotion qui propose de décomposer l'émotion en deux axes, soit la valence (le type d'émotion) et l'activation (l'intensité de l'émotion). En se fiant sur cette théorie, deux grandes familles d'outils pour mesurer l'émotion ont été analysées : les outils autodéclarés et les outils physiologiques. L'échelle SAM et le curseur affectif sont des méthodes pour mesurer la valence et l'activation sur la base d'une réponse soumise par un participant à une étude. L'analyse faciale par FaceReader, l'EDA et le HRV sont des mesures physiologiques pouvant mesurer la valence et l'activation de manière non verbale et pendant qu'un stimulus est présenté à une personne. Finalement, deux bases de données permettant de générer des émotions selon des valeurs standardisées, IAPS et GAPED, ont été présentées.

Chapitre 3

Données initiales et méthodologie

Résumé

Dans le contexte de ce projet de recherche, les données d'une étude précédente du Tech3Lab, qui agit comme point de départ du projet actuel, ont dû être transformées et adaptées afin d'être utilisables. Par la suite, ces données ont été utilisées selon des techniques particulières dans le but de répondre aux questions de recherche. Ce chapitre donne le détail sur l'origine des données utilisées, les modifications qui ont été nécessaires à effectuer et la méthodologie utilisée pour réaliser ledit projet.

3.1 Description de l'index d'expérience émotionnelle

3.1.1 Présentation de l'indice

L'index d'expérience émotionnelle (l'indice) a été développé en partenariat entre la Chaire de recherche CRSNG-Prompt en expérience utilisateur, le Tech3Lab et un partenaire industriel de la Chaire en 2018. L'idée derrière l'indice est de trouver « une manière novatrice de quantifier de manière unidimensionnelle la réaction émotionnelle, permettant ainsi de créer un classement des interfaces en fonction de leur capacité à générer des réactions émotionnelles. » (Léger et al., 2018) La méthode retenue pour mesurer la réaction émotionnelle d'une personne est l'utilisation de signaux physiologiques pour décrire les deux composantes de l'émotion, la valence et l'activation. Le signal retenu pour décrire la valence est l'émotion selon une analyse faciale automatique par FaceReader, tandis que le signal retenu pour exprimer l'activation est l'EDA combinant le signal phasique et tonique. Afin de mesurer l'émotion de manière autodéclarée, le curseur affectif a été retenu comme outil à utiliser.

Il est important de souligner que l'utilisation originale de l'indice n'était pas liée à une industrie en particulier et a été utilisée pour analyser des domaines tels que l'industrie financière, les supermarchés, l'industrie des télécommunications, etc. Donc, l'origine de l'indice est pour simplifier la présentation de l'émotion d'une personne dans un contexte généralisé.

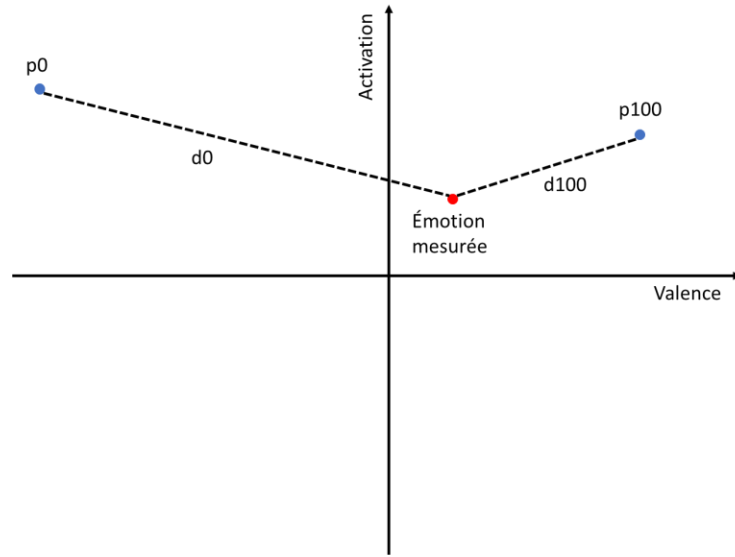
3.1.2 Calcul de l'indice

Maintenant que le principe général de l'indice a été présenté, il est possible de décrire le calcul de l'index d'expérience émotionnelle. La logique principale de l'indice est de convertir les deux composantes de l'émotion (valence et activation) en une seule valeur.

La première étape pour calculer l'indice correspond à enregistrer les composantes de l'émotion pour des données de phase de calibration. Ces données serviront à standardiser les données d'un participant et pour établir l'émotion la plus négative (le point avec la combinaison de l'activation la plus élevée et la valence la plus petite), dite point 0 (p0), et l'émotion la plus positive (le point avec la combinaison de l'activation la plus élevée et la valence la plus grande), dite point 100 (p100), vécue par un participant. Comme mentionné au chapitre 2, l'identification de p0 et p100 se base sur la logique que le quadrant I représente la région la plus désirable pour une émotion (valence positive et activation élevée) et que le point le plus loin des coordonnées (0,0) représente la version la plus désirable d'une émotion positive. Pareillement, le quadrant II représente la région la moins désirable pour une émotion (valence négative et activation élevée) et le point le plus loin des coordonnées (0,0) représente la version la moins désirable d'une émotion négative.

Par la suite, un participant est exposé à un stimulus et les valeurs d'activation et de valence sont enregistrées afin de pouvoir calculer la distance euclidienne entre l'émotion mesurée et p0 et p100 respectivement. En dernier lieu, l'indice est calculé en identifiant le ratio que la distance entre l'émotion et p0 (d0) représente sur la somme de d0 et la distance entre l'émotion et p100 (d100). La figure 3.1 illustre la relation entre p0, p100, une émotion mesurée, d0 et d100.

Figure 3.1 — Exemple de la relation entre les composantes du calcul de l'indice



Notons :

V_e = La valence de l'émotion générée par le stimulus

A_e = L'activation de l'émotion générée par le stimulus

V_{p0} = La valence au point p0

A_{p0} = L'activation au point p0

V_{p100} = La valence au point p100

A_{p100} = L'activation au point p100

$d0$ = Distance euclidienne de l'émotion à p0

$d100$ = Distance euclidienne de l'émotion à p100

Avec ces points de référence établis, il est possible de déterminer, pour une émotion vécue par une personne pendant un stimulus, la distance euclidienne de cette émotion de p0 et de p100 :

$$d0 = \sqrt{(V_e - V_{p0})^2 + (A_e - A_{p0})^2} \quad (2)$$

$$d100 = \sqrt{(V_e - V_{p100})^2 + (A_e - A_{p100})^2} \quad (3)$$

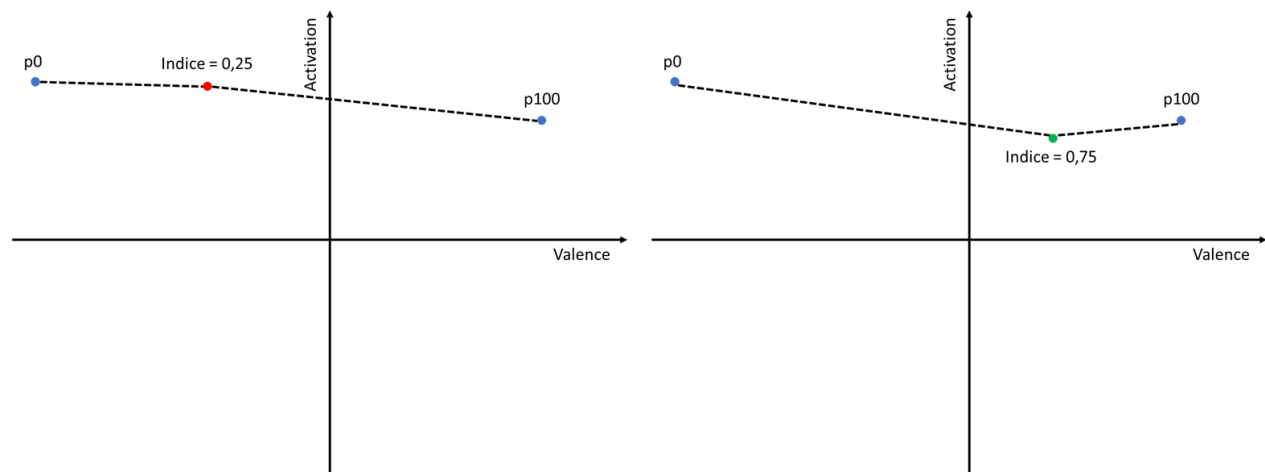
Finalement, il est possible d'établir la valeur de l'indice en faisant la proportion que représente d_0 sur la somme des distances :

$$i = d_0 / (d_0 + d_{100}) \quad (4)$$

Où i représente la valeur de l'indice.

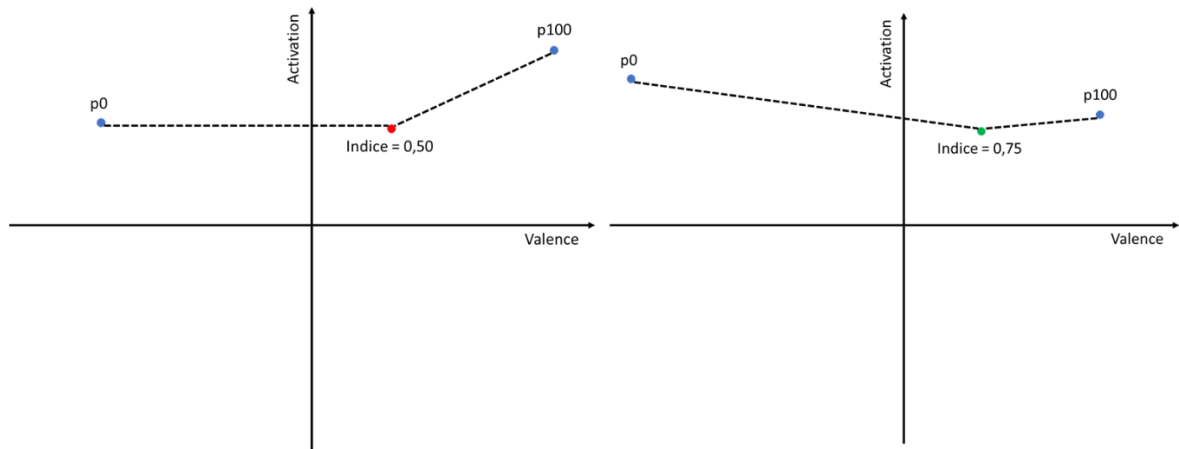
Par exemple, si une émotion vécue se retrouve à équidistance de d_0 et d_{100} , l'indice sera à 0,5; si l'émotion se trouve à être au point p_{100} , l'équation reviendra à d_0/d_0 , qui donne 1; et si l'émotion se trouve à être au point p_0 , l'équation reviendra à $0/d_{100}$, qui donne 0. La figure 3.2 illustre un exemple d'une valeur avec un indice faible et un exemple d'une valeur avec un indice élevé.

Figure 3.2 — Exemples d'émotions mesurées dans une étude et leur valeur de l'indice (faible à gauche; forte à droite)



L'indice permet d'aplatir deux mesures qui peuvent être difficilement interprétées ensemble tout en incorporant un élément de calibration qui reflète la variabilité de l'émotion vécue par une personne au lieu d'une standardisation sur tous les participants d'une étude. La figure 3.3 illustre ce fait en comparant la même émotion mesurée en relation avec deux « participants », qui ont des p_0 et p_{100} différents.

**Figure 3.3 — Exemples de variations de l'indice en fonction des points p0 et p100
(indice faible à gauche; indice fort à droite)**

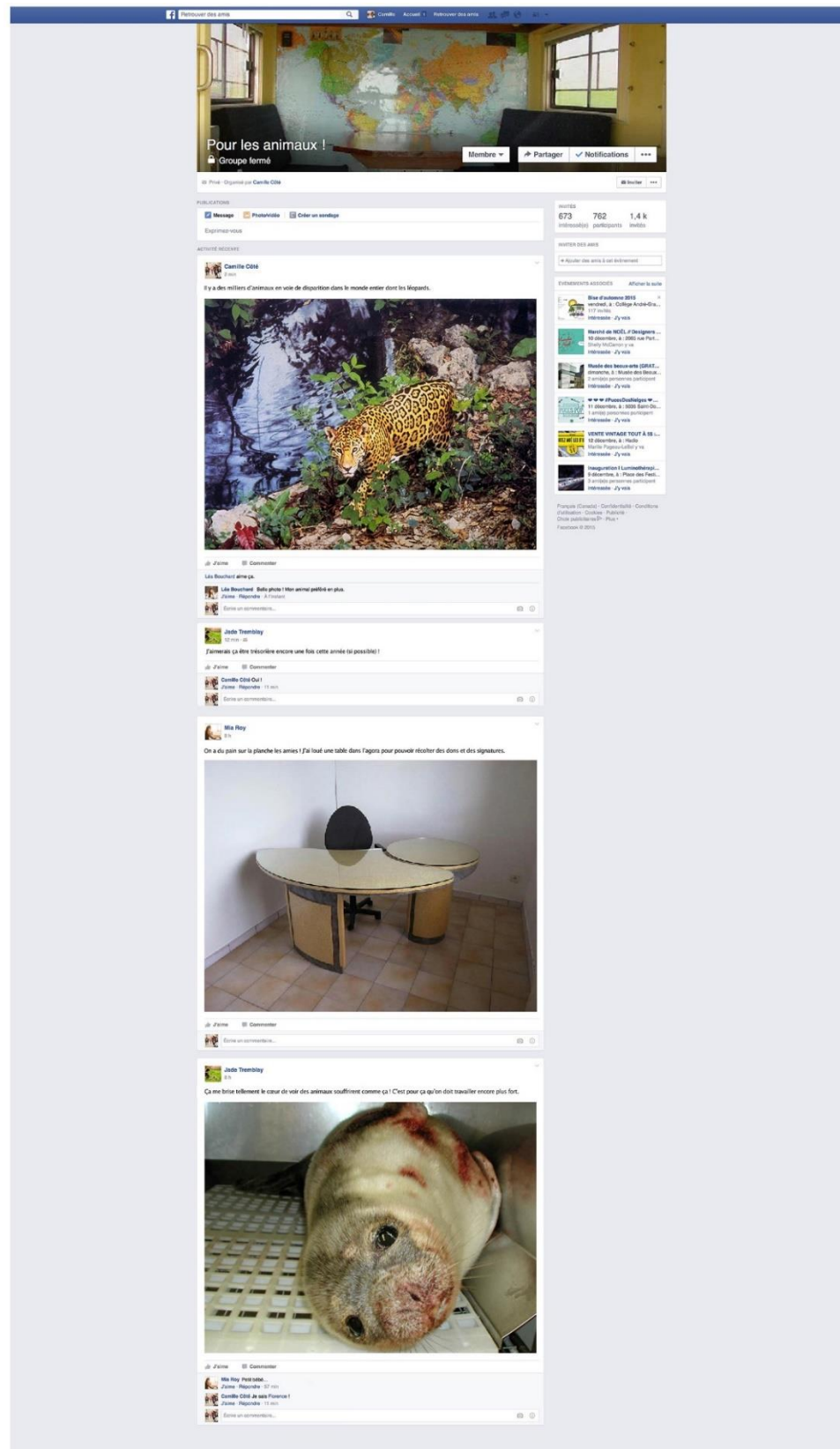


3.1.3 Présentation de l'étude originale de l'indice

Afin de contextualiser les données initiales utilisées pour cette étude, décrivons le protocole qui a permis d'obtenir ces données. L'étude initiale avait comme but de tenter de comprendre, du point de vue des éléments ergonomiques d'une page web et d'une page Instagram, ce qui différencie les pages générant une forte valeur de l'indice de celles qui génèrent une valeur faible (Léger et al., 2018). Le design expérimental de l'étude se divise en cinq parties. La première étape consiste en une calibration émotionnelle de chaque participant. Dans le but de générer une valeur de l'indice qui reflète la volatilité émotionnelle de chaque personne de manière unique, les participants de l'étude ont été exposés à douze images des bases de données IAPS et GAPED afin de mesurer le spectre d'émotions attendu dans une étude sur des marques et des pages internet corporatives. Les images étaient présentées en triades que le participant regardait pendant 45 secondes. Ces pages avaient l'apparence d'un fil d'actualité Facebook. La figure 3.4 illustre un exemple d'une page utilisé pendant cette phase. Cela étant dit, un participant pouvait regarder une image de la page sans avoir de restriction de temps (ex. 15 secondes par image de la triade). L'annexe 1 présente toutes les images de calibration utilisées avec leur banque d'images d'origine, leur code pour la banque d'images et les valeurs théoriques de la valence et de l'activation. L'annexe 2 présente toutes les triades d'images présentées dans l'étude originale. Par la suite, les participants devaient donner leur émotion selon le curseur affectif, permettant d'obtenir l'émotion rapportée par le participant. La seconde

étape de l'étude était d'obtenir l'opinion des participants (n = 50) des 27 marques qui étaient utilisés pour le projet. Pour avoir cette information, les participants ont été exposés à chaque logo pendant cinq secondes afin de capturer les données physiologiques. Après la présentation du logo, les participants ont utilisé le curseur affectif afin de donner leur émotion de manière autodéclarée. De plus, des questions vérifiant leur connaissance et expérience avec la marque ont été posées. La troisième étape de l'étude était de faire la capture de l'expérience émotionnelle des participants envers la page d'accueil des sites internet des marques. Les participants étaient exposés à chaque page pour une période de vingt secondes. Par la suite, l'émotion rapportée par le participant était enregistrée avec le curseur affectif et la question « Feriez-vous confiance à cette compagnie si vous aviez besoin de vous procurer leurs produits et/ou services? » était posée. La quatrième étape de l'étude utilisait la même structure que la phase des pages d'accueil, mais le contenu présenté était celui des comptes Instagram des marques. En dernier lieu, la cinquième étape donnait la chance à tous les participants de se connecter à leur compte Instagram personnel et de naviguer de manière habituelle pour une période de trois minutes.

Figure 3.4 — Exemple d’une page de la phase de calibration



En somme, l'étude a permis d'obtenir une quantité significative de données, tant au niveau des données de calibration que celles pour l'étude à proprement dit, afin de pouvoir mesurer l'expérience émotionnelle des participants et de mettre en œuvre l'indice.

3.2 Présentation des données initiales, accès aux données et manipulations effectuées

Afin de pouvoir contextualiser les démarches faites lors de ce projet, il est nécessaire d'expliquer la structure des données initiales ainsi que le travail qui a été fait afin d'obtenir des données utilisables dans la perspective du projet actuel. Comme illustré dans la section précédente, les participants étaient exposés aux stimuli de la phase de calibration sans durée fixe (mais les stimuli étaient présentés en triades pour quarante-cinq secondes), aux stimuli des logos pour une période de cinq secondes et aux stimuli des phases de site web pour une période de vingt secondes. Par conséquent, l'entièreté des données physiologiques concernant les images de calibration a été enregistrée à un intervalle d'une moyenne mobile prise à chaque 200 millisecondes (ms) pour tous les participants et ce même type d'information a été capturé à une fréquence d'une seconde lors des phases d'expérimentation (logos et sites web). Une autre réalité du jeu de données initial est que certaines données pouvaient être manquantes pour une variété de raisons. Dans un premier cas, les participants ont été exposés à seulement 50 % des pages web étudiées dans le projet initial afin de maintenir la pleine concentration et attention des participants pendant l'expérimentation. Il est donc attendu que, pour chaque stimulus de la phase d'expérimentation, il y aura des participants qui n'ont pas de données. Un deuxième cas concerne les données manquantes. Pour une variété de raisons hors du contrôle des chercheurs (erreur avec les outils pour faire les captures de données, le regard d'un participant était hors de la zone d'analyse pour l'analyse faciale, un capteur se décolle, etc.), il est possible de ne pas avoir de données pour une ou toutes les mesures physiologiques pour un ou plusieurs signaux d'un participant. Finalement, seulement les données de l'étude originale de l'indice ont été retenues, car ce sont les seules données qui contiennent toutes les informations nécessaires pour analyser une phase de calibration et les phases d'expérimentation avec l'indice.

Une autre réalité de ce projet est le fait que l'accès à plusieurs types de données a demandé des démarches supplémentaires. Comme l'étude originale se concentrait uniquement sur la valence mesurée par FaceReader et l'activation mesurée par le signal tonique et phasique de l'EDA, il a été nécessaire de faire une nouvelle extraction de données afin de pouvoir séparer le signal phasique de l'EDA pour l'analyser individuellement et pour obtenir le signal pour le HRV. En parallèle, comme les valeurs autodéclarées n'ont pas été considérées lors de l'étude originale, il a été nécessaire d'extraire ces données des sources brutes et de faire les transformations nécessaires (principalement la transposition des données et la standardisation des données) afin qu'elles puissent être utilisables dans le contexte de la présente étude. Donc, plusieurs sources de données ont dû être consultées et de nouvelles extractions de données ont dû être exécutées pour obtenir toute l'information nécessaire pour ce projet de recherche.

Par rapport aux manipulations nécessaires pour utiliser les données initiales dans le contexte de cette nouvelle étude, plusieurs changements ont été faits. En premier lieu, lorsqu'il y avait des données manquantes non intentionnelles (c.-à-d. pas planifié selon le protocole d'expérimentation initial), les données concernant le participant et le stimulus affectés ont été retiré du jeu de données pour se concentrer uniquement sur les informations complètes, donc qui peuvent être analysées. Par la suite, afin de pouvoir compléter certaines analyses, il a été nécessaire de concaténer des données de plusieurs sources différentes dans un même jeu de données. Ceci implique tant la concaténation verticale (donc l'ajout de variables de différentes sources associées à une même observation) ainsi que la concaténation horizontale (l'ajout d'observations de différentes sources afin de pouvoir faire des comparaisons entre elles). De plus, certaines sources de données ne s'alignaient pas de manière identique en raison de structures de données différentes, donc une resynchronisation des données a été nécessaire afin de pouvoir comparer les données des différents signaux d'intérêt pour le projet. Finalement, certaines mesures comme les données obtenues à l'aide du curseur affectif ont seulement un point par stimulus (ex. l'émotion ressentie par le participant après le stimulus). Dans cette situation, il est essentiel d'agréger les données des autres signaux physiologiques afin de pouvoir faire des comparaisons et analyses équitables. La méthode d'agrégation dépendra de l'analyse et sera explicitée dans la prochaine section.

En somme, une fois que les données initiales ont été obtenues, analysées et leur structure comprise, il a été possible de faire les ajustements nécessaires afin de réaliser cette nouvelle étude visant à approfondir les connaissances sur l'indice.

3.3 Présentation des analyses effectuées

Dans le contexte des données initiales et des manipulations effectuées, le présent projet utilise ces données afin d'analyser l'indice sous trois angles d'intérêt : tenter de trouver une manière de retirer la phase de calibration de la génération des valeurs de l'indice afin de simplifier le processus de collecte de données, comprendre la relation entre les données de l'étude qui étaient autodéclarées par les participants et les données des signaux physiologiques et vérifier que les données émotionnelles générées lors des différents stimuli de l'étude initiale démontre réellement des valeurs d'émotion variées. De plus, au début du projet, certaines pistes de recherche avaient été proposées. Ces pistes seront présentées en amont de la section afin de contextualiser les analyses qui ont été effectuées. Par la suite, cette section présentera la méthodologie utilisée afin de trouver une réponse à ces pistes d'études.

3.3.1 Pistes de recherche originales de l'étude

La première piste de recherche originalement envisagée était de pouvoir retirer la phase de calibration du processus pour obtenir une valeur de l'indice. Cette piste était la première analyse effectuée et sera présentée en plus grand détail dans la section suivante. La prochaine piste de recherche était de trouver une manière d'approfondir l'indice en ajoutant une nouvelle dimension : la charge cognitive, qui représente l'effort mental nécessaire pour effectuer une tâche (Paas et al., 1994) et peut être évaluée physiologiquement par l'électroencéphalographie (Mutlu-Bayraktar et al., 2020). L'idée derrière cette approche était de pouvoir incorporer la complexité de la tâche dans le calcul de l'indice. Deux pistes ont été proposées qui avaient pour but de modifier le calcul des distances pour évaluer la valeur de l'indice. La première idée était d'ajouter une plus grande importance aux émotions mesurées qui se retrouvent plus près du premier quadrant. Par exemple, si deux émotions mesurées sont à des distances euclidiennes identiques des points p_0 et p_{100} , mais que la première émotion se retrouve dans le premier

quadrant et que la seconde est dans le quatrième quadrant, le nouveau calcul de l'indice favoriserait la première émotion mesurée. La seconde piste était de tenir en compte des valeurs exceptionnelles. L'idée aurait été de donner un plus grand poids aux valeurs qui sont extrêmes, donc qui ne suivent pas la tendance. La distance de Mahalanobis aurait été incorporée dans le calcul de l'indice.

Notons :

$\vec{x}$ = Le vecteur où $\vec{x} = (x_1, x_2, \dots, x_N)^T$

$\vec{u}$ = Le vecteur des valeurs moyennes où $\vec{u} = (u_1, u_2, \dots, u_N)^T$

S = Une matrice de covariance

$D_M(\vec{x})$ = La distance de Mahalanobis pour le vecteur $\vec{x}$

En considérant ces éléments, la distance de Mahalanobis se calcul ainsi :

$$D_M(\vec{x}) = \sqrt{(\vec{x} - \vec{u})^T S^{-1} (\vec{x} - \vec{u})} . \quad (5)$$

La dernière idée proposée était de développer une stratégie de *benchmarking* pour établir une zone désirée pour un projet. S'il est possible d'identifier une zone cible pour un certain produit, en faisant une analyse comparative avec des produits concurrents, il pourrait être possible d'ajuster l'indice afin d'accorder une plus grande valeur aux émotions qui se trouvent dans la cible. En somme, plusieurs idées ont été proposées en amont afin de guider les recherches et analyses effectuées dans ce projet.

3.3.2 Comparaison entre l'indice original et des méthodes ne se fiant pas à la calibration

Comme mentionné au début de la section, la première grande piste de réflexion était de trouver une méthode pour simplifier la collecte de données d'une étude future en utilisant un indice qui se détermine exclusivement avec les données de la phase d'expérimentation d'une étude, écartant la nécessité de réaliser une phase de calibration des données. Un corolaire de cet objectif serait que ce nouvel indice pourrait être appliqué non seulement à de nouvelles études, mais pourrait être appliqué rétroactivement à des études qui n'avaient jamais eu de phase de calibration dans leur protocole. Deux méthodes

candidates ont été retenues afin de comparer avec la méthode originale pour calculer l'indice, soit d'utiliser les points extrêmes d'un participant et l'utilisation des 5^e et 95^e centiles d'une personne. Plus précisément, la méthode des points extrêmes utilise comme p0 le point avec la valence le plus faible et l'activation la plus haute de tous les points de la phase d'expérimentation d'un participant et le p100 correspond au point avec la valence la plus grande et l'activation la plus grande pendant cette même phase de l'étude. Pour la méthode avec les centiles, les variables sont triées de manière croissante, donc le minimum correspond au point avec la valeur la plus faible et le maximum correspond au point avec la valeur maximale. Par conséquent, le p0 correspond au 5^e centile de la valence et au 95^e centile de l'activation en analysant que les émotions à valence négative et le p100 correspond au 95^e centile de la valence et de l'activation en analysant les émotions à valence positive.

Le test pour vérifier l'utilité de ces mesures alternatives sera un test sur le coefficient de corrélation de Pearson entre la méthode originale de l'indice et les deux méthodes alternatives. Les hypothèses des tests sont les suivantes :

Pour la corrélation entre la méthode originale et la méthode des extrêmes :

$$H_0 : \rho_1 = 0$$

$$H_1 : \rho_1 \neq 0$$

où ρ_1 correspond au coefficient de corrélation entre la méthode originale pour calculer l'indice et la méthode des points extrêmes.

Pour la corrélation entre la méthode originale et la méthode des centiles :

$$H_0 : \rho_2 = 0$$

$$H_1 : \rho_2 \neq 0$$

Où ρ_2 correspond au coefficient de corrélation entre la méthode originale pour calculer l'indice et la méthode des centiles.

Concernant la granularité des données, comme il y a des données expérimentales à toutes les secondes et il est possible de calculer une valeur de l'indice pour tout moment où il existe une valeur de la valence et de l'activation, la granularité restera la même que celle des données initiales. En supplément à cette analyse, et pour simplifier la visualisation, une comparaison des méthodes pour une granularité d'une valeur par stimuli par participant sera également faite.

Il est à noter que, même si ce test est effectué, les données pour chaque méthode étudiée (méthode originale, méthode des extrêmes et méthode des centiles) ne sont pas indépendantes entre-elles parce qu'elles reposent sur les données des mêmes participants à l'étude. Dans ce cas, ce n'est pas le résultat du test d'hypothèse qui est l'élément à évaluer, mais la valeur du coefficient de corrélation de Pearson qui peut donner une indication sur la force et la direction de la relation. Si cette analyse primaire présente des résultats intéressants, des analyses plus robustes et valides, considérant la structure des données, pourraient être envisagées.

Bref, il sera possible de voir s'il existe une possibilité de retirer la phase de calibration de l'indice tout en assurant que cette mesure garde sa portée originale.

3.3.3 Analyse des signaux physiologiques et autodéclarés

La seconde grande piste d'analyse visait de mieux comprendre les liens entre les signaux physiologiques utilisés dans l'étude initiale, les données autodéclarées par le curseur affectif et les émotions qui devraient être générées par les images utilisées lors de la phase de calibration. Afin de comprendre les relations entre ces différentes données, deux analyses ont été effectuées : la première est de faire une analyse des corrélations entre les méthodes par un test sur le coefficient de corrélation de Pearson et la deuxième est une analyse des distances euclidienne entre les émotions obtenues lors de l'étude initiale et des émotions attendues selon les banques de données sources des images de calibration. La structure des hypothèses pour le test sur le coefficient de corrélation est la suivante :

$$H0 : \rho_{ij} = 0$$

$$H1 : \rho_{ij} \neq 0$$

Où ρ_{ij} correspond au coefficient de corrélation entre les données de la méthode i et les données de la méthode j . Les données utilisées pour cette analyse sont agrégées afin d'avoir une mesure par combinaison d'image et de participant parce que les données autodéclarées ont seulement une donnée disponible pour la combinaison d'intérêt, donc les autres mesures sont agrégées afin de faire une comparaison adéquate. Il faut encore noter que, en raison du lien de dépendance entre les données, le but de cette expérimentation est de faire une analyse du coefficient de corrélation afin de mieux comprendre la relation générale entre les méthodes.

En ce qui concernant l'analyse des distances euclidiennes, les distances se calculent en fonction du point de valence et d'activation moyen pour tous les participants en fonction de la même image en stimuli avec le point de la valence et de l'activation attendue selon la banque de données source. Donc, voici la structure des calculs de distance :

Notons :

V_{em} = La valence selon le signal m généré par le stimulus

A_{em} = L'activation selon le signal m généré par le stimulus

V_t = La valence théorique du stimulus

A_t = L'activation théorique du stimulus

d_m = Distance euclidienne de l'émotion générée par rapport à l'émotion théorique

En considérant ces éléments, la distance euclidienne est la suivante :

$$d_m = \sqrt{(V_{em} - V_t)^2 + (A_{em} - A_t)^2} . \quad (6)$$

Comme la valeur moyenne des émotions est utilisée, il y aura seulement un point calculé par image de calibration par participant.

Bref, ces analyses permettront de mieux comprendre les relations entre les différentes mesures d'intérêt et les émotions attendues selon la phase de calibration.

3.3.4 Analyse des variances des émotions générées lors de la phase de calibration

La dernière grande piste d'analyse était de cerner si les signaux physiologiques utilisés pour décrire les émotions avaient généré des données, lors de l'étude initiale, qui pouvaient cerner des différences significatives entre les images présentées lors de la phase de calibration. De plus, les données du curseur affectif concernant les émotions vécues et décrites par les participants ont également été analysées lors de cette partie de l'étude. Le test effectué était une analyse des variances (ANOVA) des valeurs générées par les images de calibration, limité aux données d'un type de mesure par ANOVA. En second temps, si l'ANOVA se révèle significative, une analyse des données deux à deux sera faite pour identifier les images de calibration qui se distinguent significativement des autres. Avant de présenter les hypothèses, trois éléments doivent être clarifiés :

1. Afin de sélectionner le type d'ANOVA le plus approprié (assumant/n'assumant pas l'égalité des variances entre les groupes étudiés), un test de Levene sur l'homogénéité des variances a été effectué. De manière générale, les hypothèses du test de Levene sont les suivantes :

H0 : Les variances sont homogènes

H1 : Les variances ne sont pas homogènes

2. Comme les données d'un même participant sont utilisées dans les tests, il est nécessaire de tenir en compte le fait que ces données seront corrélées entre-elles. Dans ce cas, un effet aléatoire sur l'ordonnée à l'origine a été ajouté dans les tests de l'ANOVA au niveau de chaque participant.

3. Lorsque des tests deux à deux sont effectués, afin de contrôler la probabilité globale de commettre une erreur de type I, une correction de Tukey-Kramer a été effectuée sur les seuils expérimentaux.

Concernant les hypothèses des ANOVAs effectuées, la structure générale est la suivante :

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6 = \mu_7 = \mu_8 = \mu_9 = \mu_{10} = \mu_{11} = \mu_{12}$$

H1 : au moins 2 moyennes sont différentes

où μ_i correspond à la moyenne d'une mesure de l'émotion pour la $i^{\text{ième}}$ image de la phase de calibration.

Étant donné que la granularité des données diffère dans les analyses (comme rappel, il y a seulement une mesure de valence et d'activation par image et par participant pour la méthode du curseur affectif et une donnée par 200 ms pour les méthodes physiologiques), les ANOVAs sont effectuées avec le plus grand grain, soit une mesure par image et par participant. De plus, comme les participants pouvaient regarder les stimuli sans durée fixe, il y a certaines données de très courte durée. Afin de contrôler pour des participants qui n'auraient pas regardé pour un moment suffisamment long, cette analyse a été refaite en ne gardant que des observations où le participant aurait regardé l'image pour au moins 1 seconde. Cette valeur a été choisie parce qu'elle permet d'écarter les observations des participants qui n'ont presque pas regardé l'image et permet d'assurer une quantité de données nécessaires pour effectuer l'ANOVA.

En somme, ces tests permettront d'établir si les mesures physiologiques ont généré des mesures significativement différentes les unes des autres ou si les mesures obtenues s'apparentent plus à du bruit dans les données.

3.4 Conclusion

Les sections précédentes de ce chapitre ont permis de comprendre l'origine de l'indice, le contexte et la méthodologie de l'étude initiale de l'indice, les données générées par cette étude ainsi que les manipulations nécessaires afin de pouvoir les utiliser dans l'étude actuelle et les grandes analyses de ce projet. La première grande analyse visait à tenter d'élargir la portée de l'indice en cherchant une méthode pour simplifier son calcul. La deuxième grande analyse avait pour but de tenter de mieux comprendre les signaux physiologiques et données autodéclarées qui ont été générés lors de l'étude et leurs relations avec les émotions attendues pour les stimuli de la phase de calibration. La troisième analyse voulait démontrer que les données générées dans l'étude pouvaient illustrer des différences significatives dans les émotions vécues par les personnes étant exposées aux stimuli de la phase de calibration.

Chapitre 4

Analyse

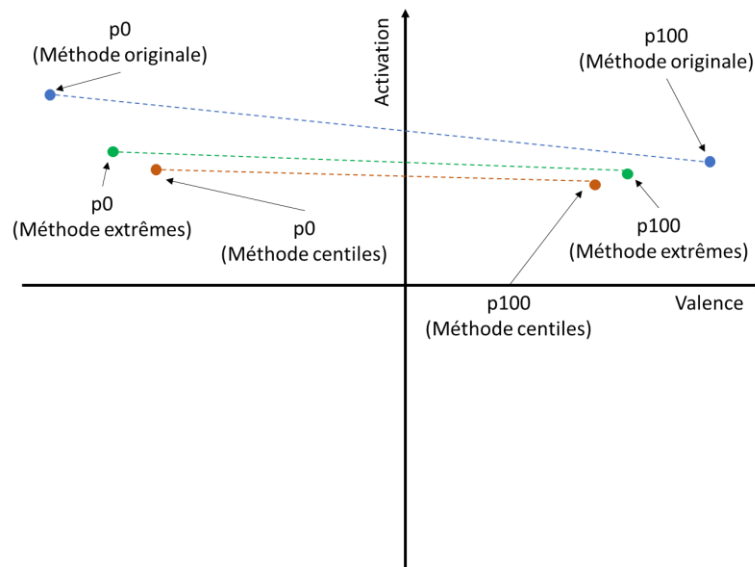
Résumé

Ce chapitre vise à présenter les résultats des analyses effectuées lors du projet de recherche. Les faits saillants des analyses seront présentés afin d'assurer une compréhension complète des résultats obtenus. La première analyse décrite concerne la comparaison entre la méthode originale pour calculer l'indice et les méthodes candidates qui n'utilisent pas la phase de calibration pour calculer l'indice et est analysée à deux niveaux de granularité, soit à un niveau plus granulaire où il y a une valeur de l'indice calculé à chaque seconde par stimulus et par participant et à un niveau moins granulaire où c'est la valeur moyenne de l'indice par stimulus et par participant qui est considérée. La deuxième analyse est à l'égard du lien entre les mesures autodéclarées et les mesures physiologiques des composantes de l'émotion, tant d'un point de vue de la corrélation entre les différents types de mesure que d'une perspective descriptive des données de l'étude originale. La dernière analyse consiste à effectuer une analyse des variances afin d'identifier si les différentes méthodes permettent d'identifier des différences émotionnelles statistiquement significatives entre les images de la phase de calibration.

4.1 Comparaison entre l'indice original et des méthodes ne se fiant pas à la calibration

En guise de rappel, la première grande piste de réflexion de ce projet était d'explorer la possibilité de simplifier le processus d'utilisation de l'indice en éliminant le processus de calibration des émotions. Comme mentionné au chapitre précédent, deux méthodes candidates ont été retenues pour les analyses, soit l'utilisation des points extrêmes (les émotions les plus intenses mesurées lors de la phase d'expérimentation) ou l'utilisation de centiles (le 95^e centile de la valence et de l'activation pour l'émotion positive et le 5^e centile de la valence et le 95^e centile de l'activation pour l'émotion négative). La figure 4.1 illustre un exemple hypothétique de la relation entre les trois méthodes pour calculer l'indice (la méthode originale, la méthode des extrêmes et la méthode des centiles).

Figure 4.1 — Illustration des points p_0 et p_{100} pour les trois méthodes de calcul de l'indice étudiées



Le test retenu afin d'évaluer s'il existe un lien entre la méthode originale pour calculer l'indice et les candidats pour une méthode simplifiée était le test du coefficient de corrélation de Pearson. La corrélation des indices a été évaluée à l'aide de données à deux niveaux de granularité différents, soit en analysant toutes les données (une observation par seconde, par stimulus et par participant) ou en agrégeant les données à une valeur moyenne de l'indice par stimulus et par participant.

4.1.1 Analyse des données de l'indice agrégées par seconde, par stimulus et par participant

Les premières analyses de la corrélation entre les méthodes de calcul de l'indice étaient effectuées avec les données avec le grain le plus précis possible. Dans ce cas, c'était de calculer la valeur de l'indice à chaque moment que les informations physiologiques étaient enregistrées lors de la phase d'expérimentation. Ceci veut dire qu'il y a une donnée chaque seconde qu'un participant était exposé à un stimulus (dans le cas de l'étude originale, ceci représente des logos, images de pages web et images de comptes Instagram). Le tableau 4.1 illustre les statistiques descriptives pour la valeur de l'indice pour chaque méthode. Comme rappel, les valeurs de l'indice se situent entre 0 et 1, inclusivement. Il est possible de remarquer que, pour la médiane, la moyenne et les

mesures des extrêmes (minimum et maximum), la méthode candidate utilisant les centiles se rapproche plus des statistiques de la méthode de calcul originale de l'indice. À l'inverse, les statistiques pour les premier et troisième quartiles indiquent que la méthode candidate utilisant les points extrêmes est plus proche des valeurs de l'indice original. Il est à noter que le N de 34541 représente les données des 27 logos visionnés pendant 5 secondes avec l'émotion mesurée au début de la période de visionnement (donc 6 points de données) et des 26 pages web et Instagram visionnées par chaque participant avec 21 points de données (20 secondes de stimulation avec l'état initial) pour les 50 participants à l'étude, tout en retirant les données manquantes.

Tableau 4.1 — Statistiques descriptives des indices avec les données agrégées à la seconde

Nombre d'observations (N = 34541)

Méthode	Min	1er Quartile	Médiane	Moyenne	Écart-Type	3e Quartile	Max
Indice original	0,0082	0,4165	0,4970	0,4955	0,1444	0,5654	0,9991
Méthode des extrêmes	0	0,3328	0,4445	0,4549	0,1647	0,5662	1,0000
Méthode des centiles	0,0006	0,3217	0,4812	0,4805	0,1995	0,6259	0,9979

En effectuant le test du coefficient de corrélation de Pearson, il est possible d'identifier s'il existe une corrélation entre les données des méthodes. Le tableau 4.2 illustre les résultats.

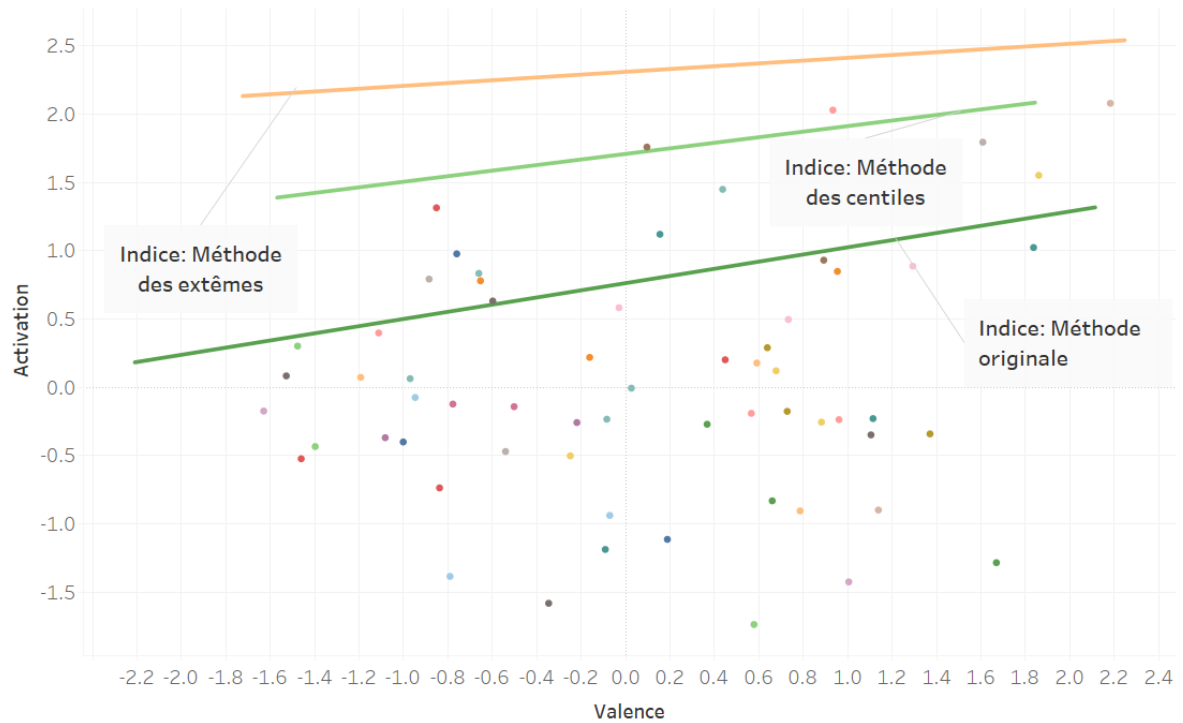
Tableau 4.2 — Résultats du test du coefficient de corrélation entre les méthodes candidates pour calculer l'indice et la méthode originale avec les données agrégées à la seconde

Nombre d'observations (N = 34541)

	Indice original et la méthode des extrêmes	Indice original et la méthode des centiles
Coefficient de corrélation de Pearson (ρ)	0,4251	0,5152
P-value	<0,0001	<0,0001

Avant de discuter des résultats, il est important d'expliciter le fait que les p-values obtenues ne sont pas fiables étant donné que le test de corrélation de Pearson suppose l'indépendance des échantillons. Cela étant dit, l'analyse des coefficients, afin de cerner la présence d'une forte corrélation entre les méthodes, demeure pertinente. La relation entre les méthodes, dans les deux cas, est positive. Ceci veut dire que les valeurs de l'indice évoluent dans la même direction (la méthode candidate tend à être positive lorsque la valeur de l'indice originale est positive et vice-versa). Cela étant dit, la force de la corrélation entre les méthodes demeure relativement faible. Dans le cas de la relation entre l'indice original et la méthode candidate utilisant les points extrêmes, la valeur du coefficient de corrélation de Pearson est de 0,43 et, dans le cas de la relation entre la méthode originale et la méthode candidate utilisant les centiles, elle est de 0,52. Afin de compléter l'analyse, la corrélation entre les deux méthodes candidates a été étudiée et le coefficient de Pearson est de 0,91. Ce résultat est attendu, car le calcul de l'indice pour les deux méthodes candidates demeure semblable (l'utilisation du minimum/maximum et l'utilisation des 5^e et 95^e centiles respectivement). La figure 4.2 permet de visualiser les différences des p0 et p100 pour les trois méthodes analysées pour un participant de l'étude.

Figure 4.2 — p0 et p100 calculés selon les différentes méthodes de calcul de l'indice pour un participant (utilisant les données agrégées à la seconde)



Bref, les valeurs obtenues sont trop petites afin de conclure que les méthodes candidates seraient des substitues pour la technique originale.

4.1.2 Analyse des données de l'indice agrégées par stimulus et par participant

Par la suite, l'analyse de la corrélation entre les méthodes a été refaite, mais cette fois-ci avec des données d'une granularité différente. Au lieu d'utiliser toutes les données possibles (donc à chaque seconde pour chaque stimulus pour chaque participant), les informations ont été agrégées par stimulus et par participant afin d'être plus représentatives de l'utilisation des données de l'indice (présenté par participant et stimulus ou même simplement par stimulus pour simplifier l'information à présenter). Il y a donc seulement une donnée pour le stimulus visionné par un participant, représentant la moyenne des valeurs de l'émotion mesurées physiologiquement. Le tableau 4.3 présente les statistiques descriptives pour la valeur de l'indice selon les trois méthodes étudiées. De manière semblable aux données plus granulaires, la médiane et la moyenne de la méthode candidate utilisant la méthode des centiles sont plus proches des valeurs de

l'indice original. À l'opposé, le minimum et le maximum de la méthode candidate utilisant les points extrêmes sont plus proches de la valeur de l'indice originale, ce qui diverge des données utilisant l'information moins agrégée. Il est à noter que le N de 2531 représente les données des 27 logos visionnés et des 26 pages web et Instagram visionnées par chaque participant pour les 50 participants à l'étude, tout en retirant les données manquantes.

Tableau 4.3 — Statistiques descriptives des indices avec les données agrégées au stimulus
Nombre d'observations (N = 2531)

Méthode	Min	1er Quartile	Médiane	Moyenne	Écart-Type	3e Quartile	Max
Indice original	0,0351	0,4268	0,5043	0,5055	0,1434	0,5794	0,9630
Méthode des extrêmes	0,0059	0,3578	0,4618	0,4692	0,1586	0,5719	0,9881
Méthode des centiles	0,0027	0,3579	0,4989	0,4998	0,1942	0,6346	0,9978

Malgré ces légères différences, le tableau 4.4 illustre les résultats du test sur le coefficient de corrélation des méthodes candidates avec l'indice original pour les données agrégées au stimulus et les différences sont minimales comparées à la version moins agrégée. Les faits saillants de l'analyse précédente restent pareils : la corrélation est positive (si la valeur de l'indice original est positive, la valeur de la méthode candidate aura tendance à être positive et vice-versa) et la force de la corrélation est relativement faible, ce qui correspond à un coefficient de corrélation de 0,42 pour la corrélation entre l'indice original et la méthode candidate utilisant les points extrêmes (une diminution de la corrélation d'environ 0,01 vis-à-vis l'analyse précédente) et un coefficient de 0,52 pour la corrélation entre l'indice original et la méthode candidate utilisant les centiles (reste approximativement stable vis-à-vis l'analyse précédente). Comme pour l'analyse plus granulaire, les p-values ne sont valides étant donnée la nature non indépendante des

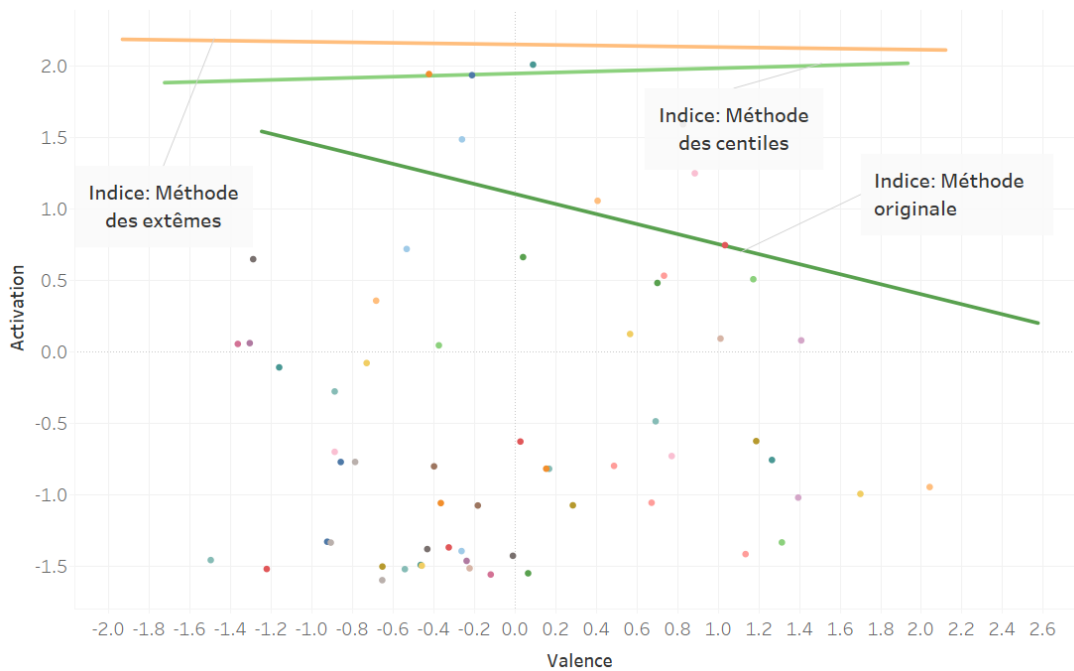
échantillons. De plus, la corrélation entre les deux méthodes candidates demeure forte lorsque les données agrégées au stimulus sont étudiées. Le coefficient de corrélation est de 0,89. Encore, ce résultat est attendu en raison du lien qui existe dans le calcul de l'indice pour les deux méthodes. La figure 4.3 illustre les p_0 et p_{100} pour les trois méthodes analysées pour un participant de l'étude.

Tableau 4.4 — Résultats du test du coefficient de corrélation entre les méthodes candidates pour calculer l'indice et la méthode originale pour les données agrégées au stimulus

Nombre d'observations (N = 2531)

	Indice original et la méthode des extrêmes	Indice original et la méthode des centiles
Coefficient de corrélation de Pearson (ρ)	0,4172	0,5158
P-value	<0,0001	<0,0001

Figure 4.3 — p0 et p100 calculés selon les différentes méthodes de calcul de l'indice pour un participant (utilisant les données agrégées au stimulus)



En somme, l'analyse de la corrélation entre la méthode de calcul originale de l'indice et de deux méthodes candidates qui n'utilisent pas les données de la phase de calibration révèle que, malgré la présence d'une relation positive, elle demeure trop faible pour que les méthodes candidates substituent la méthode initiale tout en s'appuyant sur les apprentissages faits lors de l'étude initiale de l'indice d'expérience émotionnelle.

4.2 Analyse des signaux physiologiques et autodéclarés

À la suite de la première grande piste d'analyse, une exploration plus en profondeur a été effectuée sur les différents signaux physiologiques qui existent pour mesurer l'activation et le signal utilisé pour mesurer la valence. Comme mentionné au chapitre précédent, le but de cette analyse était de mieux comprendre le lien entre ces signaux et les données autodéclarées par le moyen du curseur affectif. Afin de répondre à cette question, un test sur le coefficient de corrélation de Pearson a été effectué pour voir le lien entre les signaux physiologiques et les valeurs autodéclarées et, dans un second temps, une analyse descriptive des signaux a été effectuée afin de mieux comprendre la relation entre les

mesures des émotions, tant physiologiques qu'autodéclarées, obtenues lors de l'étude originale et les valeurs attendues selon les différentes banques d'images utilisées.

Les données utilisées pour cette partie de l'analyse viennent de la phase de calibration de l'étude originale parce que ce sont ces données qui peuvent être comparées avec des valeurs théoriques attendues qui sont fiables. De plus, il a été mentionné au chapitre 3 que les données sont agrégées à l'image visionnée par chaque participant, car les valeurs autodéclarées sont mesurées seulement une fois, soit après que l'image a été vue par le participant.

4.2.1 Analyse de la corrélation entre les signaux physiologiques et autodéclarés

La première analyse effectuée, dans le but de mieux comprendre le lien entre les mesures de l'émotion obtenues par des moyens physiologiques et celles obtenues par le curseur affectif, était une analyse de la corrélation entre les différents types de signaux. Évidemment, il est important de comparer uniquement des mesures qui rapportent la même composante de l'émotion, donc les deux signaux de la valence seront comparés, soit la valence mesurée par une analyse faciale et la valence mesurée par le curseur affectif. En ce qui concerne l'activation, quatre signaux sont analysés. Le premier est par le moyen de l'activité électrodermale (EDA) utilisant une combinaison des signaux phasiques et toniques. Le second est en analysant uniquement l'EDA phasique d'un individu. Le troisième est l'utilisation de la variance de la fréquence cardiaque (HRV). Le quatrième signal est l'activation mesurée par le curseur affectif.

Le tableau 4.5 contient les valeurs des statistiques descriptives pour les signaux de la valence et le tableau 4.6 contient les valeurs des statistiques descriptives pour les signaux de l'activation. Il est important de noter que les valeurs différentes pour le nombre d'observations découlent du fait que certaines informations sont manquantes soit lors de la phase d'expérimentation (ex. un participant oublie d'enregistrer une valeur dans le curseur affectif) ou soit lors de l'extraction des données (ex. l'outil utilisé pour extraire les données n'a pu extraire une valeur pour un certain signal d'un participant).

Tableau 4.5 — Statistiques descriptives des signaux mesurant la valence

Signal	N	Min	1er Quartile	Médiane	Moyenne	Écart-type	3e Quartile	Max
Valence par analyse faciale	500	-1.3486	-0.9780	0.2450	0.0667	1.0152	0.9573	3.6181
Valence par curseur affectif	546	-1.3404	-0.8698	-0.1457	0.0000	1.0000	0.7955	2.2799

En analysant le tableau 4.5, il est possible de remarquer que les valeurs de la moyenne sont très proches de 0 et les valeurs de l'écart-type sont proches de 1. La raison pour ceci est parce que les données ont été standardisées par participant, pour obtenir une moyenne de 0 et un écart-type de 1 pour les données de chaque participant, lorsque les données étaient au niveau le plus granulaire (ex. pour le curseur affectif, pour chaque image de calibration par participant et pour le l'analyse faciale avec FaceReader, pour chaque seconde qu'un participant visionnait une image de calibration). En agrégeant les données, la moyenne et l'écart-type se sont modifiés légèrement. Également, les données de la valence mesurées avec l'analyse faciale ont une étendue plus faible en raison d'un minimum plus grand et d'un maximum plus petit que les données obtenues par l'analyse faciale. L'étendue pour la valence avec le curseur affectif est d'environ 3,62 et l'étendue pour la valence mesurée avec l'analyse faciale est de 4,97.

Tableau 4.6 — Statistiques descriptives des signaux mesurant l'activation

Signal	N	Min	1er Quartile	Médiane	Moyenne	Écart-type	3e Quartile	Max
EDA (phasique et tonique)	496	-1.2832	-0.6634	0.0159	0.0606	0.9791	0.5344	3.8375
EDA (phasique seulement)	423	-0.3544	-0.3272	-0.2416	0.2052	1.5740	0.0632	17.1150
HRV	455	-2.1233	-0.5295	-0.1114	0.0198	0.9594	0.3609	8.8953
Activation par curseur affectif	568	-1.6212	-0.8570	0.2394	0.0000	1.0000	0.7710	1.7014

Concernant le tableau 4.6, le même processus de standardisation a été effectué que décrit précédemment, ceci explique le fait que les moyennes sont proches de 0 et les écarts-types sont proches de 1. En analysant les étendues, il est possible de remarquer que le curseur affectif demeure une méthode avec une petite valeur (3,32). Par la suite, la méthode physiologique plus traditionnelle de l'EDA combinant le signal phasique et tonique a la deuxième plus petite étendue avec une valeur de 5,12, la mesure de l'activation utilisant le HRV présente une étendue plus grande (11,02) et la version utilisant uniquement le signal phasique de l'EDA a la plus grande étendue avec 17,47. Les différences dans les étendues semblent être en raison d'une combinaison de la valeur minimum et de la valeur maximum observées. Par exemple, la mesure du curseur affectif pour évaluer l'activation a le deuxième plus grand minimum observé et le plus petit maximum tandis que le signal utilisant uniquement la composante phasique de l'EDA a le plus petit minimum observé, mais le plus grand maximum (avec 17,12). Comme la standardisation a été effectuée au niveau du jeu de données au complet, ces valeurs extrêmes sont le signe que les participants à l'étude ont eu des variabilités émotionnelles différentes.

Avec l'analyse descriptive effectuée, il est possible d'évaluer le coefficient de corrélation pour les différents signaux mesurant la valence afin de déterminer s'il existe une corrélation dans les données. Le tableau 4.7 présente les résultats du test.

Tableau 4.7 — Résultats du test du coefficient de corrélation entre les signaux mesurant la valence

	Analyse faciale et le curseur affectif
Coefficient de corrélation de Pearson (ρ)	0.01928
N	462
P-value	0.6794

Comme avec la première analyse, comme les échantillons ne sont pas indépendants, la p-value n'est pas pertinente. Cependant, l'important, dans ce cas, est d'analyser la valeur du coefficient de corrélation de Pearson, qui illustre une valeur de presque 0. Ceci permet de dire que les chances qu'il existe véritablement une corrélation entre l'utilisation de l'analyse faciale pour mesurer la valence et l'utilisation du curseur affectif pour mesurer cette même composante de l'émotion sont presque nulles.

Par la suite, afin d'évaluer la corrélation entre les différents signaux pouvant mesurer l'activation, le test du coefficient de corrélation a été effectué pour ces signaux afin d'explorer le potentiel d'une relation entre les mesures. Le tableau 4.8 présente les résultats pour la corrélation entre les données autodéclarées par le curseur affectif et les trois signaux physiologiques.

Tableau 4.8 — Résultats du test du coefficient de corrélation entre les signaux mesurant l'activation

	EDA (phasique et tonique) et le curseur affectif	EDA (phasique) et le curseur affectif	HRV et le curseur affectif	EDA (phasique et tonique) et EDA (phasique)	EDA (phasique et tonique) et HRV	EDA (phasique) et HRV
ρ	-0.0296	0.0430	0.0026	0.1195	0.0134	-0.0368
N	479	407	439	423	450	423
P- value	0.5177	0.3873	0.9562	0.0139	0.7777	0.45

Les résultats de cette analyse exploratoire révèlent que pour la relation entre le curseur affectif mesurant l'activation et les trois mesures physiologiques de l'activation, les valeurs sont très petites, proches de 0, ce qui permet de comprendre qu'il y a probablement peu de lien entre la méthode autodéclarée et les méthodes physiologiques pour mesurer l'activation. La seule exception est la relation entre l'EDA combinant les signaux phasiques et toniques et l'EDA utilisant uniquement le signal tonique, mais, comme mentionné précédemment, les résultats de cette analyse sont seulement pour étudier la force de la relation, et non la significativité.

En regardant les résultats des analyses pour les signaux mesurant la valence et l'activation, la conclusion reste la même, ceci laisse penser qu'il n'y a pas de lien entre les données obtenues par le curseur affectif et les données obtenues par l'analyse des signaux physiologiques. Bref, cette analyse laisse penser que le comportement des mesures autodéclarées est différent que celui des signaux physiologiques.

4.2.2 Analyse des statistiques descriptives de la distance entre les signaux du point théorique

À la suite de l'analyse sur la force de corrélation entre les signaux physiologiques et autodéclarés, une analyse plus descriptive a été effectuée. En utilisant les données de la phase de calibration, l'objectif de cette analyse est de comparer les distances euclidiennes entre les valeurs moyennes des émotions, pour chaque méthode de mesure de l'émotion,

et la valeur théorique des stimuli de la phase de calibration. Le calcul utilisé est décrit à la section 3.3.2 du chapitre 3. Comme l'exposition aux images de calibration devrait générer des valences et activations près des valeurs théoriques calculées par les créateurs des banques d'images (car ces banques ont été démontrées comme fiables pour générer des émotions particulières), les mesures de l'émotion, qu'elles soient physiologiques ou autodéclarées, devraient enregistrer des valeurs qui s'alignent à ces valeurs théoriques. Cette analyse aide à mieux comprendre si les différences entre les données autodéclarées et les données physiologiques ont un impact sur leur performance à bien estimer les valeurs théoriques des images présentées lors de la phase de calibration.

Comme décrit à la section 4.2.1 de ce chapitre, les données de la valence et de l'activation ont été standardisées à l'échelle du jeu de données afin d'obtenir une moyenne de 0 et un écart-type de 1. Avec cette clarification effectuée, le tableau 4.9 présente les statistiques descriptives des calculs des distances euclidiennes entre une méthode de calcul de l'émotion (valence et activation) et les valeurs théoriques de l'image de calibration qui devraient être générées chez un participant. Notez que le tableau complet présentant les valeurs émotionnelles moyennes pour chaque image et chaque méthode, ainsi que la valeur de la distance euclidienne est présenté à annexe 3.

Tableau 4.9 — Statistiques descriptives des distances entre les méthodes pour mesurer l’émotion et les valeurs théoriques

Méthode	N	Min	1er Quartile	Médiane	Moyenne	Écart-type	3e Quartile	Max
Analyse faciale + EDA (phasique et tonique)	495	0.0776	1.3050	2.0303	2.0477	0.9729	2.6773	5.7469
Analyse faciale + EDA (phasique seulement)	422	0.0709	1.3380	1.9339	2.1233	1.4738	2.5391	18.3064
Analyse faciale + HRV	454	0.1066	1.4003	1.9891	2.0479	0.9114	2.5892	6.9286
Données Autodéclarées	544	0.0292	0.8799	1.3608	1.4326	0.7393	1.8688	4.0191

En analysant les statistiques descriptives obtenues, il est possible de voir une tendance claire que les distances euclidiennes entre les données obtenues par le curseur affectif et les valeurs théoriques pour les images de calibration sont toujours les plus faibles des quatre méthodes évaluées. Cela permet de constater que, de manière générale, les émotions mesurées par le curseur affectif s’alignent mieux avec ce qui est attendu selon les banques d’images que les mesures de l’émotion physiologiques. De plus, comme l’écart-type pour la méthode avec les données autodéclarées est le plus faible, il est possible de conclure qu’il y a moins de variabilité dans les distances mesurées avec cette méthode, donc il devrait avoir une erreur plus prédictible en utilisant cette méthode. Pour les méthodes utilisant des données obtenues par signaux physiologiques, la méthode utilisant le HRV pour mesurer l’activation présente les pires résultats pour le minimum et le premier quartile, mais présente la deuxième plus petite valeur pour l’écart-type. Pour les autres statistiques, cette méthode est toujours la deuxième plus grande, donc elle se

démarque peu comme une mesure fiable. À l'inverse, malgré avoir les valeurs les plus faibles, pour les méthodes physiologiques, du minimum, de la médiane et du troisième quartile, la méthode utilisant uniquement le signal phasique de l'EDA pour mesurer l'activation présente la moyenne, l'écart-type et le maximum le plus élevé de toutes les méthodes. En d'autres mots, cette méthode estime l'émotion mesurée avec la plus grande variabilité du point théorique. En ce qui concerne la méthode utilisant la combinaison des signaux phasiques et toniques pour mesurer l'activation, malgré qu'elle ait les statistiques de la médiane et du troisième quartile les plus grandes, il est important de nuancer ce point parce que les valeurs observées sont quand même proches des valeurs pour les autres méthodes. La médiane est seulement plus grande de 0,04 de la deuxième plus grande valeur (l'activation mesurée par le HRV) et de 0,10 de la meilleure valeur observée, celle pour l'EDA phasique. Pour le troisième quartile, l'EDA phasique et tonique est plus grande de 0,09 de la valeur observée avec le HRV et de 0,14 de la valeur observée avec l'EDA uniquement phasique. L'utilisation de la combinaison des signaux de l'EDA phasiques et toniques pour mesurer l'activation présente les meilleurs résultats observés pour le premier quartile, la moyenne et le maximum. Bref, cette analyse démontre que, dans le contexte de cette étude, l'utilisation des données physiologiques pour mesurer la valence et l'activation chez un participant génère des valeurs qui s'alignent mal avec ce qui est attendu selon les valeurs théoriques émotionnelles pour les images utilisées lors de la phase de calibration.

En plus de se fier sur les statistiques descriptives pour analyser la situation, il est possible de représenter visuellement les distances pour mieux cerner les performances des différentes méthodes pour mesurer l'émotion des participants.

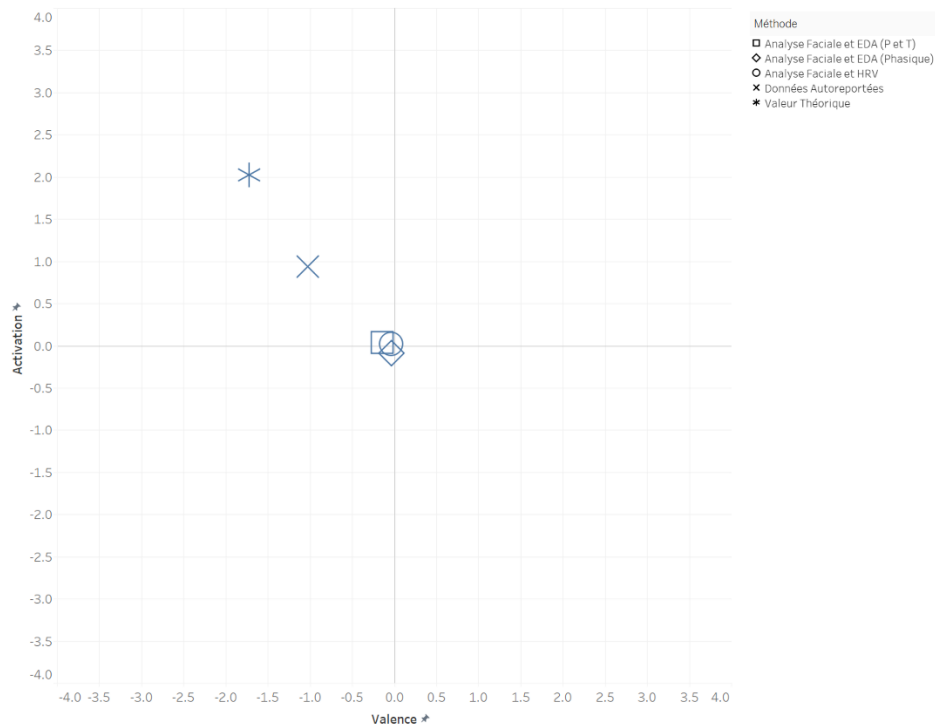
La figure 4.4 illustre l'image de calibration codée au nom «4_3» (représentant la 3^e image de la 4^e triade d'images de calibration) et la figure 4.5 illustre visuellement les distances euclidiennes moyennes des différentes méthodes pour mesurer l'émotion des participants avec la valeur théorique de la banque d'images IAPS.

Figure 4.4 — Image de calibration 4_3



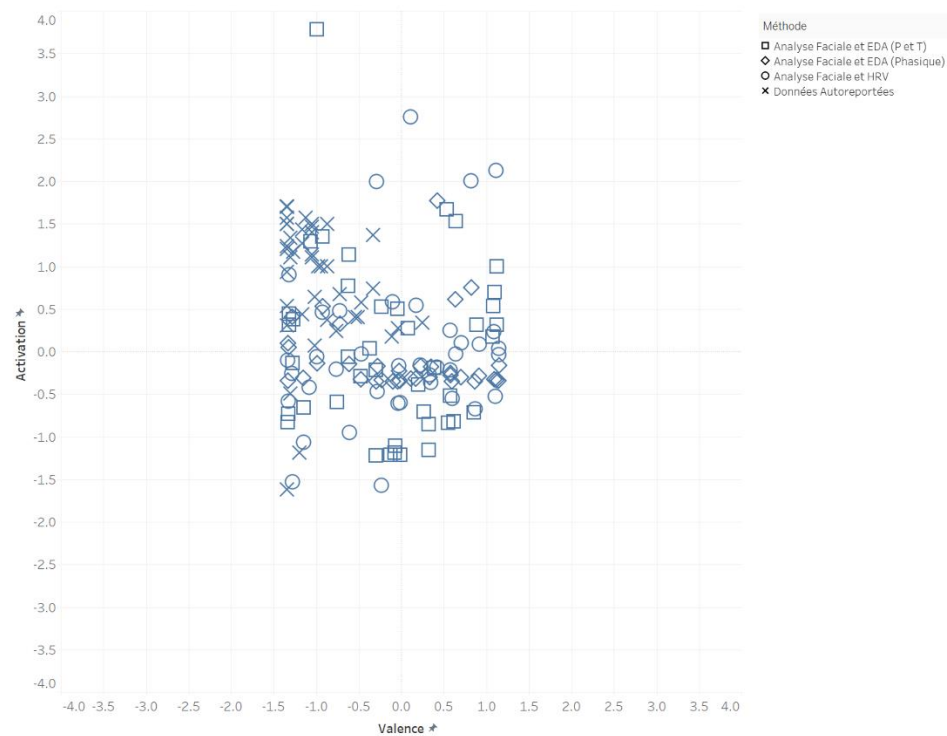
La figure 4.5, ci-dessous, illustre le fait que les données autodéclarées par le moyen du curseur affectif estiment la valeur théorique le mieux vis-à-vis les autres méthodes. La moyenne des émotions mesurées (valence de -1,02 et activation de 0,93) se situe dans le cadran représentant les émotions à valence négative et à activation élevée. À titre informatif, les distances euclidiennes mesurées sont de 1,30 pour le curseur affectif, 2,54 pour l'analyse faciale avec l'EDA phasique et tonique (le maximum observé pour cette méthode), 2,70 pour l'analyse faciale avec l'EDA phasique et 2,62 pour l'analyse faciale avec le HRV (le maximum observé pour cette méthode). L'émotion théorique pour cette image est une valence de -1,72 et une activation de 2,02, ce qui correspond à une image qui provoque une réaction très négative d'une grande intensité. Notez que les petites variations pour la valence avec les différentes méthodes autodéclarées sont causées par le fait qu'il y a des observations manquantes qui varient de méthode en méthode. Ceci s'applique pour tous les moyennes mesurées.

Figure 4.5 — Graphique illustrant l'émotion (valence et activation) moyenne mesurée pour chaque méthode en comparaison avec la valeur théorique pour l'image de calibration 4_3



Il est important de souligner que les valeurs extrêmes ne semblent pas être assez intenses et assez nombreuses pour exercer une influence importante sur les valeurs moyennes mesurées. La figure 4.6 illustre le nuage de points pour toutes les méthodes pour mesurer l'émotion pour chaque participant de l'étude.

Figure 4.6 — Graphique illustrant l'émotion (valence et activation) mesurée pour chaque participant par méthode pour l'image de calibration 4_3



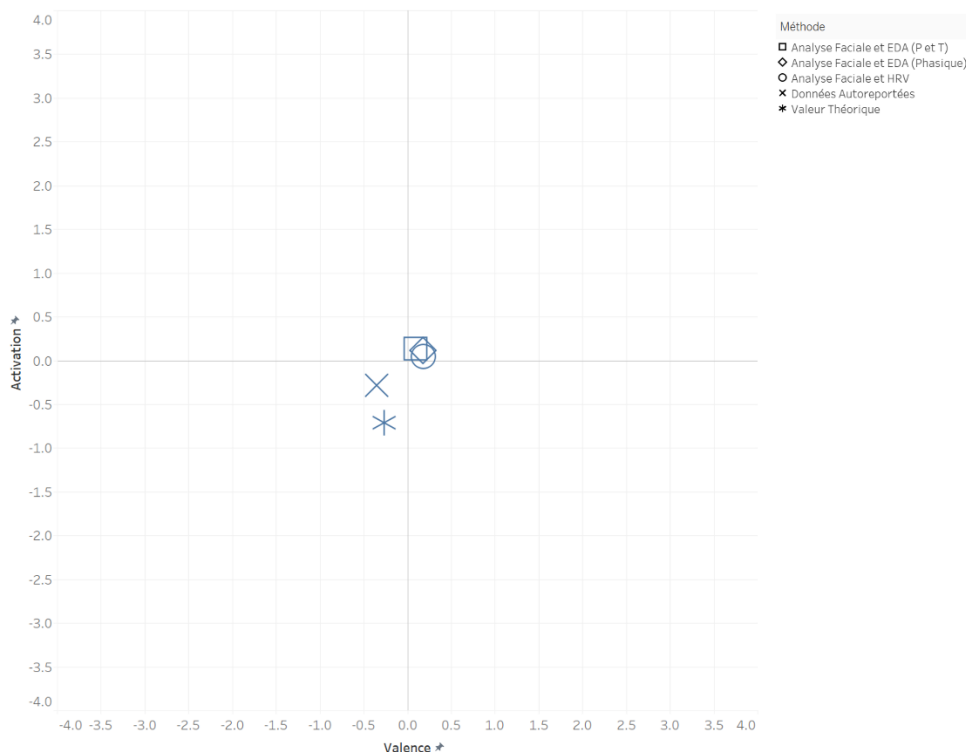
La même tendance s'observe également avec des images qui sont supposées générer des émotions de faible intensité, ne voir aucune émotion. La figure 4.7 illustre une telle image, soit l'image de calibration codée « 1_2 ».

Figure 4.7 — Image de calibration 1_2



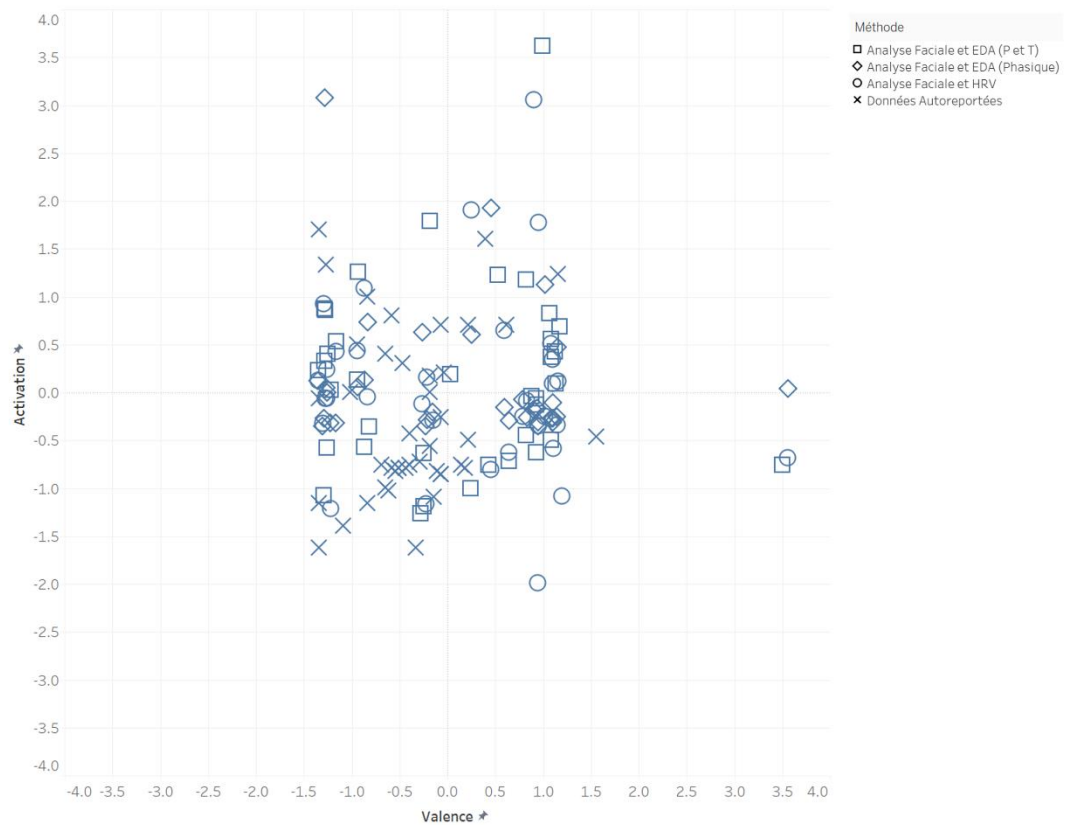
La figure 4.8 présente un graphique avec les émotions moyennes mesurées par chaque méthode avec le point de la valeur théorique également inclus pour l'image « 1_2 ». Comme avec l'image précédente, la méthode de mesure de l'émotion qui se rapproche le plus de la valeur attendue est les données autodéclarées (avec une valence de -0,35 et une activation de -0,29), ce qui représente une distance de 0,43 des coordonnées des valeurs théoriques, la plus petite distance observée pour cette méthode de mesure de l'émotion. À titre informatif, les distances euclidiennes mesurées sont de 0,92 pour l'analyse faciale avec l'EDA phasique et tonique, 0,94 pour l'analyse faciale avec l'EDA phasique et 0,88 pour l'analyse faciale avec le HRV. L'émotion théorique pour cette image est une valence de -0,27 et une activation de -0,71, ce qui correspond à une image qui provoque une réaction faiblement négative d'une faible intensité.

Figure 4.8 — Graphique illustrant l'émotion (valence et activation) moyenne mesurée pour chaque méthode en comparaison avec la valeur théorique pour l'image de calibration 1_2



De plus, les valeurs extrêmes ne semblent pas avoir un effet sur les moyennes obtenues, ce qui est présenté dans la figure 4.9.

Figure 4.9 — Graphique illustrant l'émotion (valence et activation) mesurée pour chaque participant par méthode pour l'image de calibration 1_2



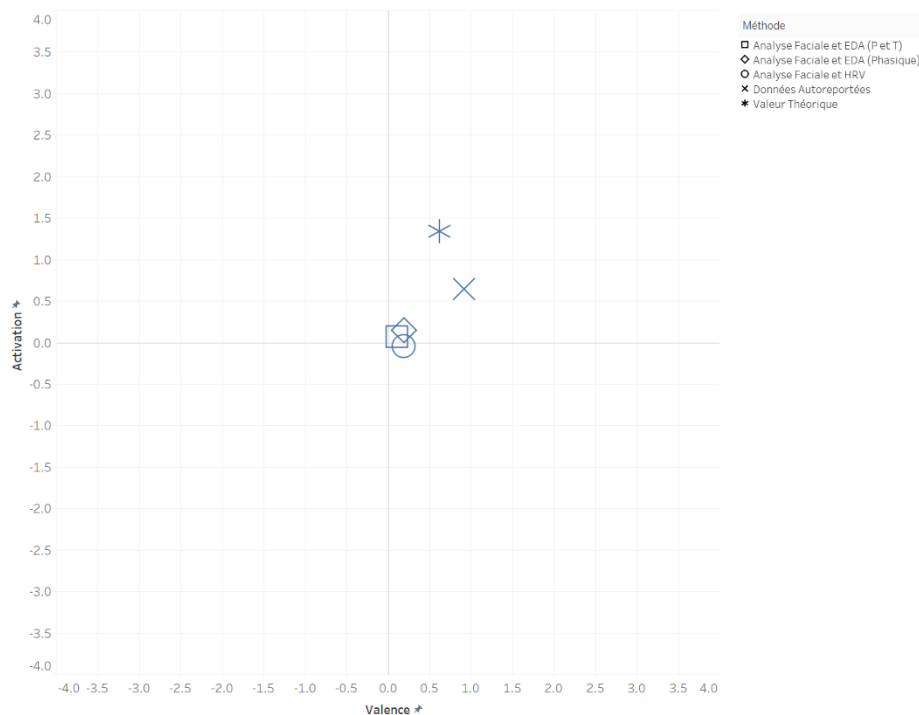
Dans un dernier temps, la tendance se maintient avec les images qui génèrent des émotions positives. La figure 4.10 présente l'image de calibration « 2_1 », qui devrait générer une émotion positive.

Figure 4.10 — Image de calibration 2_1



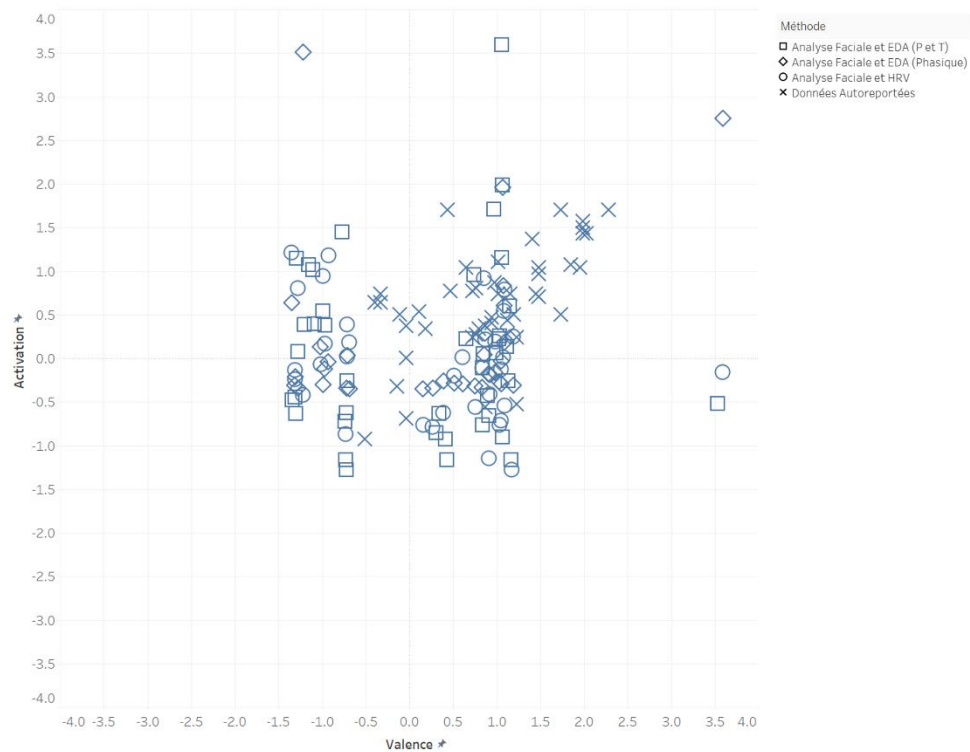
La figure 4.11 démontre les émotions moyennes mesurées pour les méthodes étudiées pour l'image de calibration «2_1». La tendance se maintient en ce qui concerne la performance de l'émotion mesurée avec les données du curseur affectif. Avec une distance euclidienne de 0,76 (en comparaison avec 1,37 pour l'analyse faciale et l'EDA phasique et tonique, 1,27 pour l'analyse faciale avec l'EDA phasique et 1,45 pour l'analyse faciale avec l'HRV), cette méthode se rapproche le plus de la valeur théorique de l'émotion générée par l'image (valence de 0,62 et activation de 1,34). L'émotion théorique représente une image générant une émotion positive d'une forte intensité.

Figure 4.11 — Graphique illustrant l'émotion (valence et activation) moyenne mesurée pour chaque méthode en comparaison avec la valeur théorique pour l'image de calibration 2_1



La figure 4.12 démontre le fait que les valeurs extrêmes n'ont pas un poids important pour le calcul des valeurs moyennes pour les mesures physiologiques.

Figure 4.12 — Graphique illustrant l'émotion (valence et activation) mesurée pour chaque participant par méthode pour l'image de calibration 2_1



En somme, avec l'analyse des statistiques descriptives et l'analyse graphique effectuée sur les données des distances euclidiennes, il est possible de cerner le fait que la mesure de l'émotion utilisant les données autodéclarées permet de mieux estimer les émotions, selon la valeur théorique attendue, que les méthodes utilisant les données physiologiques d'un participant. En considérant également les résultats du test sur le coefficient de corrélation qui présente des corrélations extrêmement faibles entre la mesure de l'émotion par le curseur affectif et les méthodes physiologiques, tant pour la valence que l'activation, il est possible de conclure que les données obtenues par le moyen du curseur affectif génèrent des informations qui mesurent les émotions de manière plus fiable et prédictible que les méthodes physiologiques.

4.3 Analyse des variances des émotions générées lors de la phase de calibration

Après la première partie de l'analyse, qui a démontré que les données expérimentales ne permettent pas d'effectuer le rôle des données de la phase de calibration, et la deuxième partie de l'analyse, qui a illustré une différence dans le comportement des données autodéclarées et physiologiques, la dernière piste d'étude a été entamée. L'objectif de cette étape est de vérifier si les différents signaux obtenus physiologiquement et de manière autodéclarée permettent de mesurer des différences significatives lorsque les participants étaient exposés à différents stimuli. Afin de pouvoir faire l'analyse, les données obtenues lors de la phase de calibration de l'étude originale ont été analysées. Les tests statistiques utilisés dans ce contexte étaient, en premier lieu, l'ANOVA et, dans un second temps, une comparaison deux à deux des différentes paires d'images de calibration. Comme les groupes comparés sont dépendants (les mêmes participants ont été analysés face aux différentes images de calibration), le type d'ANOVA à exécuter est une incluant des mesures répétées. Les ajustements faits sont présentés au chapitre précédent, à la section 3.3.3.

Concernant la structure des données utilisées, les données ont été standardisées au niveau de chaque participant dans le but de contrôler pour des participants qui auraient des réactions émotionnelles plus variables. La granularité est au niveau d'une valeur par stimulus par participant. Pour les méthodes physiologiques, la valeur est donc la moyenne observée pendant qu'un participant regardait activement le stimulus (une image de la phase de calibration). Comme les participants mesuraient leur émotion avec le curseur affectif seulement à la fin de la période d'exposition aux stimuli, il y a seulement une valeur par stimulus et par participant. Le fait d'avoir la même granularité pour toutes les variables permet de faire une comparaison directe entre les résultats obtenus. De plus, l'analyse a également été faite en écartant les observations de participants qui avaient regardé un stimulus pour moins d'une seconde. Cette analyse s'applique uniquement aux mesures physiologiques pour évaluer l'émotion.

Dans le but d’alléger la présentation des résultats, seulement les résultats significatifs (ou, dans certains cas, une sélection des résultats) sont présentés directement dans cette section. Les résultats complets des tests deux à deux sont présentés à l’annexe 4.

4.3.1 Analyse des variances pour les signaux mesurant la valence

Afin de bien contextualiser les résultats des ANOVAs et les visualisations par moyen de diagrammes en boîte, le tableau 4.10 présente les valeurs moyennes obtenues pour les douze images utilisées pour la calibration pour les deux méthodes pour mesurer la valence.

Tableau 4.10 — Moyennes observées pour chaque image de calibration pour les méthodes mesurant la valence

Méthode	Identificateur	Moyenne (Analyse Faciale)	Moyenne (Valence par le curseur affectif)
	1_1	-0,0380	-0,7455
	1_2	0,0576	-0,3722
	1_3	0,5860	-0,2875
	2_1	0,1120	0,9246

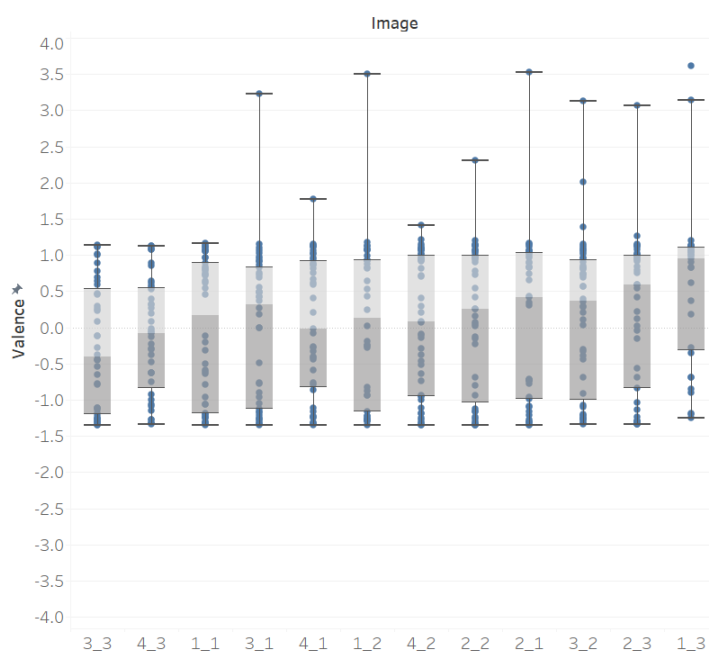
	2_2	0,1022	-0,3638
	2_3	0,2133	0,8863
	3_1	0,0037	-0,2767
	3_2	0,1661	0,9822
	3_3	-0,2815	-0,2917
	4_1	0,0267	0,9136
	4_2	0,0603	-0,3638

	4_3	-0,1399	-1,0250
---	-----	---------	---------

4.3.1.1 Analyse des variances pour la valence mesurée par l'analyse faciale

En premier lieu, les données de la valence mesurées par l'analyse faciale, classées par l'image de calibration qui avait été regardée par les participants, ont été comparées par le moyen de diagrammes en boîte (figure 4.13). En étudiant la figure, les moyennes semblent être autour de 0 avec certaines différences entre les images avec les plus petites moyennes et les plus grandes moyennes. De plus, l'ANOVA avec mesures répétées a été conçue avec un effet aléatoire sur l'ordonnée à l'origine appliqué sur les observations d'un même participant. Les résultats ajustés pour ce test sont significatifs à niveau de signification de 5 % avec une p-value de 0,01, donc il existe au moins une paire d'images qui a une valence significativement différente l'une de l'autre.

Figure 4.13 — Diagrammes en boîte de la valence mesurée physiologiquement pour chaque image de calibration (classés de manière croissante selon la moyenne)

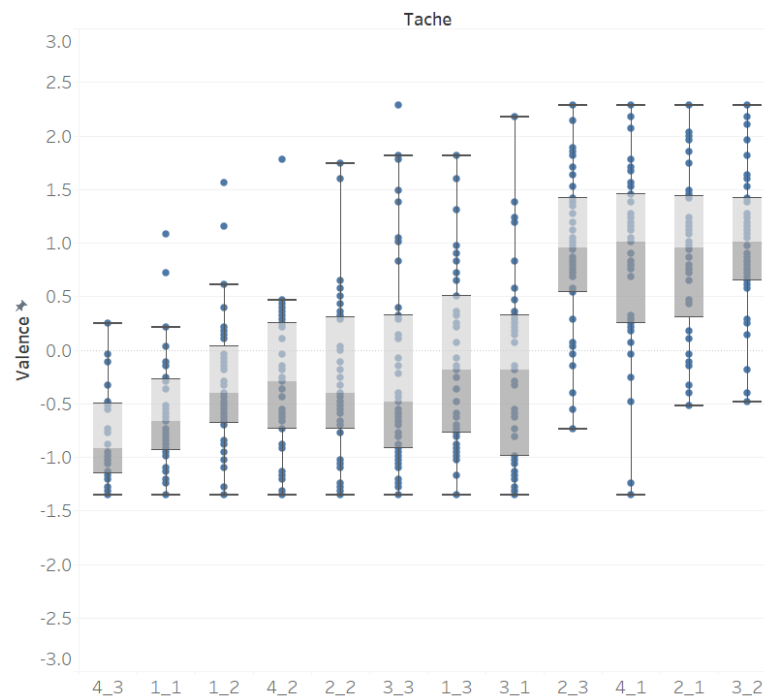


En analysant les paires d'images deux à deux, avec une correction de Tukey-Kramer, deux paires d'images sont significativement différentes à un niveau de signification de 5 %. La paire d'images (« 1_3 » et « 3_3 ») avec une p-value de 0,01 et la paire d'images (« 1_3 » et « 4_3 ») avec une p-value de 0,01. Bref, malgré la significativité de l'ANOVA pour la valence mesurée avec l'analyse faciale, les comparaisons deux à deux révèlent que très peu d'images (deux paires) génèrent une valence significativement différente des autres.

4.3.1.2 Analyse des variances pour la valence mesurées par le curseur affectif

En comparaison aux données de la valence physiologique, les diagrammes en boîte de la valence autodéclarée, pour chaque image de calibration, illustrent le fait que les images semblent générer des émotions différentes (figure 4.14). Cette analyse préliminaire est confirmée par les résultats de l'ANOVA. La p-value de l'ANOVA est inférieure à 0,0001 donc au moins une moyenne pour la valence mesurée avec le curseur affectif est significativement différente des autres à un niveau de signification de 5 %. Les tests deux à deux démontrent que 38 paires d'images sont significatives à un niveau de 5 % et deux autres paires sont significatives au niveau de signification de 10 %, donc 26 paires d'images ne sont pas significativement différentes les unes des autres à un niveau de signification de 10 %. Donc, les résultats de ces tests démontrent que la valence mesurée à l'aide du curseur affectif permet d'identifier des différences significatives dans cette composante de l'émotion vécue par les participants.

Figure 4.14 — Diagrammes en boîte de la valence mesurée par le curseur affectif pour chaque image de calibration (classés de manière croissante selon la moyenne)

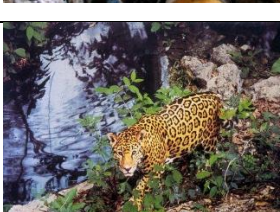


4.3.2 Analyse des variances pour les signaux mesurant l'activation

Comme pour la section précédente, le tableau 4.11 présente les valeurs moyennes obtenues pour les douze images utilisées pour la calibration pour les quatre méthodes pour mesurer l'activation. Ceci permet de mieux comprendre les résultats des ANOVAs et les diagrammes en boîte.

Tableau 4.11 — Moyennes observées pour chaque image de calibration pour les méthodes mesurant l'activation

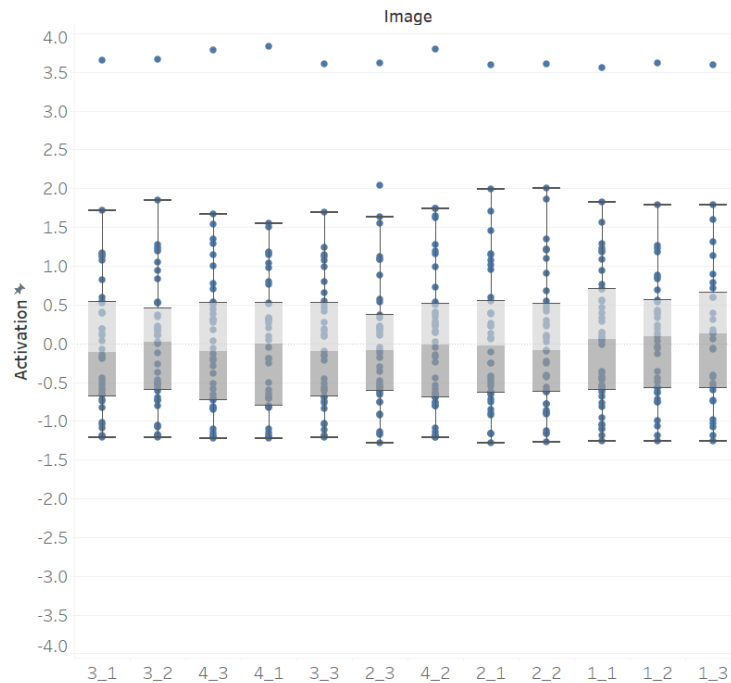
Méthode	Identificateur	Moyenne (EDA Phasique et Tonique)	Moyenne (EDA Phasique)	Moyenne (HRV)	Moyenne (Activation par le curseur affectif)
	1_1	0,0912	0,7597	0,1237	0,2819
	1_2	0,1341	0,1103	0,0413	-0,3020
	1_3	0,1506	0,1853	-0,1479	-0,5820
	2_1	0,0682	0,1366	-0,0486	0,6490
	2_2	0,0744	0,1134	-0,1272	-0,8851

	2_3	0,0537	0,2806	0,1755	0,5513
	3_1	0,0131	-0,0567	0,1870	-0,8014
	3_2	0,0212	0,7141	-0,0524	0,3112
	3_3	0,0337	0,0716	0,1159	0,0246
	4_1	0,0333	-0,0077	-0,1738	0,5853
	4_2	0,0587	0,0121	0,0918	-0,8262
	4_3	0,0331	-0,0891	0,0251	0,9332

4.3.2.1 Analyse des variances pour l'activation mesurées par l'EDA phasique et tonique

Concernant les mesures de l'activation, la première méthode étudiée est l'utilisation des signaux phasiques et toniques de l'EDA. Les diagrammes en boîte qui illustrent l'activation mesurée avec cette méthode montrent les résultats sont très homogènes entre les images (figure 4.15). Le résultat de l'ANOVA ne permet pas de démontrer qu'il existe au moins une paire d'images avec des moyennes de l'activation significativement différentes l'une de l'autre ($p\text{-value} = 0,06$) au niveau de signification 5 %. Comme il n'est pas possible de rejeter l'hypothèse nulle du test, il n'est pas pertinent d'exécuter les tests deux à deux. Bref, cette analyse ne permet pas de dire que les émotions moyennes mesurées par l'EDA phasique et tonique sont différentes.

Figure 4.15 — Diagrammes en boîte de l'activation mesurée par l'EDA phasique et tonique pour chaque image de calibration (classés de manière croissante selon la moyenne)



4.3.2.2 Analyse des variances pour l'activation mesurées par l'EDA phasique

Pour l'EDA utilisant uniquement le signal phasique, l'analyse des diagrammes en boîte des activations mesurées (figure 4.16) illustre un nombre de valeurs extrêmes très élevées. Afin de faciliter l'analyse, la figure 4.17 présente ces diagrammes en boîte en se concentrant sur les valeurs entre -4 et 4. Cette figure indique que, pour plusieurs images, les médianes des valeurs mesurées se trouvent à environ 0. Avec une p-value de 0,11, le résultat de l'ANOVA ne permet pas de conclure qu'il existe au moins une image qui a une moyenne de l'activation significativement différente des autres au niveau de signification de 5 %. Avec ce résultat, une analyse des résultats deux à deux n'est pas pertinente. Bref, en utilisant uniquement le signal phasique de l'EDA, il n'est pas possible de prouver que les différents stimuli ont généré des activations différentes chez les participants.

Figure 4.16 — Diagrammes en boîte de l'activation mesurée par l'EDA phasique pour chaque image de calibration (classés de manière croissante selon la moyenne)

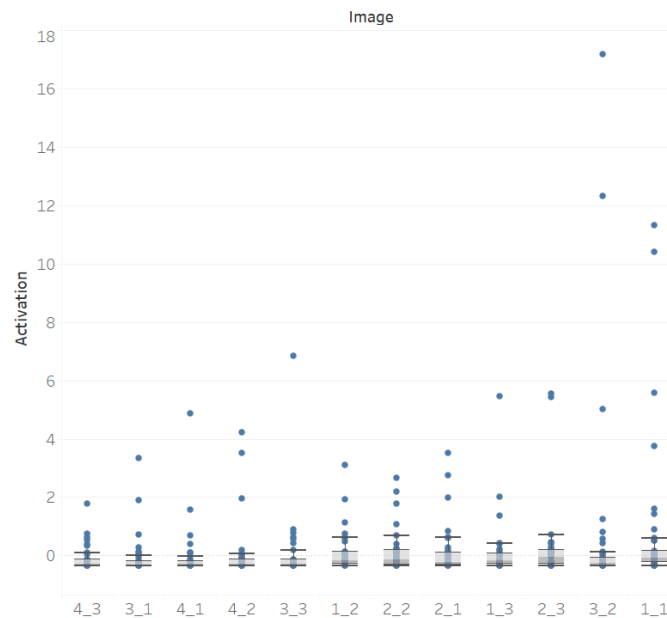
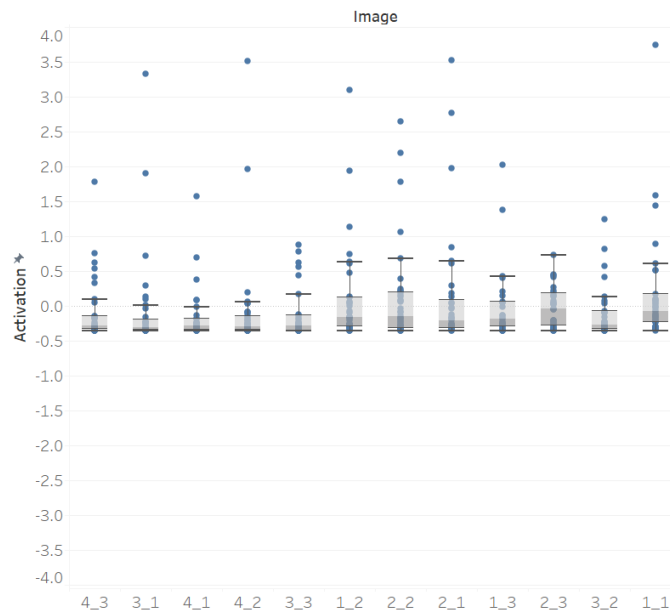


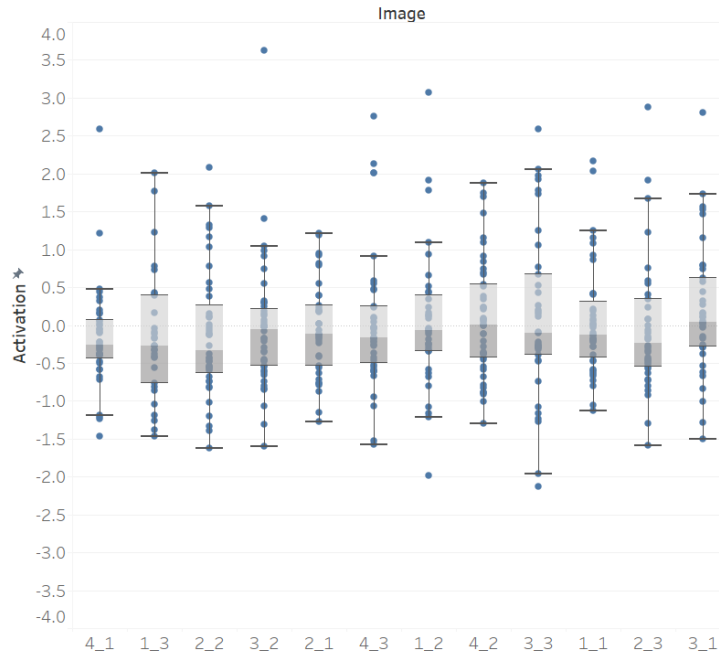
Figure 4.17 — Diagrammes en boîte de l’activation mesurée par l’EDA phasique pour chaque image de calibration (classés de manière croissante selon la moyenne; omettant les valeurs extrêmes)



4.3.2.3 Analyse des variances pour l’activation mesurées par le HRV

Concernant l’activation mesurée par le HRV, les diagrammes en boîte des activations mesurées (figure 4.18) semblent indiquer que les comportements des distributions sont tous semblables, avec des moyennes qui sont toutes proches de 0. À noter qu’une valeur extrême est omise du diagramme en boîte afin de faciliter la lecture (une valeur de 8,92 pour l’image « 2_3 »). En effectuant l’ANOVA, il n’est pas possible de rejeter l’hypothèse nulle supposant l’égalité des moyennes, car la p-value obtenue est de 0,75, ce qui est supérieur au niveau de signification de 5 % utilisé pour le test. En raison du résultat de l’ANOVA, l’exécution des tests deux à deux n’est pas pertinente. Bref, les données obtenues de l’activation mesurée par le HRV ne permettent pas de dire qu’il existe des différences significatives dans les valeurs.

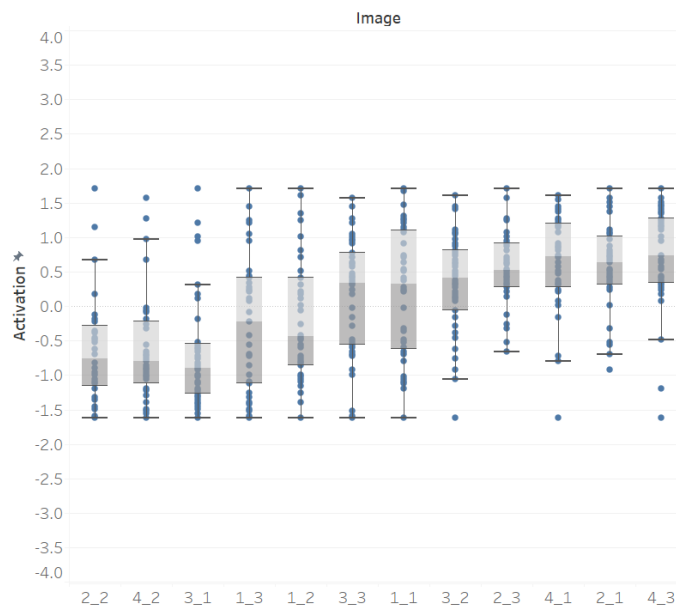
Figure 4.18 — Diagrammes en boîte de l’activation mesurée par le HRV pour chaque image de calibration (classés de manière croissante selon la moyenne; omettant une valeur extrême)



4.3.2.4 Analyse des variances pour l’activation mesurées par le curseur affectif

La dernière analyse effectuée est de regarder les valeurs de l’activation mesurée par les données autodéclarées par les participants de l’étude. Les diagrammes en boîte des distributions de l’activation (figure 4.19) semblent indiquer la présence d’une hétérogénéité des distributions. L’ANOVA, avec une p-value de moins de 0,0001, permet de confirmer qu’il existe au moins une image qui génère des émotions différentes des autres à un niveau de signification de 5 %. En effectuant les tests deux à deux, pour un niveau de signification de 5 %, il est possible d’identifier 39 paires d’images qui sont significativement différentes entre-elles. En passant à un niveau de 10 %, 3 nouvelles paires deviennent significatives, portant le total à 42 paires d’images. Ceci veut dire que, sur les 66 paires d’images possibles, seulement 24 paires ne sont pas significatives à un niveau de 10 %. Bref, des quatre méthodes pour mesurer l’activation qui ont été analysées, la méthode utilisant des données autodéclarées se démarque comme celle qui permet de différencier le plus grand nombre d’images de calibration.

Figure 4.19 — Diagrammes en boîte de l’activation mesurée par le curseur affectif pour chaque image de calibration (classés de manière croissante selon la moyenne)



4.3.3 Analyse des variances pour les signaux physiologiques de durée suffisante

Dans le but de présenter une analyse complète, qui permettra de mieux interpréter les résultats, le tableau 4.12 présente les valeurs moyennes pour tous les signaux physiologiques en ne conservant que les observations où le participant a regardé le stimulus pendant au moins 1 seconde.

Tableau 4.12 — Moyennes observées pour chaque image de calibration pour les méthodes physiologiques avec un regard total d’au moins une seconde

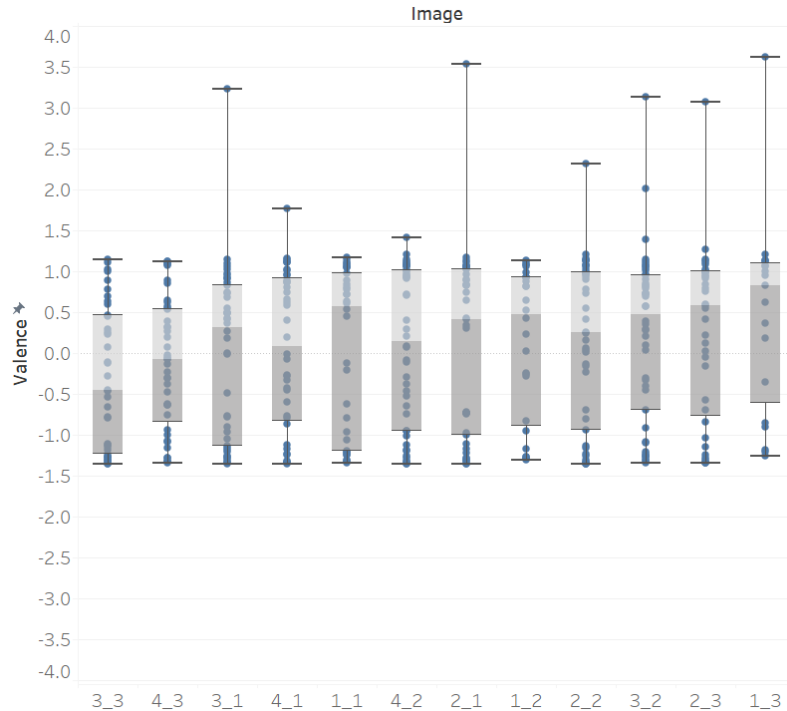
Méthode	Identificateur	Moyenne (Analyse faciale)	Moyenne (EDA Phasique et Tonique)	Moyenne (EDA Phasique)	Moyenne (HRV)
	1_1	0,0489	0,1596	0,4557	0,0711
	1_2	0,1259	0,0337	0,1082	0,1126
	1_3	0,4507	0,2130	0,1576	-0,2732
	2_1	0,1171	0,0125	0,1273	0,0652
	2_2	0,1332	0,1094	0,1474	-0,0806

	2_3	0,2288	0,0363	0,3336	0,0707
	3_1	0,0042	0,0161	-0,0499	0,1851
	3_2	0,2278	0,0001	0,5252	-0,0102
	3_3	-0,3160	-0,0186	0,1671	0,1062
	4_1	0,0441	-0,0158	-0,0404	-0,1615
	4_2	0,0711	0,1006	0,0916	0,1283
	4_3	-0,1397	0,0361	-0,0681	0,0252

4.3.3.1 Analyse des variances pour la valence mesurée par l'analyse faciale

En écartant les observations de courte durée, le diagramme en boîte des valeurs mesurées par l'analyse faciale, démontre une homogénéité générale, avec quelques images qui semblent présenter des moyennes différentes (figure 4.20). Le résultat ajusté pour ce test est significatif à niveau de signification de 5 % avec une p-value de 0,02, ce qui permet de dire qu'il existe au moins une paire d'images qui a une valence significativement différente l'une de l'autre.

Figure 4.20 — Diagrammes en boîte de l'analyse faciale pour chaque image de calibration en gardant que les observations où le participant a regardé pour au moins une seconde (classés de manière croissante selon la moyenne)

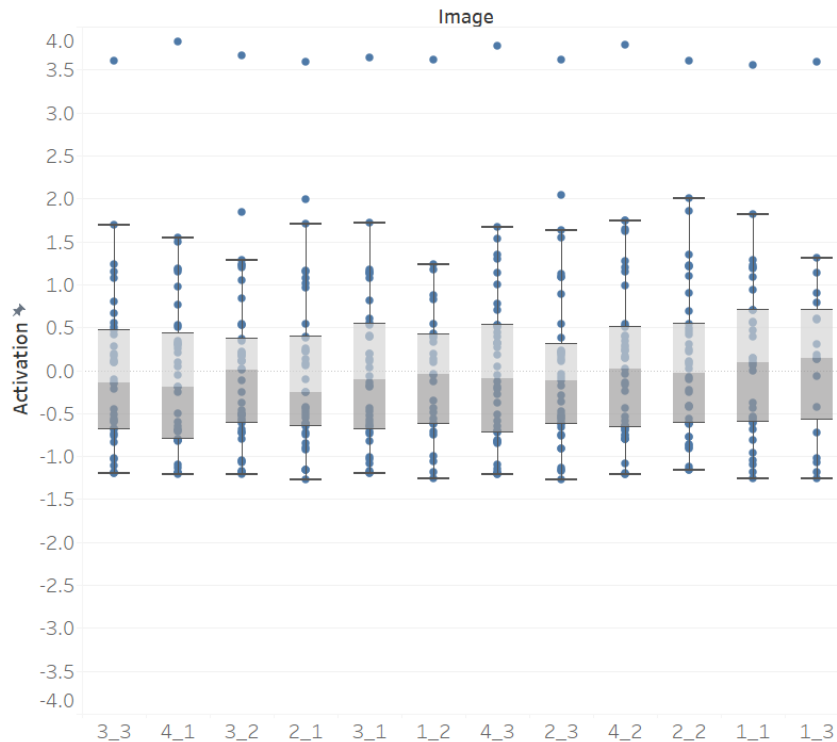


L'analyse des images deux à deux révèle que ce sont les mêmes paires d'images qui sont significatives que pour l'analyse conservant toutes les observations à un niveau de signification de 5 % (la paire d'images [« 1_3 » et « 3_3 »] avec une p-value de 0,02 et la paire d'images [« 1_3 » et « 4_3 »] avec une p-value de 0,03). Il est à souligner que la taille de l'échantillon pour l'image « 1_3 » est petite, soit de 19. Bref, les résultats se démarquent peu des résultats généraux.

4.3.3.2 Analyse des variances pour l'activation mesurée par l'EDA phasique et tonique

En ce qui concerne les données pour la combinaison des signaux phasiques et toniques de l'EDA, le diagramme en boîte démontre une tendance homogène pour les différents stimuli (figure 4.21). L'ANOVA ne permet pas de rejeter l'hypothèse nulle car la p-value est de 0,25 pour niveau de signification de 5 %. Aussi, comme pour l'analyse de la valence mesurée par l'analyse faciale, la taille de l'échantillon pour l'image « 1_3 » est de 19. Bref, comme pour l'analyse régulière de l'EDA phasique et tonique, il n'est pas possible de conclure que cette méthode mesure des activations significativement différentes entre-elles pour les stimuli utilisés dans l'étude.

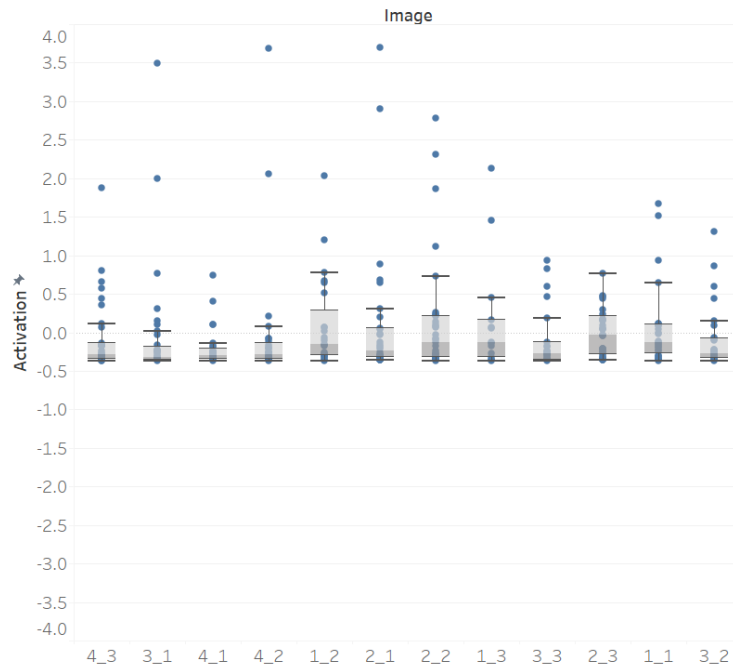
Figure 4.21 — Diagrammes en boîte de l'EDA phasique et tonique pour chaque image de calibration en gardant que les observations où le participant a regardé pour au moins une seconde (classés de manière croissante selon la moyenne)



4.3.3.3 Analyse des variances pour l'activation mesurée par l'EDA phasique

Pour l'EDA utilisant uniquement le signal phasique, l'analyse des diagrammes en boîte des activations mesurées se limite aux valeurs de -4 à 4 afin de faciliter l'interprétation (les mêmes valeurs extrêmes sont présentes). La figure 4.22 indique que les distributions semblent stables d'image en image. Avec une p-value de 0,77, le résultat de l'ANOVA ne permet pas de conclure qu'il existe au moins une image qui a une moyenne de l'activation significativement différente des autres au niveau de signification de 5 %. Il est important de souligner qu'une image (« 1_3 ») contient un faible nombre d'observations, soit 14. Bref, en se concentrant sur les observations où les participants ont regardé le stimulus pour au moins une seconde, il n'est pas possible de dire que l'EDA phasique permet de différencier entre les activations vécues par les participants.

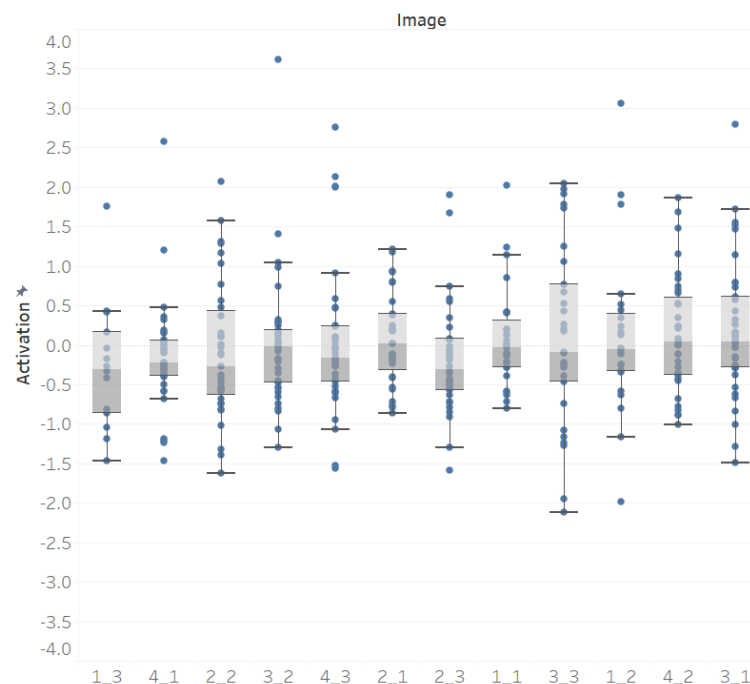
Figure 4.22 — Diagrammes en boîte de l'EDA phasique pour chaque image de calibration en gardant que les observations où le participant a regardé pour au moins une seconde (classés de manière croissante selon la moyenne; omettant les valeurs extrêmes)



4.3.3.3 Analyse des variances pour l'activation mesurée par l'HRV

Finalement, pour l'analyse des données du HRV, le diagramme en boîte démontre que les valeurs semblent être homogènes pour toutes les images présentées lors de la phase de calibration (figure 4.23) et la même valeur extrême a été omise de cette figure (8,92 pour l'image « 2_3 »). Avec une p-value de 0,94, au niveau de signification de 5 %, le résultat de l'ANOVA ne permet pas de rejeter l'hypothèse nulle, donc il n'est pas possible de dire qu'il y a au moins une image qui génère une activation significativement différente des autres. Aussi, avec un n de 14, l'image « 1_3 » a une taille d'échantillon petite. Donc, l'HRV, même en n'analysant que les observations où le participant a regardé une image pour au moins une seconde, ne permet pas d'identifier des différences générales dans les activations vécues par les participants.

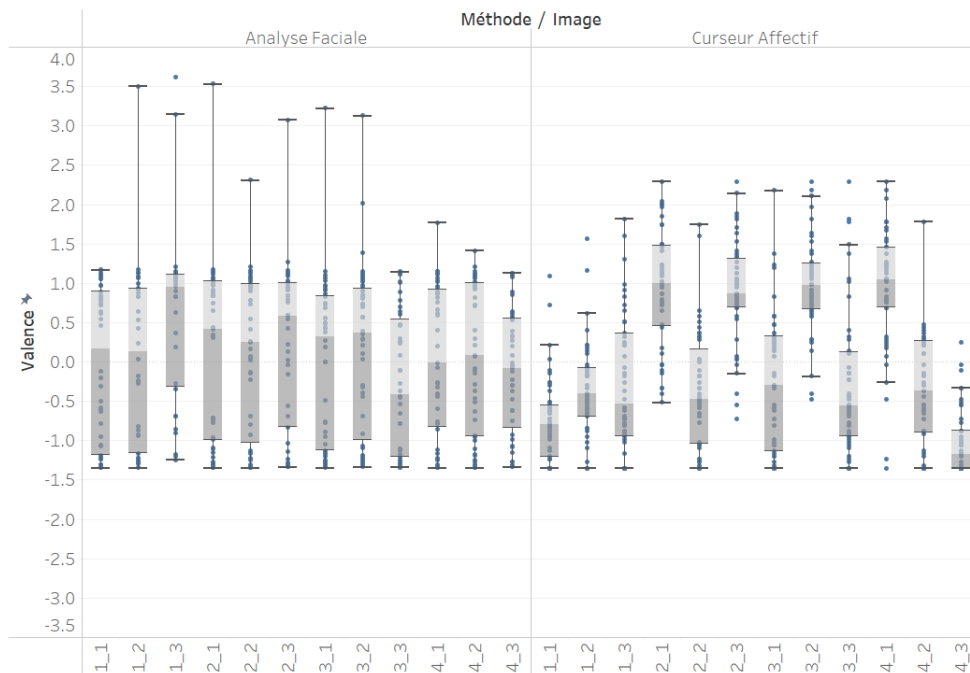
Figure 4.23 — Diagrammes en boîte de l'HRV pour chaque image de calibration en gardant que les observations où le participant a regardé pour au moins une seconde (classés de manière croissante selon la moyenne; omettant une valeur extrême)



4.3.4 Conclusion de l'analyse

En somme, autant pour la valence que pour l'activation, ce sont les méthodes utilisant le curseur affectif qui permettent d'identifier des différences significatives dans les émotions mesurées. Tant pour la mesure de l'activation que la valence, le curseur affectif permet d'identifier des valeurs significativement différentes dans la majorité des cas. En comparaison, la méthode physiologique pour identifier la valence (l'analyse faciale avec FaceReader) identifie les différences les plus prononcées dans les données et les méthodes physiologiques pour mesurer l'activation ne peuvent pas identifier de différences dans les émotions mesurées, donc elles semblent générer en majorité du bruit. Cette tendance peut être observée en comparant directement les distributions des valeurs mesurées pour chaque méthode. La figure 4.24 présente les valeurs mesurées par les techniques pour évaluer la valence.

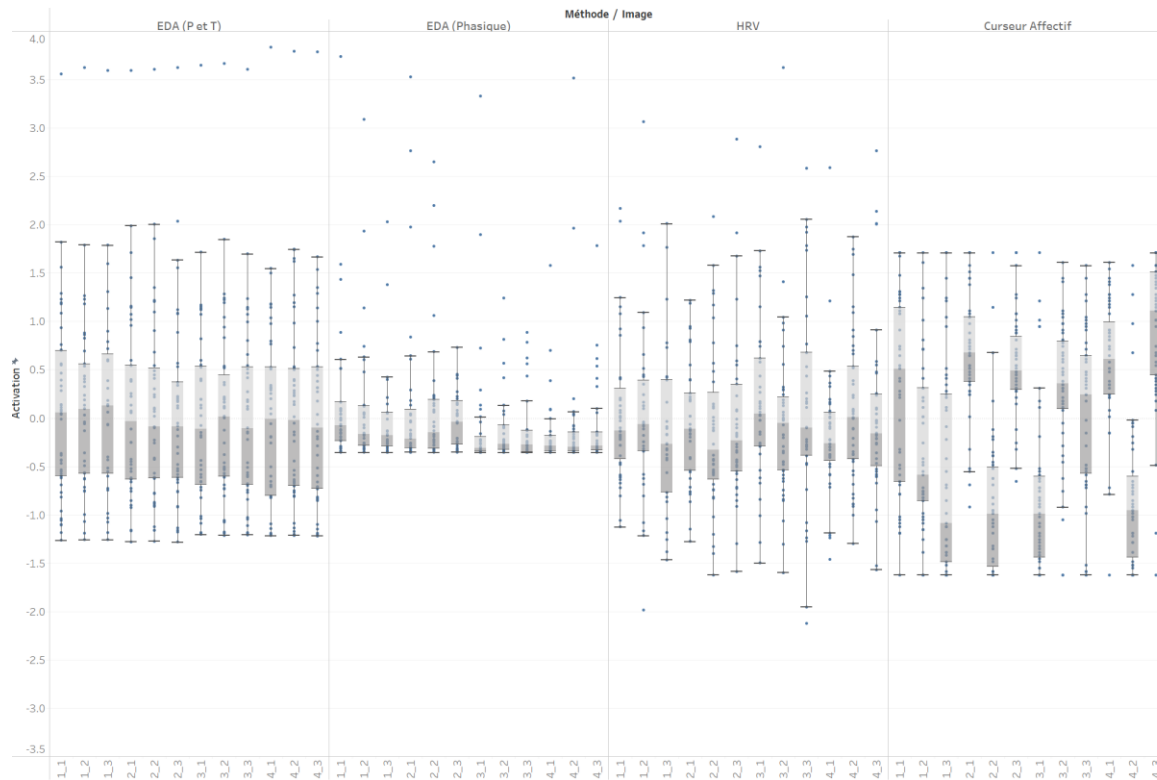
Figure 4.24 — Diagrammes en boîte des méthodes mesurant la valence



En observant les émotions mesurées, il est possible de constater que les distributions pour le curseur affectif sont souvent plus concentrées autour des médianes tandis que les distributions sont plus larges pour l'analyse faciale. Il est donc plus difficile de cerner des

différences pour la méthode physiologique, ce qui est démontré par le résultat de l'ANOVA.

Figure 4.25 — Diagrammes en boîte des méthodes mesurant l'activation (omettant les valeurs extrêmes)



La figure 4.25 présente les diagrammes en boîte pour chaque image de calibration et pour chaque méthode d'évaluation de l'activation, tout en écartant les valeurs les plus extrêmes (se limitant aux observations entre -3,5 et 4). Avec ces visualisations, il est possible de constater que le curseur affectif cerne des différences réelles entre les activations vécues par les participants (en présentant des distributions avec des moyennes clairement différentes entre les images), tandis que les méthodes physiologiques sont beaucoup plus centrées autour de 0.

En combinant les mesures de l'activation et de la valence dans un schéma en deux dimensions tout en utilisant les paires de techniques appropriées (le curseur affectif pour mesurer la valence et l'activation, l'analyse faciale pour la valence et l'EDA phasique et tonique pour l'activation, l'analyse faciale pour la valence et l'EDA phasique pour

l'activation et l'analyse faciale pour la valence et l'HRV pour l'activation), le problème illustré par l'analyse précédente devient plus clair. Les figures 4.26 (le curseur affectif pour mesurer la valence et l'activation), 4.27 (l'analyse faciale pour la valence et l'EDA phasique et tonique pour l'activation), 4.28 (l'analyse faciale pour la valence et l'EDA phasique pour l'activation) et 4.29 (l'analyse faciale pour la valence et le HRV pour l'activation) présentent ces plans cartésiens en deux dimensions. Il est possible de remarquer que les émotions mesurées avec les techniques physiologiques se centrent autour de l'origine, tandis que les émotions mesurées avec les mesures autodéclarées sont beaucoup plus variées se rapprochant plus des émotions qui devraient être générées par les images de calibration.

Figure 4.26 — Graphique des valences et activations moyennes par image de calibration pour la technique du curseur affectif

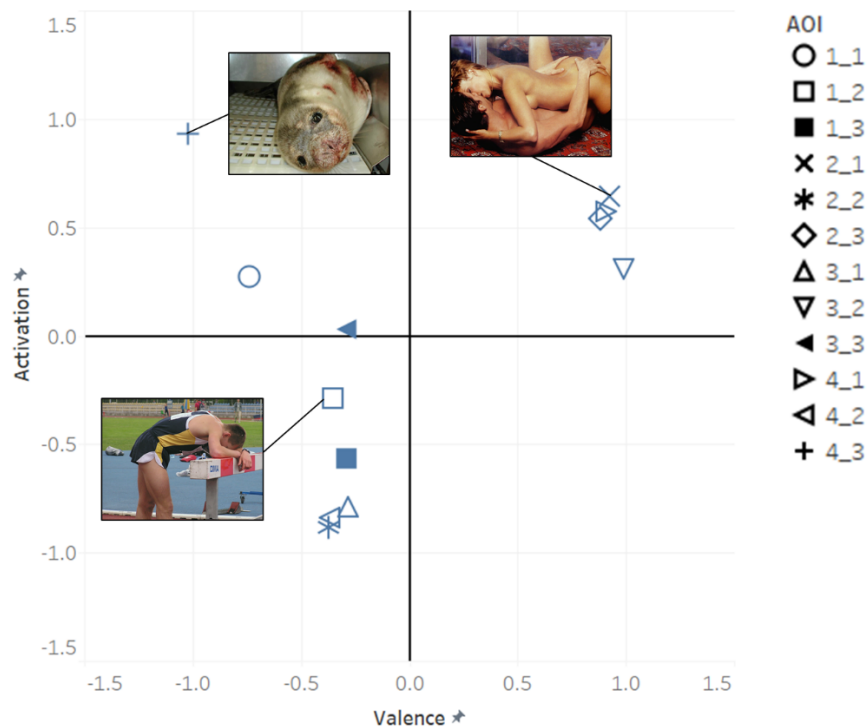


Figure 4.27 — Graphique des valences et activations moyennes par image de calibration pour la technique de l'analyse faciale avec l'EDA phasique et tonique

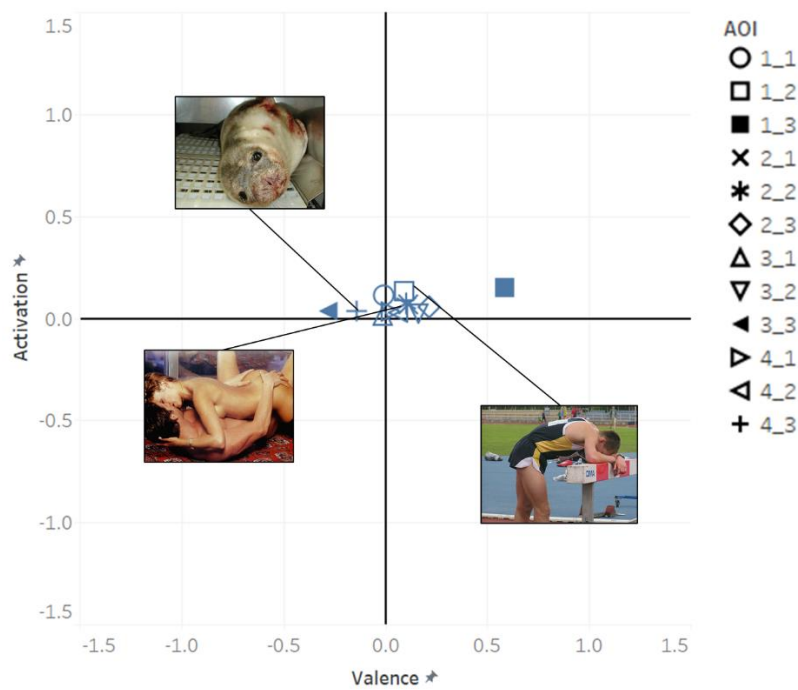


Figure 4.28 — Graphique des valences et activations moyennes par image de calibration pour la technique de l'analyse faciale avec l'EDA phasique

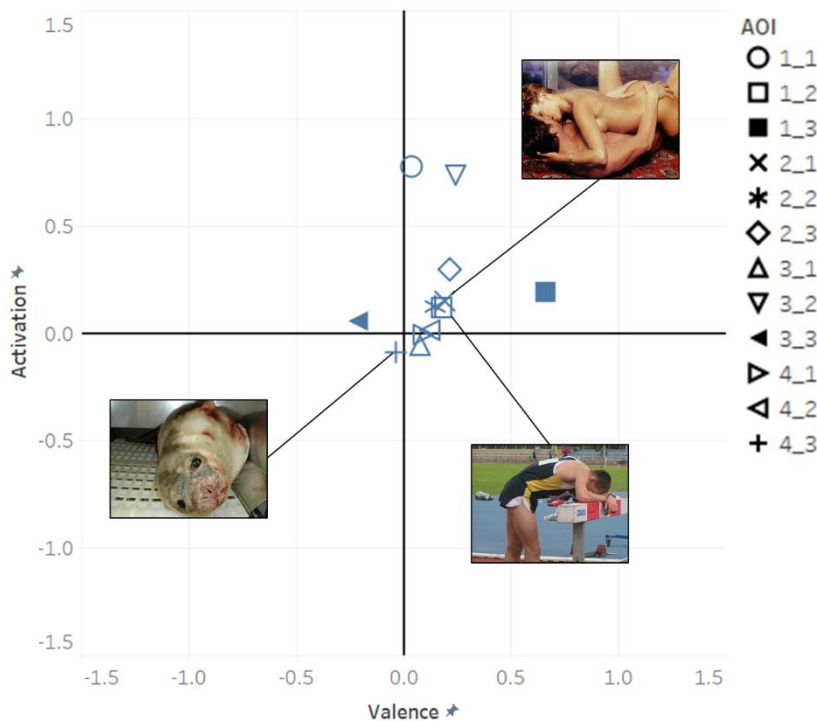
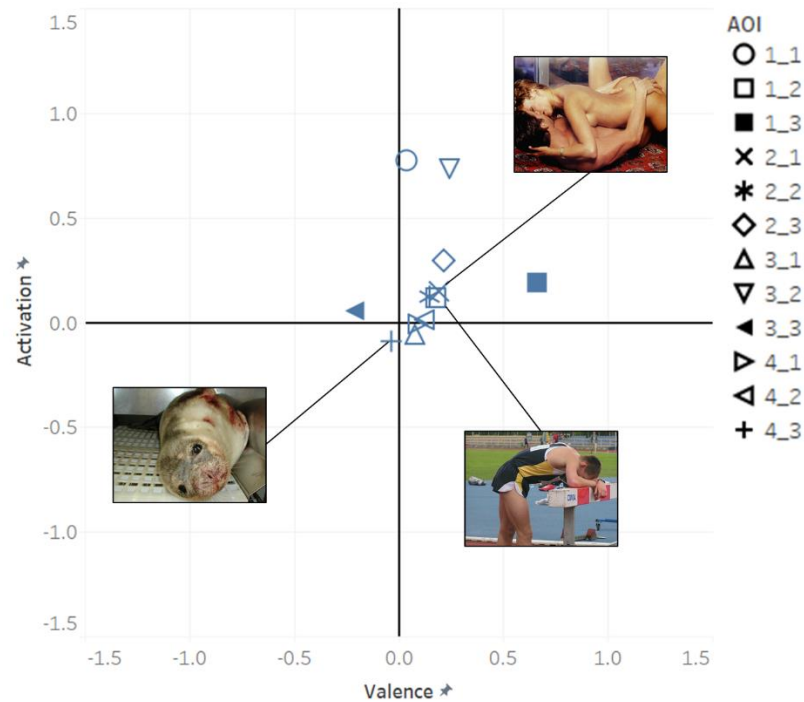


Figure 4.29 — Graphique des valences et activations moyennes par image de calibration pour la technique de l'analyse faciale avec l'HRV



Bref, les résultats démontrent la capacité des méthodes autodéclarées à identifier des différences réelles entre les émotions, la capacité de l'analyse faciale à identifier les différences les plus extrêmes et l'incapacité des méthodes physiologiques pour mesurer l'activation de cerner des différences entre les différentes images de calibration. De plus, ces constats s'appliquent que la durée des regards aux stimuli soit considérée ou non.

4.4 Conclusion

Dans ce chapitre, trois grandes analyses ont été effectuées. Dans un premier temps, des méthodes candidates pour calculer l'indice sans l'utilisation de la phase de calibration ont été comparées à la méthode originale à l'aide d'un test sur le coefficient de corrélation de Pearson. Les résultats ont démontré la présence d'une relation qui est plus ou moins forte. Par la suite, la relation entre les différents signaux mesurant l'activation et la valence a été étudiée et ensuite comparée avec les valeurs théoriques attendues des images utilisées lors de la phase de calibration. Les tests de corrélation ont démontré qu'il n'est pas possible de prouver l'existence d'une relation entre les méthodes autodéclarées pour

mesurer l'émotion et les méthodes physiologiques. De plus, une analyse descriptive de la relation entre les mesures de l'émotion et les valeurs théoriques indique que la méthode la plus fiable et la moins volatile pour mesurer les émotions semble être la méthode des données autodéclarées. Finalement, une analyse des moyennes des émotions mesurées par les méthodes discutées précédemment, par le moyen de l'ANOVA, démontre que la méthode la plus robuste pour cerner des différences dans les émotions, tant pour la valence que l'activation, est l'utilisation du curseur affectif. En considérant tous ces résultats, il est possible de conclure qu'il n'était pas nécessaire d'explorer les autres pistes de recherche originellement proposées pour l'étude, car ces résultats démontrent des problèmes avec les données physiologiques de l'étude originale. Bref, les analyses démontrent clairement la pertinence d'utiliser la méthodologie originale afin d'avoir des résultats valides dans de futures analyses utilisant l'index d'expérience émotionnelle.

Chapitre 5

Discussion

Résumé

Après la réalisation des analyses sur les données du projet de recherche original, certaines tendances émergent des résultats obtenus. Dans ce chapitre, les résultats seront interprétés dans le but de cerner les connaissances apprises lors du présent projet de recherche. De plus, les limites de l'étude actuelle seront soulignées afin de contextualiser ce qui a été appris et d'éviter des applications des faits qui sont incorrects.

5.1 Interprétation des résultats

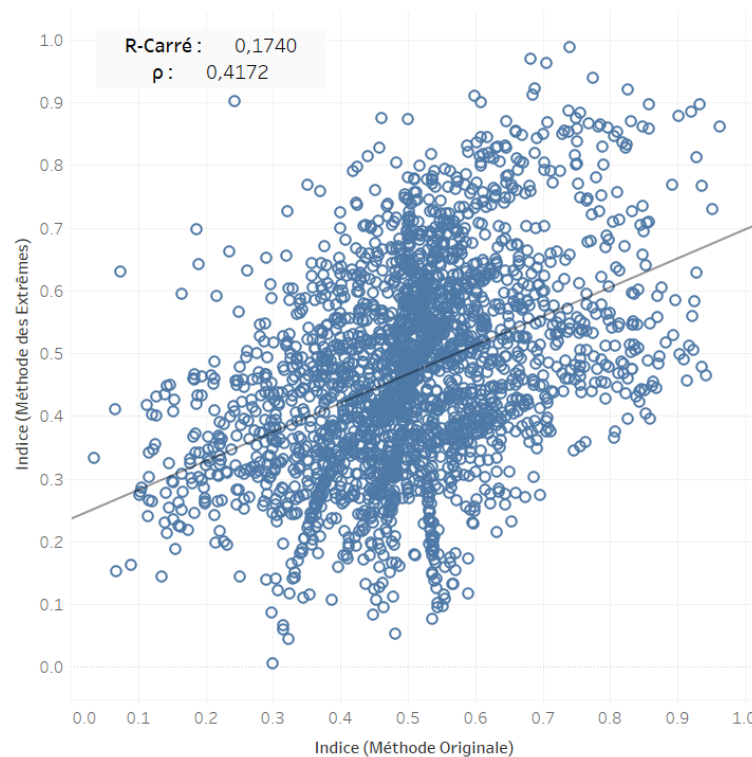
Les trois grandes analyses de ce projet de recherche, soit l'analyse de la corrélation entre des méthodes candidates pour mesurer l'indice sans la phase de calibration et la méthode de calcul originale, l'analyse de la relation entre les signaux physiologiques pour mesurer les composantes de l'émotion et la méthode autodéclarée et l'analyse des variances des valeurs obtenues par les différents signaux pour mesurer les composantes de l'émotion, ont tous démontré des résultats permettant de conclure des faits sur l'indice et les signaux pour mesurer les composantes de l'émotion. Les acquis des trois analyses seront présentés dans le même ordre que les analyses au chapitre précédent.

5.1.1 Interprétation de l'analyse de la corrélation entre des méthodes de calcul de l'indice sans la phase de calibration et la méthode originale

En guise de rappel, les résultats de l'analyse sur le coefficient de corrélation pour les méthodes qui n'utilisent pas la phase de calibration et la méthode originale qui utilise cette phase étaient mitigés. Malgré le fait que les corrélations étaient positives, la corrélation entre la méthode originale et la méthode utilisant les points extrêmes mesurés lors de la phase d'expérimentation avait un coefficient de 0,43 en utilisant les données plus granulaires (une valeur par seconde, stimulus reçu et participant) et un coefficient de 0,42 en utilisant les données plus agrégées (une valeur par stimulus reçu et par participant). Pour la corrélation entre la méthode originale et la méthode utilisant les 5^e et 95^e centiles des émotions mesurées lors de la phase d'expérimentation, le coefficient était de 0,52 en

utilisant les données plus granulaires et de 0,52 en utilisant les données plus agrégées. Ces données démontrent que la force de la relation entre les méthodes de calculs de l'indice avec et sans la phase de calibration analysée est seulement d'intensité moyenne. Il est donc possible de comprendre que le comportement des méthodes candidates n'est pas toujours le même que celui de la méthode originale. La figure 5.1 illustre un nuage de points mettant en relation les indices calculés pour la méthode originale et la méthode des points extrêmes.

Figure 5.1 — Nuage de points des méthodes de calcul de l'indice original et avec les points extrêmes (données agrégées au stimulus et au participant)



En observant le nuage de points, il est possible de voir que, malgré la tendance positive dans les données, les valeurs pour l'indice mesuré avec les points extrêmes, pour une même valeur de l'indice original, varient grandement. Par exemple, pour les données à environ 0,5 pour la méthode originale, les valeurs pour l'indice mesuré avec la méthode des extrêmes varient entre 0,27 et 0,87, ce qui représente une étendue de 0,6. Une tendance semblable se dégage de la relation entre l'indice original et l'indice calculé avec la méthode des centiles. Dans le cas de cette relation, pour les données à environ 0,5 pour la

méthode originale, l'étendue est de 0,74 (un minimum de 0,2 et un maximum observé de 0,94).

Cette relation peu prédictible permet donc de conclure qu'il n'est pas possible de simplement retirer la phase de calibration du calcul de l'indice sans devoir refaire une étude complète sur ce nouvel indice. Comme le fonctionnement de l'indice repose sur une étude initiale qui se fie sur l'utilisation de données de calibration, générées par les images des banques de données IAPS et GAPED, le retrait de cette phase dans le processus d'expérimentation voudrait dire que tout ce qui a été compris lors du projet de recherche original ne serait plus valide, d'où la nécessité de refaire un nouveau projet de recherche avec expérimentation afin de trouver une méthode de calcul sans indice. Bref, les données démontrent que, afin de pouvoir reposer sur les découvertes faites lors du premier projet de recherche sur l'indice, il est nécessaire d'utiliser les données d'une phase de calibration afin de calculer la valeur de l'indice.

5.1.2 Interprétation de l'analyse sur la relation entre les signaux physiologiques et autodéclarés pour mesurer la valence et l'activation

En ce qui concerne l'analyse des signaux qui mesurent les composantes de l'émotion, les résultats ont démontré l'absence d'une tendance claire concernant la relation entre les valeurs obtenues par le curseur affectif et les autres méthodes. En faisant une analyse des coefficients de corrélation de Pearson, les valeurs étaient très faibles et proches de zéro, donc il y a peu d'indications qu'il existe une relation entre ces différentes méthodes pour mesurer la valence et l'activation. Cette absence d'une relation mesurable entre les méthodes physiologiques et la méthode du curseur affectif amène à penser que ces variables, qui mesurent la même situation (dans ce cas, la valence et l'activation mesurées lors de la phase de calibration du projet de recherche original), ne sont pas directement comparables. En somme, les résultats de l'expérimentation permettent de croire que la méthode autodéclarée se démarque des méthodes physiologiques, car elle mesure l'émotion différemment.

Cette piste de réflexion est approfondie lors de l'analyse descriptive qui a découlé de l'analyse de la corrélation entre les types de signaux. Cette analyse a démontré que, de manière générale, la méthode la plus fiable (plus petite erreur d'estimation moyenne) et

la moins volatile (plus petit écart-type) pour estimer une émotion vécue par un participant pendant la phase de calibration était en utilisant les données autodéclarées. Comme rappel, la mesure étudiée était la distance euclidienne entre l'émotion (valence et activation) moyenne mesurée et la valeur de l'émotion théorique attendue selon les banques de données sources pour les images de calibration. Par le fait que ces valeurs théoriques ont été utilisées afin de déterminer les images pour la phase de calibration et que ces valeurs sont valides, il est attendu que les composantes de l'émotion mesurées seront proches de ces valeurs théoriques. Il est donc sous-entendu que le signal qui mesure l'émotion la plus près de la valeur attendue est la plus fiable. Bref, l'analyse des statistiques descriptives de la distance euclidienne indique que l'émotion mesurée par le curseur affectif donne des résultats qui s'alignent avec ce qui est attendu.

En combinant les apprentissages de la première et de la deuxième partie de cette analyse, il est possible de faire un constat dans le contexte de l'étude originale : la valence et l'activation mesurées par le curseur affectif ne sont pas directement comparables aux méthodes physiologiques pour mesurer la valence et l'activation. Comme indiqué précédemment, ni la valence ni l'activation mesurée par le curseur affectif ne possède une corrélation forte avec les méthodes autodéclarées. Cette incompatibilité entre les méthodes est renforcée lors de l'analyse des statistiques descriptives. Il faut rappeler le fait que les valeurs théoriques pour les images de calibration ont été mesurées avec l'échelle SAM, un outil de mesure émotionnelle autodéclarée. De plus, comme précisée au chapitre deux, l'échelle SAM est comparable au curseur affectif utilisé lors de la collecte de données. Il est donc possible de conclure que les méthodes physiologiques et la méthode autodéclarée, pour mesurer la valence et l'activation, ne sont pas directement comparables et que les attentes envers les émotions qui seraient mesurées lors d'une étude devraient être différentes dépendamment que si l'étude se fie sur une collecte de données avec le curseur affectif ou si ce sera une approche physiologique qui sera utilisée.

5.1.3 Interprétation de l'analyse des variances des émotions générées lors de la phase de calibration

La dernière piste d'analyse était de faire l'ANOVA comparant les émotions générées pendant que les participants à l'étude originale étaient exposés aux différentes images de

calibration. Le but principal était de déterminer s'il était possible de conclure des différences significatives dans les moyennes mesurées selon les différentes images. Étant donné que les images sélectionnées pour la phase de calibration avaient été choisies dans le but de générer des émotions particulières (ex. valence positive et activation forte; valence faible et activation faible), il est attendu de détecter des moyennes statistiquement différentes, au moins entre les images générant des émotions différentes. Par la suite, une analyse plus granulaire en faisant l'analyse des tests deux à deux pour toutes les paires d'images utilisées a été faite. Avec la méthodologie utilisée, il est possible d'identifier quelles images génèrent des émotions différentes.

Pour les données de la valence obtenues par le moyen de l'analyse faciale, les résultats sont mitigés. En analysant les images qui sont significativement différentes entre elles, seulement deux paires étaient différentes à un niveau de signification de 5 %. Ces paires d'images avaient l'image « 1_3 » (qui devrait générer une émotion faiblement positive avec une valence théorique de 0,28) et l'autre image de la paire était soit « 3_3 » (une valence théorique négative de -1,39) ou « 4_3 » (l'image avec la valence théorique la plus faible, soit -1,72). Les paires comparent une image d'une valence positive avec une image de valence négative, mais il faut mentionner qu'il y avait d'autres images qui génèrent des valences plus positives que les tests n'ont pas identifiées comme significativement différentes d'images à valence négative (par exemple, l'image « 3_2 » avait la valence la plus élevée des images utilisées et elle n'a jamais été identifiée comme une image générant une valence différente des autres). Le faible nombre de paires d'images significativement différentes entre elles indiquent que les données génèrent beaucoup de « bruit », donc peu d'information réellement utile pour différencier les émotions.

En ce qui concerne les données de l'activation, les résultats sont pires que ceux observés pour les données de la valence. Pour les trois signaux pouvant mesurer l'activation de manière physiologique, la conclusion est la même, il n'est pas possible d'identifier des différences significatives entre les activations mesurées lors de la phase de calibration. Bref, les techniques physiologiques n'ont pas été capables de différencier les activations vécues par les participants lors de la phase de calibration.

À l'opposé des résultats décrits dans les paragraphes précédents, les émotions mesurées par le curseur affectif sont beaucoup plus distinctes afin d'identifier quels stimuli ont généré des valeurs significativement différentes. Pour la valence, comme mentionné au chapitre des résultats, 39 paires d'images étaient significativement différentes entre-elles à un niveau de 10 %, ce qui représente environ 60 % des paires d'images possibles. De manière générale, les paires significatives correspondent aux paires qui devraient générer des émotions différentes, par exemple, une image générant une valence très négative et une image générant une valence très positive. De plus, les paires d'images non significatives correspondent aux images qui devraient générer des valences d'une intensité semblable. Un exemple de ceci serait les images « 2_1 » et « 2_3 », les deux images génèrent normalement des valences positives. Cette tendance se maintient en analysant les activations mesurées par le curseur affectif. Avec un taux de paires d'images significativement différentes, à un niveau de signification de 10 %, qui est de 64 %, les activations mesurées par la technique autodéclarée identifient clairement des différences de l'intensité de l'émotion vécue par les participants. Pour des raisons semblables à celles de la valence, les paires d'images non significatives sont typiquement des images qui devraient générer la même intensité d'activation, donc ce résultat serait attendu. Bref, les résultats de la méthode autodéclarée pour mesurer les composantes de l'émotion permet d'identifier les émotions différentes tant pour les émotions qui varient grandement que celles qui sont plus nuancées.

En analysant les résultats de cette expérimentation, il est possible d'identifier une interprétation possible : que la méthode autodéclarée pour mesurer l'émotion génère moins de « bruit » dans ses données que les méthodes physiologiques. Ceci veut donc dire que le curseur affectif serait un outil plus utile afin de détecter le niveau de la valence et de l'activation vécue par un participant suivant un stimulus. En comparaison, une méthode physiologique comme l'analyse faciale serait capable de mesurer les plus grandes différences de la valence, mais aurait plus de difficulté à identifier des émotions plus nuancées, donc les données auraient plus l'apparence d'être composées de « bruit » que d'information utilisable pour les analyses. Concernant l'activation, les résultats de l'analyse semblent indiquer que ce n'est que du « bruit » qui est généré. En somme, le

curseur affectif serait une technique plus résiliente et flexible pour mesurer l'émotion d'une personne en comparaison aux méthodes physiologiques analysées.

Cette première section du chapitre permet de constater deux grands éléments. Le premier étant que la phase de calibration est une partie intégrante de l'indice d'expérience émotionnelle et est nécessaire afin de pouvoir se fier sur les apprentissages faits lors du projet de recherche original. Le deuxième, par les analyses sur les signaux mesurant les composantes de l'émotion, est que la mesure de l'émotion par le curseur affectif génère des valeurs plus alignées avec ce qui est attendu selon les valeurs émotionnelles théoriques et permet de plus facilement identifier, tant pour la valence que l'activation, l'intensité et la direction de l'émotion vécue par une personne. En conclusion, il devient évident que les mesures autodéclarées demeurent pertinentes pour analyser les composantes de l'émotion.

5.2 Limites

En premier lieu, il est important de souligner le fait que les analyses de ce mémoire reposent entièrement sur des données déjà collectées et aucune nouvelle donnée n'a été générée pour ce mémoire. Toutes les données viennent du projet de recherche original de l'indice, ce qui limite la portée des analyses. Par exemple, en utilisant les données du projet original, il a été possible de réaliser qu'il existe peu de lien entre l'émotion mesurée par le curseur affectif et l'émotion mesurée physiologiquement. Cependant, il faudrait un nouveau projet de recherche pour savoir s'il est possible de traduire les données physiologiques pour être plus compatibles avec les données autodéclarées.

Une autre limite à souligner, qui découle également du fait que cette étude se fie uniquement sur des données déjà recueillies, est que les données ont un certain niveau de qualité qui reste fixe. Par exemple, l'analyse de la durée du regard pour les images de calibration démontre que la durée moyenne est de 4,16 secondes avec une médiane de 3,2 secondes. En plus, comme mentionné au chapitre 4, le nombre d'observations pour l'image « 1_3 » était petit et ceci peut être dû au fait que la durée du regard moyen pour cette image est de 1,6 seconde. En augmentant le seuil, certaines images de calibration

n'auraient pas pu être analysées, car il n'y aurait soit trop peu d'observations ou aucune observation.

5.3 Conclusion

En interprétant les résultats obtenus lors de ce projet de recherche, il a été possible d'analyser l'information afin de mieux comprendre les enjeux. En premier lieu, l'importance de la phase de calibration au calcul de l'indice a pu être remarquée, ce qui force son utilisation si les constats de l'étude originale veulent être utilisés. Par la suite, une analyse de la relation entre les signaux mesurant l'émotion a permis de réaliser que les méthodes physiologiques ne sont pas directement comparables à la méthode du curseur affectif, expliquant pourquoi les réponses autodéclarées permettent de mieux estimer les émotions vécues par un individu. La dernière analyse a révélé le fait que les réponses autodéclarées sont mieux adaptées pour identifier les différences dans l'émotion vécue. À l'inverse, les mesures physiologiques génèrent principalement du bruit qui fait en sorte que seulement les différences les plus flagrantes sont identifiables. En ce qui concerne les limites de ce projet de recherche, l'utilisation de données historiques, obtenues lors de l'étude originale, fait en sorte que certaines pistes d'analyse doivent être retenues pour des études futures.

Chapitre 6

Conclusion

Ce projet de recherche avait pour but original d'approfondir l'index d'expérience émotionnelle afin de le rendre plus sensible pour l'industrie de l'assurance. Plusieurs avenues de recherche avaient été proposées afin de guider ce projet et la première de ces pistes, le retrait de la phase de calibration dans l'identification des points p0 et p100 utilisés pour le calculer l'indice, avait été explorée. Les résultats, qui avaient démontré la nécessité de garder la phase de calibration afin de pouvoir se fier sur les connaissances du fonctionnement de l'indice ont mené à une exploration des différents signaux physiologiques et des émotions mesurées de manière autodéclarée. Ces analyses supplémentaires révèlent que les signaux physiologiques ont peu de corrélation avec le curseur affectif, tant pour la valence que l'activation, dans le cadre de l'étude originale, que le curseur affectif mesure des valeurs qui sont plus alignées avec les valeurs théoriques émotionnelles pour les images de calibration utilisées dans l'étude originale et que les données physiologiques de l'étude originale illustrent principalement du bruit, en opposition aux données autodéclarées qui permettent de distinguer entre les émotions (tant valence qu'activation) vécues par les participants lors de la phase de calibration de l'étude originale.

Ces résultats et leurs interprétations permettent d'identifier une avenue de recherche théorique potentielle et une contribution au domaine industriel.

6.1 Piste de recherche future

Comme expliqué au chapitre 5, ce projet se fie exclusivement aux données du projet original de l'indice, les avenues de recherche possibles pour ce projet sont limitées. Donc, comme les résultats de cette étude ont illustré qu'il y a peu de relation entre les mesures physiologiques de la valence et de l'activation et les mesures de ces composantes obtenues par le curseur affectif, il serait intéressant d'explorer la relation entre les données physiologiques et les banques d'images utilisées dans l'étude originale (IAPS et GAPED).

Comme souligné lors de la revue de la littérature, les mesures physiologiques, notamment l'EDA, capturent des émotions qui reflètent typiquement les mêmes tendances que les émotions capturées par l'échelle SAM (Bradley et Lang, 2007). Dans cette optique, la piste proposée précédemment mérite d'être explorée en plus grande profondeur. Actuellement, les banques d'images IAPS et GAPED fixent les valeurs attendues de la valence et de l'activation pour chaque image uniquement en fonction de valeurs autodéclarées (Bradley et Lang, 2007; Dan-Glauser et Scherer, 2011). Donc, une idée possible, pour permettre une utilisation de ces banques d'images dans un contexte où les données physiologiques sont capturées, serait d'établir des valeurs théoriques pour chaque image de la banque qui repose sur des valeurs obtenues physiologiquement.

6.2 Contribution industrielle

Étant donné les résultats obtenus, une conclusion de cette étude qui pourra aider l'implémentation d'un projet étudiant les émotions de personnes en industrie serait l'utilisation de techniques autodéclarées pour mesurer la valence et l'activation d'une personne. Dans un premier temps, cette observation se fie directement sur les résultats de cette étude, principalement l'analyse des variances pour les signaux physiologiques et les réponses autodéclarées. Comme ces résultats démontrent que les signaux physiologiques étaient principalement du bruit et que les méthodes autodéclarées permettent de mieux différencier entre la valence et l'activation vécue par une personne, l'utilisation d'un outil qui capte des réponses autodéclarées de participants serait mieux adaptée dans un contexte où la priorité est de mesurer la valence et l'activation d'une personne.

En parallèle, la technique autodéclarée qui a été utilisée lors de l'expérimentation, soit le curseur affectif, a un avantage intéressant : comme sa construction s'inspire des principes d'interfaces utilisateurs, il n'est pas nécessaire de fournir des instructions sur le fonctionnement de l'outil afin qu'il puisse être utilisé (Betella et Verschure, 2016). Cet avantage pourrait permettre un déploiement plus portatif de l'outil dans un contexte d'évaluation des émotions, car une personne peut fournir ses réponses avec cet outil sans avoir à interagir avec un humain ou lire des instructions.

Bref, il pourrait être intéressant, lorsqu'une entreprise désire explorer les réponses émotionnelles de la valence et de l'activation que leur produit ou service génère chez une personne, d'utiliser un outil autodéclaré comme le curseur affectif afin de maximiser la qualité des données et de simplifier le protocole d'expérimentation.

Bibliographie

- Betella, Alberto et Paul F.M.J. Verschure (2016). « The Affective Slider: A Digital Self-Assessment Scale for the Measurement of Human Emotions », *PLoS ONE*, vol. 11, p. 1-11.
- Bradley, Margaret M. et Peter J. Lang (1994). « Measuring emotion: the Self-Assessment Manikin and the Semantic Differential », *Journal of Behavior Therapy and Experimental Psychiatry*, vol. 25, no 1, p. 49-59.
- Bradley, Margaret M. et Peter J. Lang (2007). « The International Affective Picture System (IAPS) in the study of emotion and attention » dans James A. Coan et John J. B. Allen (dir.), *Handbook of emotion elicitation and assessment*, Oxford, Royaume-Uni, Oxford University Press, p. 29-46.
- Braithwaite, Jason J., Derrick G. Watson, Robert Jones, Mickey Rowe (2015). *A Guide for Analysing Electrodermal Activity (EDA) & Skin Conductance Responses (SCRs) for Psychological Experiments*, rapport technique, Birmingham, Royaume-Uni, University of Birmingham.
- Dan-Glauser, Elise et Klaus Scherer (2011). « The Geneva affective picture database (GAPED): A new 730-picture database focusing on valence and normative significance », *Behavior Research Methods*, vol. 43, no 2, p. 468-477.
- Dawson, Michael E., Anne M. Schell et Diane L. Filion (2007). « The Electrodermal System » dans John T. Cacioppo, Louis G. Tassinary et Gary G. Berntson (dir.), *Handbook of Psychophysiology*, Cambridge, Royaume-Uni, Cambridge University Press, p. 159-181.
- Dufey, Michele, Ana Fernandez et Rocío Mayol Troncoso. (2011). « Adding support to cross-cultural emotional assessment: Validation of the International Affective Picture System in a Chilean sample », *Universitas Psychologica*, vol. 10, no 2, p. 521 - 533.
- Gong, Xianmin et Duhua Wang (2016). « Applicability of the International Affective Picture System in Chinese older adults: A validation study », *PsyCh Journal*, vol 5, no 2, p. 117-124.
- Greco, Alberto, Gaetano Valenza, Antonio Lanata, Enzo Pasquale Scilingo et Luca Citi (2016). « cvxEDA: A Convex Optimization Approach to Electrodermal Activity Processing », *IEEE Transactions of Biomedical Engineering*, vol. 63, no 4, p. 797-804.
- Larsen, Jeff T., Catherine J. Norris et John T. Cacioppo (2003). « Effects of positive and negative affect on electromyographic activity over zygomaticus major and corrugator supercilii », *Psychophysiology*, vol. 40, no 5, p. 776-785.

- Labonte-LeMoyne, Élise, François Courtemanche, Marc Fredette et Pierre-Majorique Léger (2018). « How Wild is Too Wild: Lessons Learned and Recommendations for Ecological Validity in Physiological Computing Research », *International Conference on Physiological Computing Systems*, p. 123–130.
- Lane, Richard D., Kateri McRae, Eric M. Reiman, Kewei Chen, Geoffery L. Ahern et Julian F. Thayer (2009). « Neural correlates of heart rate variability during emotion », *NeuroImage*, vol. 44, no 1, p. 213-222.
- Lang, Peter J., Margaret M. Bradley et Bruce N. Cuthbert (2005). « International affective picture system (IAPS): affective ratings of pictures and instruction manual », rapport technique, Gainesville, États-Unis, University of Florida.
- Léger, Pierre-Majorique, Sylvain Sénécal, Marc Fredette, Laurence Dumont, François Courtemanche, Shang Lin Chen et Zi-Su Zhao (2018). *Index d'expérience émotionnelle : facteurs clés de succès des interfaces numériques canadiennes*, rapport scientifique, Montréal, Canada, HEC Montréal.
- Loijens, Leanne et Olga Krips (2018). *FaceReader Methodology Note*, white paper, Wageningen, Pays-Bas, Noldus.
- Loijens, Leanne, Olga Krips, Hans van Kuilenburg, Marten den Uyl et Paul Iva (2015). *Reference Manual: FaceReader Version 6.1*, Wageningen, Pays-Bas, Noldus.
- Mačiukaitė, Laura, Arvydas Kuzinas et Osvaldas Rukšėnas (2015). « The universality of the international affective picture system: ratings from a sample of Lithuanian students », *International journal of psychology: a biopsychosocial approach*, vol. 16, no 1, p. 111-126.
- Malik, Marek, J. Thomas Bigger, A. John Camm, Robert E. Kleiger, Alberto Malliani, Arthur J. Moss et Peter J. Schwartz (1996). « Heart rate variability: Standards of measurement, physiological interpretation, and clinical use », *European Heart Journal*, vol. 17, no 3, p. 354–381.
- Mutlu-Bayraktar, Duygu, Ozel Pinar, Fatih Altindis et Bulent Yilmaz (2020). « Relationship between objective and subjective cognitive load measurements in multimedia learning », *Interactive Learning Environments*. Récupéré de <https://www-tandfonline-com.proxy2.hec.ca/doi/full/10.1080/10494820.2020.1833042>
- Nickel, Peter et Friedhelm Nachreiner (2003). « Sensitivity and Diagnosticity of the 0.1-Hz Component of Heart Rate Variability as an Indicator of Mental Workload », *Human Factors*, vol. 45, no 4, p. 575-590.
- Paas, Fred G. W. C., Jeroen J. G. van Merriënboer et Jos J. Adam (1994). « Measurement of Cognitive Load in Instructional Research », *Perceptual and Motor Skills*, vol. 79, no 1, p. 419–430.

Posada-Quintero, Hugo F., and Ki H. Chon (2020). « Innovations in Electrodermal Activity Data Collection and Signal Processing: A Systematic Review », *Sensors*, vol. 20, no 2, p. 479-497.

Soares, Ana Paula, Ana P. Pinheiro, Ana Costa, Carla Sofia Frade, Montserrat Comesaña, Rita Pureza (2015). « Adaptation of the International Affective Picture System (IAPS) for European Portuguese », *Behavior Research Methods*, vol. 47, p. 1159-1177.

Russell, James A. (1980). « A Circumplex Model of Affect », *Journal of Personality and Social Psychology*, vol. 39, no 6, p. 1161-1178.

Russell, James A. et Albert Mehrabian (1977). « Evidence for a three-factor theory of emotions », *Journal of Research in Personality*, vol. 11, no 3, p. 273-294.

Zaman, Bieke et Tara Shrimpton-Smith (2006). « The FaceReader: measuring instant fun of use », *NordiCHI '06: Proceedings of the 4th Nordic conference on Human-computer interaction: changing roles*, p. 457-460

Annexes

Annexe 1 : Images IAPS et GAPED utilisées lors de la phase de calibration

Figure A1.1 – Image de calibration 1_1



Banque d'images : IAPS

Code de la banque d'images : 8230

Valence théorique standardisée : -1,3793

Activation théorique standardisée : 1.1324

Figure A1.2 – Image de calibration 1_2



Banque d'images : IAPS

Code de la banque d'images : 8121

Valence théorique standardisée: -0,2653

Activation théorique standardisée: -0,7094

Figure A1.3 – Image de calibration 1_3



Banque d'images : GAPED

Code de la banque d'images : N089

Valence théorique standardisée: 0.2846

Activation théorique standardisée: -2,0991

Figure A1.4 – Image de calibration 2_1



Banque d'images : IAPS

Code de la banque d'images : 4668

Valence théorique standardisée: 0,6232

Activation théorique standardisée: 1,3405

Figure A1.5 – Image de calibration 2_2



Banque d'images : GAPED

Code de la banque d'images : N024

Valence théorique standardisée: 0,5835

Activation théorique standardisée: -1,8993

Figure A1.6 – Image de calibration 2_3



Banque d'images : IAPS

Code de la banque d'images : 4698

Valence théorique standardisée: 0,9746

Activation théorique standardisée: 1,9753

Figure A1.7 – Image de calibration 3_1



Banque d'images : GAPED

Code de la banque d'images : N041

Valence théorique standardisée: 0,5251

Activation théorique standardisée: -1,4051

Figure A1.8 – Image de calibration 3_2



Banque d'images : GAPED

Code de la banque d'images : P129

Valence théorique standardisée: 2,1865

Activation théorique standardisée: -1.1909

Figure A1.9 – Image de calibration 3_3



Banque d'images : IAPS

Code de la banque d'images : 9832

Valence théorique standardisée: -1,3859

Activation théorique standardisée: -0,3764

Figure A1.10 – Image de calibration 4_1



Banque d'images : IAPS

Code de la banque d'images : 1650

Valence théorique standardisée: 1,0741

Activation théorique standardisée: 1,4654

Figure A1.11 – Image de calibration 4_2



Banque d'images : GAPED

Code de la banque d'images : N004

Valence théorique standardisée: -0,1744

Activation théorique standardisée: -0,4447

Figure A1.12 – Image de calibration 4_3



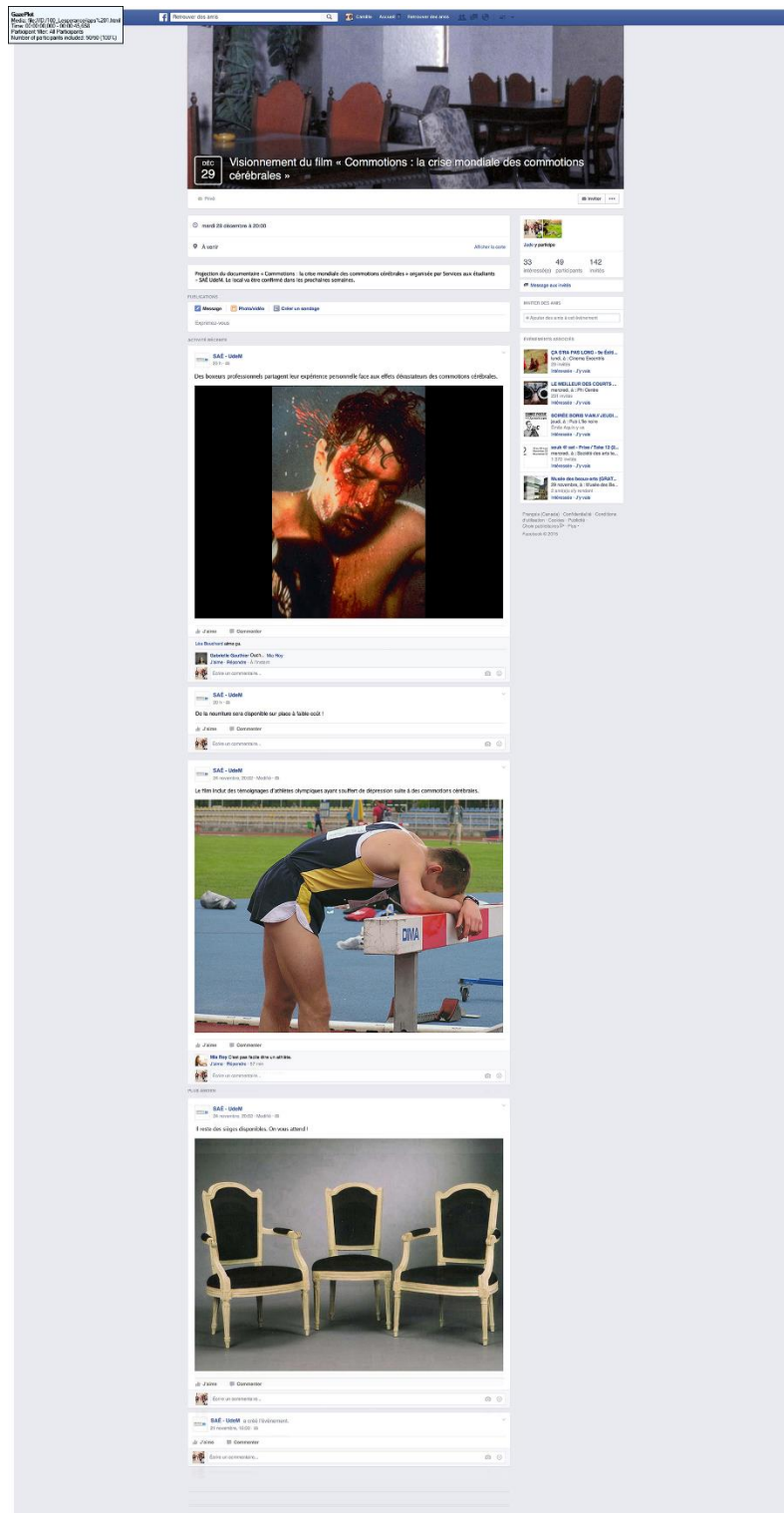
Banque d'images : GAPED

Code de la banque d'images : A007

Valence théorique standardisée: -1,7189

Activation théorique standardisée: 2,0232

Annexe 2 : Images de calibration dans leur contexte de présentation lors de l'étude (fil d'actualité Facebook)



xvi

Annexe 3 : Moyennes des valences et activations pour les images de calibration selon la méthode de mesure de l'émotion et distance euclidienne avec la valeur théorique

Méthode	Image	Valence	Activation	Distance de la valeur théorique
Curseur affectif	1_1	-0.7398	0.2727	1.0715
	1_2	-0.3531	-0.2887	0.4297
	1_3	-0.2914	-0.5670	1.6368
	2_1	0.9219	0.6437	0.7582
	2_2	-0.3733	-0.8818	1.3967
	2_3	0.8821	0.5413	1.4370
	3_1	-0.2815	-0.7906	1.0140
	3_2	0.9889	0.3073	1.9180
	3_3	-0.2890	0.0323	1.1706
	4_1	0.9074	0.5738	0.9070
	4_2	-0.3699	-0.8383	0.4394
	4_3	-1.0237	0.9337	1.2925
Analyse faciale + EDA (P et T)	1_1	-0.0085	0.1151	1.7070
	1_2	0.0934	0.1346	0.9170
	1_3	0.5869	0.1510	2.2704
	2_1	0.1132	0.0687	1.3703

	2_2	0.1034	0.0748	2.0316
	2_3	0.2144	0.0541	2.0661
	3_1	-0.0131	0.0134	1.5172
	3_2	0.1634	0.0216	2.3586
	3_3	-0.2801	0.0341	1.1795
	4_1	0.0280	0.0337	1.7731
	4_2	0.0615	0.0253	0.5258
	4_3	-0.1385	0.0335	2.5410
Analyse faciale + EDA (P)	1_1	0.0381	0.7750	1.4617
	1_2	0.1799	0.1176	0.9392
	1_3	0.6631	0.1887	2.3189
	2_1	0.1916	0.1488	1.2675
	2_2	0.1499	0.1214	2.0667
	2_3	0.2175	0.2944	1.8436
	3_1	0.0789	-0.0589	1.4182
	3_2	0.2414	0.7344	2.7368
	3_3	-0.2119	0.0543	1.2505
	4_1	0.0916	-0.0082	1.7710
	4_2	0.1221	0.0139	0.5461
	4_3	-0.0337	-0.0892	2.7023

Analyse faciale + HRV	1_1	0.0381	0.1233	1.7399
	1_2	0.1799	0.0461	0.8769
	1_3	0.6631	-0.1432	1.9922
	2_1	0.1916	-0.0468	1.4530
	2_2	0.1499	-0.1302	1.8214
	2_3	0.2175	0.1905	1.9387
	3_1	0.0789	0.2044	1.6702
	3_2	0.2414	-0.0558	2.2520
	3_3	-0.2119	0.1252	1.2766
	4_1	0.0916	-0.1750	1.9121
	4_2	0.1221	0.0536	0.5798
	4_3	-0.0337	0.0233	2.6152

Annexe 4 : Résultats des tests deux à deux de l'ANOVA

Note : Les p-values ajustées en vert sont significatives à un niveau de signification de 5%
et les p-values ajustées en jaune sont significatives à un niveau de signification de 10%

Résultats des tests deux à deux pour la valence mesurée par le curseur affectif

Images		Estimate	Standard Error	DF	t Value	Pr > t	Adj P
1_1	1_2	-0.3719	0.1134	68	-3.28	0.0016	0.0659
1_1	1_3	-0.4623	0.1468	68.6	-3.15	0.0024	0.0916
1_1	2_1	-1.6709	0.1276	77.6	-13.1	<.0001	<.0001
1_1	2_2	-0.3844	0.1354	68.8	-2.84	0.0059	0.1873
1_1	2_3	-1.6288	0.1197	78.5	-13.6	<.0001	<.0001
1_1	3_1	-0.4658	0.1499	65.2	-3.11	0.0028	0.1025
1_1	3_2	-1.7258	0.1135	83.8	-15.21	<.0001	<.0001
1_1	3_3	-0.4518	0.1532	70.1	-2.95	0.0043	0.1464
1_1	4_1	-1.6587	0.1349	71.1	-12.29	<.0001	<.0001
1_1	4_2	-0.3838	0.1293	70.7	-2.97	0.0041	0.1403
1_1	4_3	0.2805	0.09117	78.5	3.08	0.0029	0.1068
1_2	1_3	-0.09038	0.1557	76.2	-0.58	0.5634	1.0000
1_2	2_1	-1.2989	0.1378	82.2	-9.43	<.0001	<.0001
1_2	2_2	-0.01248	0.1451	76.3	-0.09	0.9317	1.0000
1_2	2_3	-1.2569	0.1306	80.5	-9.62	<.0001	<.0001
1_2	3_1	-0.09388	0.1587	72.3	-0.59	0.5559	1.0000
1_2	3_2	-1.3538	0.1249	81.7	-10.84	<.0001	<.0001
1_2	3_3	-0.07989	0.1618	77.8	-0.49	0.6228	1.0000
1_2	4_1	-1.2867	0.1445	76.3	-8.9	<.0001	<.0001
1_2	4_2	-0.01188	0.1394	76.4	-0.09	0.9323	1.0000
1_2	4_3	0.6524	0.105	65.8	6.22	<.0001	<.0001
1_3	2_1	-1.2086	0.1663	85.7	-7.27	<.0001	<.0001
1_3	2_2	0.07789	0.1724	83.9	0.45	0.6525	1.0000
1_3	2_3	-1.1665	0.1604	81.8	-7.27	<.0001	<.0001
1_3	3_1	-0.0035	0.184	86.6	-0.02	0.9849	1.0000
1_3	3_2	-1.2634	0.1558	78	-8.11	<.0001	<.0001
1_3	3_3	0.01049	0.1867	89.5	0.06	0.9553	1.0000
1_3	4_1	-1.1964	0.172	86.2	-6.95	<.0001	<.0001
1_3	4_2	0.0785	0.1676	83.1	0.47	0.6407	1.0000
1_3	4_3	0.7428	0.1404	60.8	5.29	<.0001	0.0001
2_1	2_2	1.2865	0.1564	84.2	8.23	<.0001	<.0001
2_1	2_3	0.04203	0.143	86.2	0.29	0.7696	1.0000
2_1	3_1	1.2051	0.1691	83.5	7.13	<.0001	<.0001

2_1	3_2	-0.05488	0.1378	87	-0.4	0.6915	1.0000
2_1	3_3	1.219	0.172	87.2	7.09	<.0001	<.0001
2_1	4_1	0.01221	0.156	88.4	0.08	0.9378	1.0000
2_1	4_2	1.2871	0.1511	86.1	8.52	<.0001	<.0001
2_1	4_3	1.9514	0.1202	69.7	16.24	<.0001	<.0001
2_2	2_3	-1.2444	0.1501	81	-8.29	<.0001	<.0001
2_2	3_1	-0.08139	0.175	82.4	-0.47	0.6432	1.0000
2_2	3_2	-1.3413	0.1451	79.4	-9.24	<.0001	<.0001
2_2	3_3	-0.0674	0.1779	85.9	-0.38	0.7057	1.0000
2_2	4_1	-1.2742	0.1624	84.9	-7.84	<.0001	<.0001
2_2	4_2	0.000601	0.1577	79.7	0	0.997	1.0000
2_2	4_3	0.6649	0.1285	59.8	5.18	<.0001	0.0002
2_3	3_1	1.163	0.1632	78.8	7.12	<.0001	<.0001
2_3	3_2	-0.09691	0.1306	87	-0.74	0.4601	0.9998
2_3	3_3	1.177	0.1663	82.8	7.08	<.0001	<.0001
2_3	4_1	-0.02982	0.1496	85.1	-0.2	0.8425	1.0000
2_3	4_2	1.245	0.1446	83	8.61	<.0001	<.0001
2_3	4_3	1.9093	0.1118	71.9	17.07	<.0001	<.0001
3_1	3_2	-1.2599	0.1586	75.3	-7.94	<.0001	<.0001
3_1	3_3	0.01399	0.1891	88.6	0.07	0.9412	1.0000
3_1	4_1	-1.1929	0.1747	84.2	-6.83	<.0001	<.0001
3_1	4_2	0.08199	0.1703	80.1	0.48	0.6315	1.0000
3_1	4_3	0.7463	0.1436	58.2	5.2	<.0001	0.0002
3_2	3_3	1.2739	0.1618	78.1	7.88	<.0001	<.0001
3_2	4_1	0.06709	0.1447	82	0.46	0.644	1.0000
3_2	4_2	1.3419	0.1394	81.4	9.63	<.0001	<.0001
3_2	4_3	2.0063	0.1051	77.1	19.1	<.0001	<.0001
3_3	4_1	-1.2068	0.1775	88.6	-6.8	<.0001	<.0001
3_3	4_2	0.068	0.1732	84.3	0.39	0.6957	1.0000
3_3	4_3	0.7323	0.147	61.3	4.98	<.0001	0.0003
4_1	4_2	1.2748	0.1573	83.7	8.1	<.0001	<.0001
4_1	4_3	1.9392	0.1279	63.6	15.16	<.0001	<.0001
4_2	4_3	0.6643	0.122	61.6	5.45	<.0001	<.0001

Résultats des tests deux à deux pour la valence mesurée par l'analyse faciale

Images		Estimate	Standard Error	DF	t Value	Pr > t	Adj P
1_1	1_2	-0.07836	0.1365	444	-0.57	0.5661	1.0000
1_1	1_3	-0.412	0.1454	444	-2.83	0.0048	0.1696
1_1	2_1	-0.09837	0.133	444	-0.74	0.46	0.9999
1_1	2_2	-0.07552	0.1323	444	-0.57	0.5684	1.0000
1_1	2_3	-0.2191	0.1339	444	-1.64	0.1026	0.8950
1_1	3_1	-0.1462	0.1329	444	-1.1	0.2719	0.9946
1_1	3_2	-0.1951	0.1281	444	-1.52	0.1285	0.9338
1_1	3_3	0.1582	0.1348	444	1.17	0.2414	0.9907
1_1	4_1	-0.06796	0.132	444	-0.51	0.6069	1.0000
1_1	4_2	-0.1301	0.1289	444	-1.01	0.3133	0.9975
1_1	4_3	0.1373	0.132	444	1.04	0.299	0.9967
1_2	1_3	-0.3336	0.1507	444	-2.21	0.0273	0.5404
1_2	2_1	-0.02001	0.1402	444	-0.14	0.8865	1.0000
1_2	2_2	0.00284	0.1395	444	0.02	0.9838	1.0000
1_2	2_3	-0.1407	0.1411	444	-1	0.319	0.9977
1_2	3_1	-0.06781	0.1395	444	-0.49	0.6272	1.0000
1_2	3_2	-0.1167	0.1355	444	-0.86	0.3897	0.9994
1_2	3_3	0.2365	0.1415	444	1.67	0.0953	0.8802
1_2	4_1	0.0104	0.1392	444	0.07	0.9405	1.0000
1_2	4_2	-0.05177	0.1364	444	-0.38	0.7044	1.0000
1_2	4_3	0.2156	0.1392	444	1.55	0.1222	0.9259
1_3	2_1	0.3136	0.1482	444	2.12	0.0349	0.6117
1_3	2_2	0.3364	0.1476	444	2.28	0.0231	0.4926
1_3	2_3	0.1929	0.1491	444	1.29	0.1965	0.9797
1_3	3_1	0.2658	0.1484	444	1.79	0.074	0.8224
1_3	3_2	0.2169	0.1445	444	1.5	0.1341	0.9399
1_3	3_3	0.5701	0.1499	444	3.8	0.0002	0.0089
1_3	4_1	0.344	0.1478	444	2.33	0.0204	0.4586
1_3	4_2	0.2818	0.145	444	1.94	0.0526	0.7309
1_3	4_3	0.5492	0.1478	444	3.72	0.0002	0.0121
2_1	2_2	0.02285	0.1344	444	0.17	0.8651	1.0000
2_1	2_3	-0.1207	0.136	444	-0.89	0.3752	0.9992
2_1	3_1	-0.04781	0.1364	444	-0.35	0.7261	1.0000
2_1	3_2	-0.09669	0.1316	444	-0.73	0.4629	0.9999
2_1	3_3	0.2565	0.138	444	1.86	0.0637	0.7839
2_1	4_1	0.03041	0.1353	444	0.22	0.8223	1.0000
2_1	4_2	-0.03177	0.1321	444	-0.24	0.8101	1.0000
2_1	4_3	0.2356	0.1353	444	1.74	0.0822	0.8476
2_2	2_3	-0.1436	0.1353	444	-1.06	0.2893	0.9960

2_2	3_1	-0.07065	0.1355	444	-0.52	0.6023	1.0000
2_2	3_2	-0.1195	0.1307	444	-0.91	0.3608	0.9990
2_2	3_3	0.2337	0.1371	444	1.7	0.089	0.8655
2_2	4_1	0.007557	0.1344	444	0.06	0.9552	1.0000
2_2	4_2	-0.05461	0.1312	444	-0.42	0.6774	1.0000
2_2	4_3	0.2128	0.1344	444	1.58	0.1139	0.9143
2_3	3_1	0.07292	0.1373	444	0.53	0.5957	1.0000
2_3	3_2	0.02403	0.1325	444	0.18	0.8562	1.0000
2_3	3_3	0.3773	0.1389	444	2.72	0.0069	0.2219
2_3	4_1	0.1511	0.1362	444	1.11	0.2677	0.9942
2_3	4_2	0.08895	0.133	444	0.67	0.504	1.0000
2_3	4_3	0.3564	0.1362	444	2.62	0.0092	0.2734
3_1	3_2	-0.04889	0.1315	444	-0.37	0.7103	1.0000
3_1	3_3	0.3043	0.137	444	2.22	0.0269	0.5355
3_1	4_1	0.07821	0.1353	444	0.58	0.5634	1.0000
3_1	4_2	0.01604	0.1324	444	0.12	0.9036	1.0000
3_1	4_3	0.2834	0.1353	444	2.1	0.0367	0.6263
3_2	3_3	0.3532	0.1335	444	2.65	0.0084	0.2575
3_2	4_1	0.1271	0.1307	444	0.97	0.3313	0.9982
3_2	4_2	0.06493	0.1273	444	0.51	0.6103	1.0000
3_2	4_3	0.3323	0.1307	444	2.54	0.0113	0.3160
3_3	4_1	-0.2261	0.1369	444	-1.65	0.0992	0.8884
3_3	4_2	-0.2883	0.134	444	-2.15	0.032	0.5859
3_3	4_3	-0.02089	0.1369	444	-0.15	0.8788	1.0000
4_1	4_2	-0.06217	0.1312	444	-0.47	0.6358	1.0000
4_1	4_3	0.2052	0.1335	444	1.54	0.125	0.9295
4_2	4_3	0.2674	0.1312	444	2.04	0.0421	0.6668

Résultats des tests deux à deux pour l'activation mesurée par le curseur affectif

Images		Estimate	Standard Error	DF	t Value	Pr > t	Adj P
1_1	1_2	0.5858	0.1831	92.9	3.2	0.0019	0.0759
1_1	1_3	0.8642	0.2021	92.7	4.28	<.0001	0.0026
1_1	2_1	-0.3682	0.1571	74.8	-2.34	0.0217	0.4556
1_1	2_2	1.1663	0.1722	89.5	6.77	<.0001	<.0001
1_1	2_3	-0.2699	0.1499	66.3	-1.8	0.0764	0.8121
1_1	3_1	1.0862	0.1835	92.2	5.92	<.0001	<.0001
1_1	3_2	-0.02834	0.1671	84.8	-0.17	0.8658	1.0000
1_1	3_3	0.2583	0.1811	90.8	1.43	0.1572	0.9551
1_1	4_1	-0.3064	0.1605	79.3	-1.91	0.0599	0.7500
1_1	4_2	1.1083	0.1782	89.5	6.22	<.0001	<.0001
1_1	4_3	-0.6525	0.1692	86.2	-3.86	0.0002	0.0112
1_2	1_3	0.2784	0.1954	88.6	1.42	0.1577	0.9554
1_2	2_1	-0.9541	0.1484	79	-6.43	<.0001	<.0001
1_2	2_2	0.5804	0.1643	88.8	3.53	0.0007	0.0301
1_2	2_3	-0.8557	0.1408	69.2	-6.08	<.0001	<.0001
1_2	3_1	0.5004	0.1762	88.8	2.84	0.0056	0.1815
1_2	3_2	-0.6142	0.159	86.3	-3.86	0.0002	0.0110
1_2	3_3	-0.3276	0.1736	90.6	-1.89	0.0623	0.7638
1_2	4_1	-0.8922	0.152	82.9	-5.87	<.0001	<.0001
1_2	4_2	0.5224	0.1706	87.2	3.06	0.0029	0.1085
1_2	4_3	-1.2383	0.1611	89.3	-7.69	<.0001	<.0001
1_3	2_1	-1.2325	0.1712	71.9	-7.2	<.0001	<.0001
1_3	2_2	0.302	0.1851	81.2	1.63	0.1066	0.8921
1_3	2_3	-1.1341	0.1647	64.3	-6.89	<.0001	<.0001
1_3	3_1	0.222	0.1957	87.2	1.13	0.2599	0.9922
1_3	3_2	-0.8926	0.1804	80	-4.95	<.0001	0.0002
1_3	3_3	-0.606	0.1934	88.9	-3.13	0.0023	0.0908
1_3	4_1	-1.1706	0.1743	74.9	-6.71	<.0001	<.0001
1_3	4_2	0.244	0.1907	83.4	1.28	0.2043	0.9795
1_3	4_3	-1.5167	0.1824	82.9	-8.32	<.0001	<.0001
2_1	2_2	1.5345	0.1345	83	11.41	<.0001	<.0001
2_1	2_3	0.09838	0.1046	64.7	0.94	0.3504	0.9984
2_1	3_1	1.4545	0.1489	76.7	9.77	<.0001	<.0001
2_1	3_2	0.3399	0.1281	80.4	2.65	0.0096	0.2685
2_1	3_3	0.6265	0.1458	72.7	4.3	<.0001	0.0029
2_1	4_1	0.06187	0.1193	76.7	0.52	0.6055	1.0000
2_1	4_2	1.4765	0.1422	80.4	10.38	<.0001	<.0001
2_1	4_3	-0.2843	0.1307	77.8	-2.17	0.0327	0.5725
2_2	2_3	-1.4361	0.1261	75.1	-11.38	<.0001	<.0001

2_2	3_1	-0.08005	0.1647	84.3	-0.49	0.6282	1.0000
2_2	3_2	-1.1946	0.1461	89.1	-8.17	<.0001	<.0001
2_2	3_3	-0.908	0.1619	88.8	-5.61	<.0001	<.0001
2_2	4_1	-1.4726	0.1386	84.7	-10.63	<.0001	<.0001
2_2	4_2	-0.058	0.1587	83	-0.37	0.7157	1.0000
2_2	4_3	-1.8188	0.1485	91	-12.24	<.0001	<.0001
2_3	3_1	1.3561	0.1413	67.4	9.6	<.0001	<.0001
2_3	3_2	0.2415	0.1192	70.8	2.03	0.0464	0.6740
2_3	3_3	0.5281	0.138	64.8	3.83	0.0003	0.0144
2_3	4_1	-0.03651	0.1097	71	-0.33	0.7404	1.0000
2_3	4_2	1.3781	0.1343	69.7	10.26	<.0001	<.0001
2_3	4_3	-0.3826	0.1221	70.3	-3.13	0.0025	0.0946
3_1	3_2	-1.1146	0.1593	82.7	-6.99	<.0001	<.0001
3_1	3_3	-0.828	0.1739	89.2	-4.76	<.0001	0.0004
3_1	4_1	-1.3926	0.1525	80.2	-9.13	<.0001	<.0001
3_1	4_2	0.02206	0.171	83.5	0.13	0.8977	1.0000
3_1	4_3	-1.7387	0.1616	87.6	-10.76	<.0001	<.0001
3_2	3_3	0.2866	0.1565	82.6	1.83	0.0706	0.7959
3_2	4_1	-0.278	0.1322	82.5	-2.1	0.0386	0.6224
3_2	4_2	1.1366	0.1532	84.9	7.42	<.0001	<.0001
3_2	4_3	-0.6242	0.1427	87.2	-4.38	<.0001	0.0019
3_3	4_1	-0.5646	0.1495	76.8	-3.78	0.0003	0.0152
3_3	4_2	0.85	0.1683	90.7	5.05	<.0001	0.0001
3_3	4_3	-0.9108	0.1588	83.2	-5.74	<.0001	<.0001
4_1	4_2	1.4146	0.1459	82	9.69	<.0001	<.0001
4_1	4_3	-0.3461	0.1348	80.2	-2.57	0.0121	0.3150
4_2	4_3	-1.7608	0.1554	88.5	-11.33	<.0001	<.0001

**Résultats des tests deux à deux pour la valence mesurée par l'analyse faciale
(observations avec un regard d'au moins une seconde)**

Images		Estimate	Standard Error	DF	t Value	Pr > t	Adj P
1_1	1_2	-0.06492	0.1578	393	-0.41	0.681	1.0000
1_1	1_3	-0.4607	0.1778	393	-2.59	0.0099	0.2886
1_1	2_1	-0.1197	0.1465	393	-0.82	0.4143	0.9996
1_1	2_2	-0.09135	0.143	393	-0.64	0.5232	1.0000
1_1	2_3	-0.2367	0.1457	393	-1.62	0.105	0.8991
1_1	3_1	-0.1088	0.1425	393	-0.76	0.4455	0.9998
1_1	3_2	-0.1902	0.14	393	-1.36	0.175	0.9704
1_1	3_3	0.1643	0.1463	393	1.12	0.262	0.9935
1_1	4_1	-0.04039	0.1448	393	-0.28	0.7804	1.0000
1_1	4_2	-0.07408	0.1415	393	-0.52	0.601	1.0000
1_1	4_3	0.1371	0.142	393	0.97	0.3347	0.9983
1_2	1_3	-0.3958	0.1839	393	-2.15	0.032	0.5852
1_2	2_1	-0.05478	0.1544	393	-0.35	0.7229	1.0000
1_2	2_2	-0.02642	0.1516	393	-0.17	0.8618	1.0000
1_2	2_3	-0.1718	0.1536	393	-1.12	0.2642	0.9938
1_2	3_1	-0.0439	0.1511	393	-0.29	0.7716	1.0000
1_2	3_2	-0.1253	0.1489	393	-0.84	0.4006	0.9995
1_2	3_3	0.2292	0.1543	393	1.49	0.1381	0.9438
1_2	4_1	0.02454	0.1527	393	0.16	0.8724	1.0000
1_2	4_2	-0.00915	0.1499	393	-0.06	0.9513	1.0000
1_2	4_3	0.202	0.1506	393	1.34	0.1804	0.9730
1_3	2_1	0.341	0.1752	393	1.95	0.0523	0.7291
1_3	2_2	0.3693	0.1731	393	2.13	0.0335	0.5989
1_3	2_3	0.224	0.1749	393	1.28	0.201	0.9811
1_3	3_1	0.3519	0.1728	393	2.04	0.0424	0.6683
1_3	3_2	0.2704	0.1709	393	1.58	0.1143	0.9145
1_3	3_3	0.625	0.1757	393	3.56	0.0004	0.0213
1_3	4_1	0.4203	0.1747	393	2.41	0.0166	0.4042
1_3	4_2	0.3866	0.1714	393	2.26	0.0246	0.5099
1_3	4_3	0.5978	0.1727	393	3.46	0.0006	0.0290
2_1	2_2	0.02836	0.1373	393	0.21	0.8364	1.0000
2_1	2_3	-0.117	0.1394	393	-0.84	0.4017	0.9995
2_1	3_1	0.01089	0.1384	393	0.08	0.9373	1.0000
2_1	3_2	-0.07054	0.1349	393	-0.52	0.6013	1.0000
2_1	3_3	0.284	0.1417	393	2	0.0457	0.6902
2_1	4_1	0.07932	0.1397	393	0.57	0.5705	1.0000
2_1	4_2	0.04563	0.1364	393	0.33	0.7381	1.0000

2_1	4_3	0.2568	0.1374	393	1.87	0.0624	0.7777
2_2	2_3	-0.1454	0.1362	393	-1.07	0.2865	0.9958
2_2	3_1	-0.01747	0.1345	393	-0.13	0.8967	1.0000
2_2	3_2	-0.0989	0.131	393	-0.76	0.4507	0.9998
2_2	3_3	0.2557	0.1381	393	1.85	0.0649	0.7882
2_2	4_1	0.05096	0.1362	393	0.37	0.7084	1.0000
2_2	4_2	0.01727	0.1326	393	0.13	0.8965	1.0000
2_2	4_3	0.2284	0.1333	393	1.71	0.0875	0.8614
2_3	3_1	0.1279	0.1375	393	0.93	0.3528	0.9988
2_3	3_2	0.04648	0.1338	393	0.35	0.7285	1.0000
2_3	3_3	0.401	0.1409	393	2.85	0.0047	0.1652
2_3	4_1	0.1963	0.1386	393	1.42	0.1574	0.9598
2_3	4_2	0.1626	0.1353	393	1.2	0.2301	0.9886
2_3	4_3	0.3738	0.1363	393	2.74	0.0064	0.2098
3_1	3_2	-0.08143	0.1313	393	-0.62	0.5355	1.0000
3_1	3_3	0.2731	0.1371	393	1.99	0.047	0.6985
3_1	4_1	0.06843	0.1363	393	0.5	0.6158	1.0000
3_1	4_2	0.03475	0.133	393	0.26	0.794	1.0000
3_1	4_3	0.2459	0.1334	393	1.84	0.0661	0.7931
3_2	3_3	0.3546	0.1353	393	2.62	0.0091	0.2722
3_2	4_1	0.1499	0.1328	393	1.13	0.2597	0.9932
3_2	4_2	0.1162	0.1293	393	0.9	0.3693	0.9991
3_2	4_3	0.3273	0.1302	393	2.51	0.0123	0.3344
3_3	4_1	-0.2047	0.1398	393	-1.46	0.144	0.9493
3_3	4_2	-0.2384	0.1366	393	-1.75	0.0817	0.8460
3_3	4_3	-0.0272	0.137	393	-0.2	0.8427	1.0000
4_1	4_2	-0.03369	0.1341	393	-0.25	0.8018	1.0000
4_1	4_3	0.1775	0.1344	393	1.32	0.1872	0.9760
4_2	4_3	0.2112	0.1317	393	1.6	0.1097	0.9073