



UNIVERSITÉ  
DE GENÈVE



UNIVERSITÉ  
DE GENÈVE

FACULTÉ DES SCIENCES

# Fully empathic agent

réalisé par Brandy Lee Nogales Simon

sous la direction du Postdoc. Marios Fanourakis et

du Prof. Guillaume Chanel

## Table des matières

1. Introduction.....	3
2. State-of-the-art.....	7
2.1 Outils actuels de reconnaissance des émotions.....	7
2.2 Outils actuels pour la création d'un avatar.....	9
3. Méthode.....	10
3.1 Architecture générale.....	10
3.2 Implémentation.....	10
3.2.1 Détecteur d'émotions.....	11
3.2.2 Agent virtuel.....	11
3.2.3 Programme final.....	14
3.3 Testing.....	16
3.4 Méthode de validation.....	16
4. Résultats.....	17
4.1 Outcomes.....	17
4.2 Performance du logiciel.....	18
4.3 Validation.....	19
5. Discussion.....	20
Références.....	21
Annexe.....	23

# 1. Introduction

L'arrivée d'agents virtuels et robots dans nos vies nous fascine depuis des décennies. Qu'ils soient d'apparence humanoïde ou complètement inspirés par la science-fiction. Les avancées de la recherche technologique et de l'intelligence artificielle nous ont permis de rendre cette ambition possible. Nous avons tous déjà interagit au moins une fois avec un agent virtuel, Siri, Alexa, ou encore les chat-bots des services-clients. Plus ces derniers évoluent plus l'envie de nous identifier à eux augmente. Par exemple, des études ont montré que certains video-gamers s'identifient plus à leur propre avatar qu'à leur soi biologique (J. M. Beer et al., 2015). C'est ainsi qu'on peut, depuis quelques années, créer des avatars à notre image et les utiliser en tant que représentation de nous-mêmes. Ces énormes progrès nous conduisent pourtant à nous poser des questions incontournables quant à l'interaction machine-humain. C'est là que les sciences affectives font un travail considérable afin de comprendre et améliorer cette relation, en étudiant ses limites et ses utilités.

C'est pour ces raisons que j'ai décidé de consacrer mon travail à la création d'un agent virtuel empathique, afin d'améliorer et enrichir le lien qui peut être créé entre un avatar et un humain. Pour comprendre le principe d'«agent virtuel empathique» il faut d'abord éclaircir les différents concepts qui le composent : l'empathie, la notion d'agent virtuel et l'interaction entre ces derniers et les humains.

Nos émotions représentent nos états d'esprit et nos réactions face à des expériences ou situations particulières, elles sont visibles à travers nos expressions faciales et dépendent de plusieurs facteurs tels que l'humeur, le tempérament, la personnalité, etc. C'est grâce aux postulats de Darwin sur l'universalité de ces expressions (C. Darwin, 1890) qu'on va chercher à définir les émotions dites primaires et fondamentales, qui sont innées et donc invariantes entre les différentes cultures (P. Ekman et W. V. Friesen, 1971). La combinaison de ces dernières aboutit aux autres émotions, dites secondaires, voir tertiaires.

Dans ce travail, je vais me concentrer sur les six émotions de base d'après Ekman (1999) :

- **Colère** : réponse à un obstacle qui interfère avec la poursuite d'un objectif qui nous tient à cœur, à la frustration ou un sentiment d'injustice. Implique souvent le désir de blesser l'objet déclencheur.
- **Peur** : réponse à la menace d'atteinte physique ou psychologique. Elle active des pulsions bloquantes («freeze») ou stimulantes («fight», «flight»). La peur déclenche souvent la colère.

- **Surprise** : réponse à un événement soudain et inattendu, qui va à l'encontre de nos attentes. En principe brève et anticipant une autre émotion.
- **Tristesse** : réaction à la perte (d'un objet ou d'une personne), à une déception, au sentiment d'impuissance. Elle se caractérise par une baisse d'énergie et de motivation.
- **Dégoût** : répulsion envers un objet, détecté par la vue, l'odeur ou le goût. Rejet ou aversion pour quelque chose ou quelqu'un.

Pour créer des agents virtuels dits empathiques, il faut d'abord définir dans quelles circonstances ils vont devoir l'être et quels sont les mécanismes nécessaires pour pouvoir éprouver de l'empathie. L'empathie est la réponse émotionnelle qu'on éprouve en réaction à l'état émotionnel d'autrui, c'est la capacité à se mettre à la place de l'autre : «je me projette mentalement dans la situation que j'observe». Cette compétence est essentielle dans la régulation des interactions sociales (Frans B.M. de Waal, 2008). La sympathie, d'un autre côté, implique le désir de faire sentir mieux l'autre, ce n'est pas simplement ressentir ce que l'autre personne ressent, mais être motivé à changer son état émotionnel. Pour ce faire, il faut des capacités cognitives supplémentaires (Eisenberg, 1990). Il est important de remarquer que même si ces deux habiletés ne sont pas facilement distinguables, il existe des différences considérables. En effet, on peut être empathique et ne pas ressentir de la sympathie et on peut être sympathique et ne pas ressentir de l'empathie (Bachoud-Lévi et al., 2004).

En prenant compte de ces définitions, on ne peut pas parler d'agents purement empathiques ou sympathiques, puisque ces derniers ne pourraient en aucun cas comprendre l'émotion ressentie par le sujet en face de lui ni la ressentir à leur tour. Ils vont simplement reproduire une émotion en fonction de ce qu'ils identifient afin de signaler de l'intérêt et du soutien. Quand on parle d'agent virtuel empathique, on va donc se concentrer sur la capacité à montrer une bienveillance à l'égard d'autrui, sans forcément comprendre précisément la situation que l'autre vit.

Pour être empathique et reproduire une émotion, il faut d'abord pouvoir identifier et déterminer correctement quelle est l'émotion de notre interlocuteur. Comme nous l'avons vu plus haut, ces six émotions primaires sont reconnues universellement. Puisque l'expression des émotions est à l'origine un moyen de communication et adaptation sociale, il n'est pas étonnant de parler de capacité innée, bien que dépendante de l'influence de l'environnement (Carroll. E. Izard, 1994).

Mais alors comment reconnaître les émotions à travers un ordinateur ? Il existe plusieurs méthodes grâce auxquelles nous pouvons déterminer l'état émotionnel d'une personne : les mesures physiologiques, l'analyse de la parole, l'analyse des mouvements corporels, etc. Du côté

de l'analyse visuelle des expressions des émotions, nous avons le Facial action coding system (FACS), développé par Paul Ekman et Wallace Friesen en 1978. Cet outil sert à coder les expressions faciales en regroupant différents mouvements (contraction ou détente) d'un ou plusieurs muscles du visage en unités d'action. Ainsi, chaque émotion correspond à l'activation plus ou moins intense d'unités d'action spécifiques.

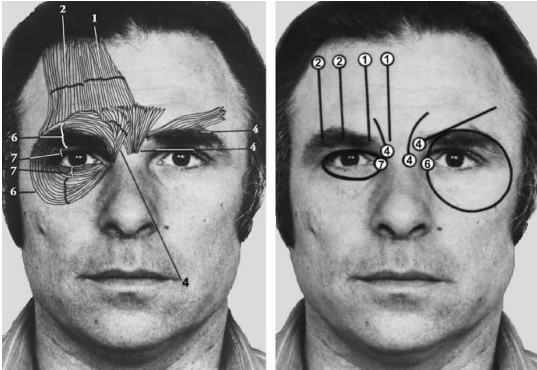
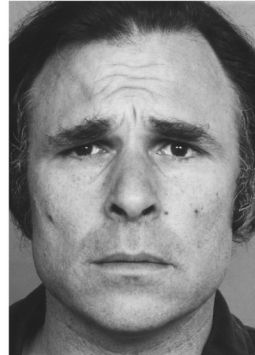
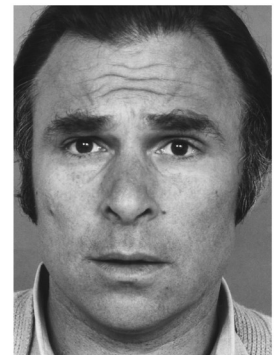


Figure 1: Facial Action Coding System (Ekman & Friesen, 1978)



**UA 1+4**



**UA 1+2+4**

Figure 2 : Facial Action Coding System (Ekman & Friesen, 1978)

Un agent virtuel est un programme logiciel qui peut interagir avec les humains. Grâce à l'intelligence artificielle et l'apprentissage machine, il peut mener une conversation ou assister la personne en fonction de la situation et des besoins. Leur introduction dans la vie quotidienne apporte un certain nombre d'avantages : un service automatisé personnalisé. Même si les agents virtuels ou les avatars ne remplacent pas pleinement l'interaction humain-humain, elle apporte une aide non négligeable dans plusieurs secteurs, soit dans l'aide à l'apprentissage ou l'enseignement, dans le commerce, ou encore dans les grandes entreprises. Les sciences affectives computationnelles cherchent aussi à les utiliser pour aider les personnes avec des troubles cognitifs ou affectifs.

Puisque ces agents virtuels sont destinés à nous faciliter la vie et nous soutenir dans des tâches spécifiques, cela signifie que le temps passé en leur présence sera de plus en plus important. C'est pour cette raison qu'il est primordial de rendre cet échange le plus proche de ce qui serait un échange avec un autre humain.

Plusieurs études montrent que pour que l'interaction soit perçue comme «sociale», il est primordial que l'agent virtuel soit présent visuellement et pas seulement auditivement (A. L. Baylor, 2009). En effet, on a plus l'impression d'interagir avec une vraie présence si on est capable de la voir physiquement. Pour ce qu'il en est de l'empathie, communiquer avec des émotions permet de créer un vrai lien entre l'avatar et la personne. Le fait de mimer

automatiquement le comportement de l'interlocuteur humain (mimicry) suscite des émotions positives dans la communication humaine (T. Numata, 2020), alors le fait de reproduire les mêmes émotions (surtout les émotions positives), rend la réponse de l'agent virtuel plus naturelle et donc à anthropomorphiser ce dernier.

Dans ce travail je lie les émotions du sujet humain à la réaction de l'agent virtuel. C'est à dire qu'il faut d'abord reconnaître l'émotion du sujet grâce à l'analyse de ses expressions faciales récupérées par la webcam pour ensuite l'envoyer au programme qui gère les mouvements faciaux de l'avatar pour qu'il reproduise cette même émotion. L'émotion de l'avatar sera une moyenne des émotions identifiées par le programme et non une mimique simultanée et automatique, afin de rendre son comportement plus empathique et pas donner un simple effet de miroir.

## 2. State-of-the-art

### 2.1 Outils actuels de reconnaissance des émotions

La reconnaissance des émotions étant un sujet au cœur de la recherche des sciences affectives computationnelles, les technologies dites FER («facial expression recognition») ne cessent d'augmenter et de se perfectionner. Il existe donc plusieurs logiciels capables à ce jour de reconnaître les émotions en se basant sur différentes modalités : on peut inférer les émotions d'un sujet à travers sa voix, grâce aux outils comme Empath (Smartmedical Corp), qui analyse les propriétés physiques de la voix et classe les émotions en utilisant différents algorithmes de machine learning. Voici quelques exemples de technologies :

- **Électroencéphalographie (EEG) :** technique électrophysiologique non invasive d'enregistrement de l'activité électrique provenant du cerveau humain grâce à des électrodes (St. Louis, E.K. et al., 2016). Pour déterminer l'émotion ressentie, il faut analyser l'activation et l'amplitude des cinq fréquences émises par différentes parties du cerveau qui sont liées à divers états émotionnels et qui donnent un indice sur la valence et l'arousal de l'émotion (Gurwinder Singh et al., 2020).
- **Électrocardiographie (ECG) :** méthode non invasive d'interprétation de l'activité électrique du cœur en temps réel, l'activité cardiaque étant liée au système central humain. L'ECG est utile pour analyser les activités corporelles mais aussi pour la reconnaissance des émotions. Il faut placer trois capteurs principaux sur le bras gauche, bras droit et sur la jambe gauche afin de mesurer quatre paramètres différents (P, PR, QRS complexe et QT/Qt) et analyser leur amplitude et durée pour déterminer l'état émotionnel (Andrius Dziedzickis et al., 2020).
- **Speech :** l'état émotionnel engendre une série de modifications physiologiques, via le système sympathique et parasympathique, qui vont à leur tour modifier les caractéristiques du tractus vocal et ainsi produire des modifications de l'expression vocale (D. Grandjean et T. Baenziger, 2019). Cette technique permet l'extraction de caractéristiques acoustiques de la voix telles que la hauteur, l'étendue de la hauteur (pitch range), l'intensité (loudness), le tempo et la précision de l'articulation afin de classer les émotions grâce à des modèles complexes de classification (GMM, HMM) (M. E. Ayadi, 2011).

D'autres outils, plus intéressants pour notre problématique, se basent plutôt sur les mouvements faciaux pour déterminer l'émotion. Ils utilisent le « deep learning », qui est basé sur les algorithmes de type réseaux neuronaux (NNs). Le « neural network » copie la structure du cortex

cérébral en créant des neurones artificiels qui vont fonctionner comme les neurones de notre cerveau. Ces derniers sont répartis en trois couches (surtout pour le deep learning) : couche « input », couche « hidden » et couche « output ». Chacune de ces couches modifie la valeur de l'input et essaie de la transformer en output cible. On peut avoir des réseaux pour la classification (apprentissage supervisé) ou pour le clustering (apprentissage non supervisé). D'abord on a le logiciel «Face Reader» de la société Noldus. qui analyse les expressions faciales des sujets, fournies par des live, des vidéo ou des images fixes, pour déterminer leurs émotions. D'un autre côté on a la société iMotions, qui intègre les moteurs de codage automatiques AFFDEX, appartenant au groupe Affectiva, et Realeyes. Pour inférer les émotions, ce logiciel utilise vingt unités d'actions et sept émotions : joie, colère, peur, dégoût, mépris, tristesse et surprise. Il prend en compte aussi des repères faciaux et des indices comportementaux tels que l'orientation de la tête et l'attention. iMotions travaille sur un output de probabilités, l'émotion inférée par le sujet est celle qui a la probabilité la plus élevée.

Ces deux logiciels sont très performants mais avec une licence privée et donc pas adaptés à la recherche d'un travail de bachelor. Je vais passer en revue les autres possibilités pour résoudre ce problème mais cette fois sous licence OpenSource. Un des projets les plus connus dans le domaine est celui de « OpenFace » créé par le chercheur Tadas Baltrusaitis de l'Université de Cambridge en 2017. Son travail permet la création d'applications interactives pour l'analyse du comportement facial, en estimant la position de la tête, la direction du regard et l'action des différentes unités d'actions. de reconnaître l'unité d'action faciale et d'estimer le regard. OpenFace a des très bonnes performances en temps réel, il fonctionne à partir d'une webcam sans nécessité de matériel supplémentaire.

D'autre part, plusieurs librairies sont disponibles en accès libre. Celle que j'ai choisie pour ce travail est la librairie python « Deepface », créée par l'ingénieur Sefik Ilkin Serengil en 2019. Elle sert d'outil de reconnaissance faciale et permet d'analyser des attributs tels que l'âge, le sexe, l'émotion et l'ethnie. Pour ce faire, ce framework mélange et intègre plusieurs technologies disponibles dans le domaine, tels que les modèles : Face, Google FaceNet, OpenFace, Facebook DeepFace, DeepID et plusieurs autres.. Cette combinaison d'outils performants dépassent même la précision de décision des humains quant à la détection des émotions.



## 2.2 Outils actuels pour la création d'un avatar

Depuis les débuts des personnages animés dans les années 1920, l'industrie de l'animation n'a cessé d'évoluer exponentiellement. Omniprésente dans le cinéma, elle intéresse depuis des décennies les chercheurs dans différents domaines : l'éducation, la santé, le marketing, les sciences biomédicales et tant d'autres.

Beaucoup de recherches étudient les interactions entre agents virtuels et humains, mais cet échange n'est pas souvent optimal, encore moins naturel. Au contraire, il est loin d'être considéré comme une interaction sociale. La difficulté concerne la réaction trop automatique et forcée de l'avatar. Je vais donc me concentrer sur le lien entre l'identification de l'émotion ressentie par le sujet et la réponse de l'agent virtuelle afin de la rendre plus authentique et naturelle. D'un part il faut que l'émotion identifiée soit conforme à la réalité et d'autre part que la réaction de l'avatar. C'est à dire l'émotion exprimée à son tour, soit spontanée mais dans un délai raisonnable, comme dans une interaction humain-humain et pas totalement simultanée, ce qui pourrait ressembler plus à une mimique et donc rendre l'échange totalement artificiel.

Dans l'actualité, l'outil permettant la création d'animations et personnages de fiction 3D le plus utilisé dans le cadre artistique et créatif est « Blender ». C'est un logiciel gratuit et Open Source, créé par la Blender Foundation en 2002. Étant un outil très complet, son utilisation peut être très complexe. Dans les logiciels opensource disponibles, il existe OpenFACS et SmartBody. OpenFACS est un outil open source d'animation de visage 3D, utilisant les unités d'action du FACS créé en 2019 par V. Cuculo et A. D'Amelio. Il permet de simuler des expressions faciales plutôt réalistes, dynamiques et en temps réel grâce à la manipulation d'unités d'action spécifiques. D'un autre côté, SmartBody est une plate-forme d'animation de personnages développée à l'USC Institute for Creative Technologies. Il permet de travailler avec le Behavioral Markup Language (BML) qui transforme les descriptions de comportement en animations en temps réel.

### 3. Méthode

#### 3.1 Architecture générale

Afin d'isoler les principales étapes à réaliser pour trouver une solution à ce problème, j'ai décidé de le séparer en trois parties : l'outil capable d'identifier les émotions du sujet à travers la webcam, celui capable de contrôler les expressions faciales de l'avatar, et en dernier l'intégration de ces programmes en un seul. Ainsi, chaque partie sera consacrée uniquement à la résolution d'une tâche précise, qui une fois accomplie, pourra se lier aux autres parties pour communiquer entre-elles et être dépendantes.

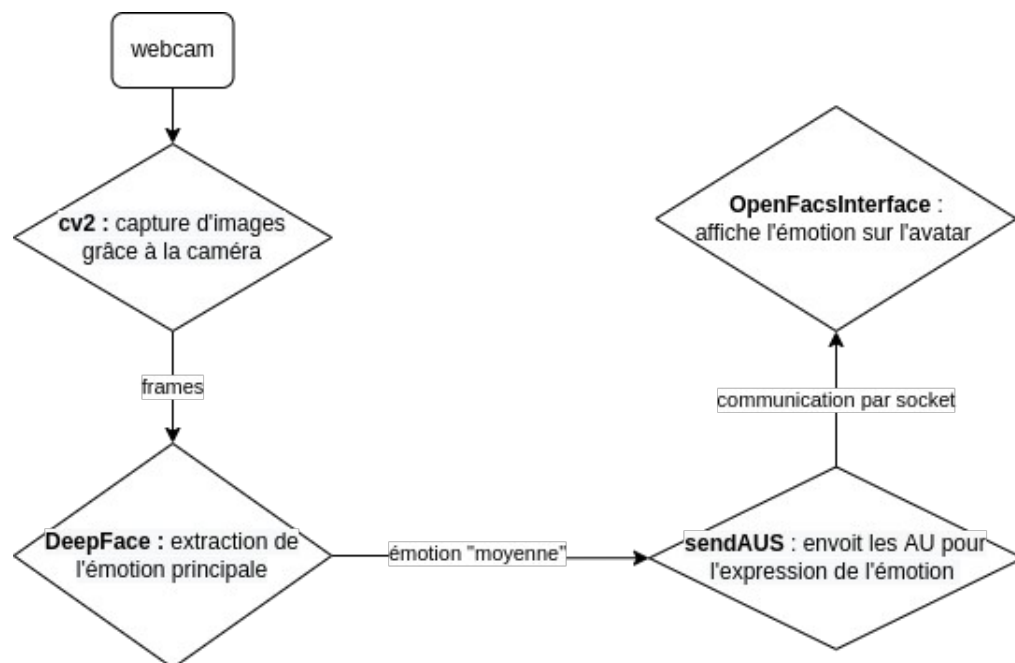


Figure 1: Schéma de l'architecture générale

#### 3.2 Implémentation

Le langage de programmation utilisé pour réaliser ce projet est python. Je vais me servir des bibliothèques unittest, socket, cv2 et deepface afin de traiter les données et de faire communiquer mon programme et l'interface de OpenFACS. Le script suit l'organisation du schéma au point 3.1, d'abord il a fallu que le programme accède à la webcam du sujet, ensuite que le live vidéo soit traité comme une suite d'images pour que la bibliothèque deepface puisse extraire l'émotion dominante en analysant chaque image. Une fois l'émotion déterminée, il faut l'envoyer au programme qui s'occupe d'afficher les expressions sur l'avatar, pour ce faire, la fonction sendAUS va utiliser des sockets afin de communiquer avec l'interface de OpenFACS.

### 3.2.1 Détecteur d'émotions

Afin d'analyser le live à travers la webcam et décider quelle est l'émotion ressentie par le sujet, j'ai dû créer un programme python. J'ai utilisé «cv2», la dernière version de la librairie OpenCV, qui contient plusieurs algorithmes de computer vision et machine learning. Elle permet de détecter et reconnaître des visages, d'identifier des objets, d'inférer des actions humaines à partir de vidéos, de suivre des mouvements de caméra, etc. Ce qui est réellement intéressant pour mon cas, c'est qu'elle permet de sampler une vidéo en temps réel pour la convertir en images appelées «frames», pour qu'ensuite je puisse réaliser des analyses sur celles-ci. Le programme va afficher l'image de la webcam en live, et dessiner un rectangle rouge autour du visage à analyser, pour éviter d'avoir deux personnes dans le cadre et se concentrer sur les émotions d'une seule personne.

C'est à partir de ce moment là que l'utilisation de «DeepFace» devient essentielle. En effet, cette librairie va nous permettre d'analyser chaque frame afin de détecter différentes caractéristiques du visage. L'attribut qui nous intéresse est celui de l'émotion. La fonction «analyze» rend la probabilité que chacune des six émotions se réalise, il faut alors que le programme retourne l'émotion qui a la plus grande probabilité, appelée «dominant\_emotion». C'est ainsi que, à chaque fois que le programme détecte une émotion, il va l'afficher au dessus du rectangle rouge pour montrer au sujet quelle est l'émotion associée à son expression faciale.

### 3.2.2 Agent virtuel

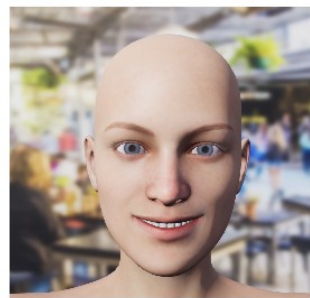
Dans un premier temps, j'ai commencé la réalisation de l'avatar avec Blender, en modélisant un buste humanoïde. Cependant, après des efforts considérables, j'ai dû abandonner cette approche puisqu'il aurait fallu une expertise et des connaissances importantes dans l'utilisation du logiciel afin de modéliser les unités d'action et les muscles. Par ces raisons, j'ai décidé donc d'utiliser l'outil « OpenFACS », tout à fait performant et adapté à mes objectifs.

À ce jour, plusieurs articles scientifiques traitent de l'activation des unités d'action dans l'expression des émotions. Il y a en effet un consensus particulièrement important quant aux unités d'action qu'il faut activer pour quelles émotions, ce qui diverge parmi les auteurs c'est plutôt le niveau d'activation. j'ai établi des valeurs spécifiques pour l'intensité des unités d'actions. En effet, pour chaque émotion des unités d'action sont activées simultanément afin de reproduire les mouvements faciaux représentant l'émotion en question.

Afin de trouver des niveaux d'activation satisfaisants pour l'expression des émotions, je me suis basée sur différents articles scientifiques : P. Ekman, W. Friesen (1976), K. Karthick et al. (2013)

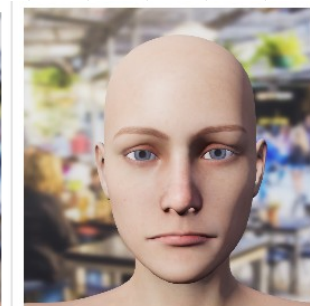
et Y. Zhang et al. (2014). Voici les unités d'action recommandées pour chaque émotion, ainsi que les outputs de ces activations sur l'avatar :

Joie	AU1	AU2	AU4	AU5	AU6	AU7	AU9	AU10	AU12	AU14	AU15	AU17	AU20	AU23	AU25	AU26	AU28	AU45
P. Ekman, W. Friesen (1976)																		
K. Karthick et al. (2013)																		
Y. Zhang et al. (2014)																		
OpenFACS					5.0				3.0						5.0			
Mon avatar					5.0				2.5						5.0			

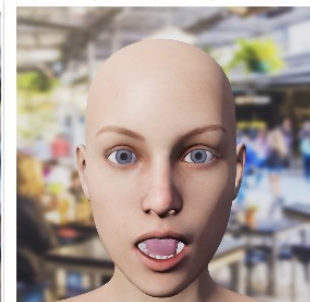


Mon avatar

Tristesse	AU1	AU2	AU4	AU5	AU6	AU7	AU9	AU10	AU12	AU14	AU15	AU17	AU20	AU23	AU25	AU26	AU28	AU45
P. Ekman, W. Friesen (1976)																		
K. Karthick et al. (2013)																		
Y. Zhang et al. (2014)																		
OpenFACS	5.0		5.0								5.0	3.0						
Mon avatar	5.0		5.0			2.0					4.5	2.0						



Mon avatar

[illegible]

## Mon avatar



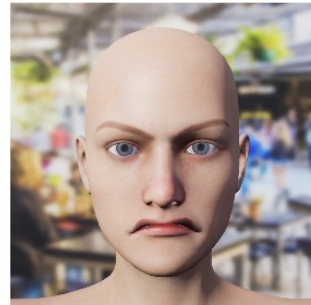
Dégoût	AU1	AU2	AU4	AU5	AU6	AU7	AU9	AU10	AU12	AU14	AU15	AU17	AU20	AU23	AU25	AU26	AU28	AU45
P. Ekman, W. Friesen (1976)																		
K. Karthick et al. (2013)																		
Y. Zhang et al. (2014)																		
OpenFACS			5.0				2.0	5.0			5.0	5.0						
Mon avatar			5.0		1.3	2.0	2.5	5.0			4.0	2.0						



Intensité maximale



Intensité moyenne

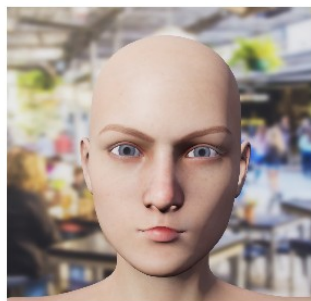


OpenFACS



Mon avatar

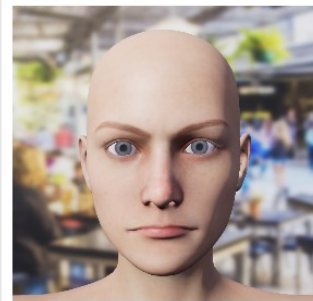
Colère	AU1	AU2	AU4	AU5	AU6	AU7	AU9	AU10	AU12	AU14	AU15	AU17	AU20	AU23	AU25	AU26	AU28	AU45
P. Ekman, W. Friesen (1976)																		
K. Karthick et al. (2013)																		
Y. Zhang et al. (2014)																		
OpenFACS			5.0	3.0							3.0	3.0						
Mon avatar			5.0	3.0		4.5	4.0				3.0	3.0	3.0	2.5				



Intensité maximale



Intensité moyenne

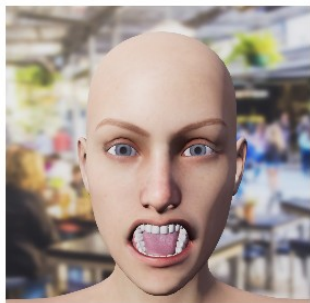


OpenFACS

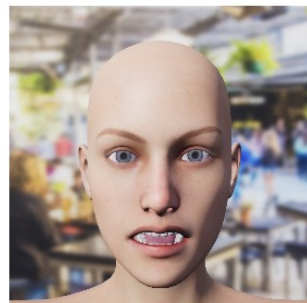


Mon avatar

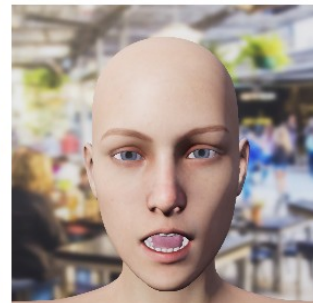
Peur	AU1	AU2	AU4	AU5	AU6	AU7	AU9	AU10	AU12	AU14	AU15	AU17	AU20	AU23	AU25	AU26	AU28	AU45
P. Ekman, W. Friesen (1976)																		
K. Karthick et al. (2013)																		
Y. Zhang et al. (2014)																		
OpenFACS	3.0		3.0			3.0							3.0	3.0	3.0	3.0		
Mon avatar	5.0	2.7	5.0	4.0		4.5							4.5		4.0	1.0		



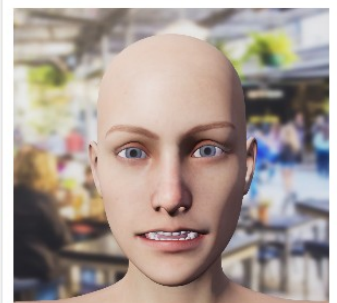
Intensité maximale



Intensité moyenne



OpenFACS



Mon avatar

Ayant essayé plusieurs combinaisons d'unités d'action différentes, je me suis décidé pour celles qui semblaient pas exagérées ou «trop» artificielles, tâche très complexe quand il s'agit de représenter des émotions sur un agent virtuel. Comme l'avatar d'OpenFACS ne possède pas tous les muscles nécessaires pour représenter l'intégralité des unités d'actions, j'ai du moyenner certaines unités d'action et en ajouter d'autres, absentes dans les descriptions des articles vus plus haut, comme par exemple l'AU 9, 15 et 20 pour l'émotion de colère.



Figure 2: Expression "neutre"

Comme on peut voir sur la figure 2 l'expression «neutre» représente l'expression où aucune unité d'action n'est activée. Comme nous pouvons le voir, les unités 28 («lip suck») et 45 (clignement des yeux) ne sont utilisées dans aucune des émotions car ces mouvements ne contribuaient pas à une bonne imitation de l'émotion.

OpenFACS se compose de deux parties, la première qui contient l'avatar, qui gère tous les mouvements (côté animation 3D). La deuxième consiste d'un code python «OpenFACS.py» qui contient des fonctions permettant d'envoyer, sous forme de fichier json, les valeurs d'unités d'action qui vont être activées sur le visage de l'avatar. Donc chaque fichier json envoyé avec la fonction «sendAUS» contient une émotion et une vitesse. L'émotion est déterminée par liste contenant 18 valeurs (voir tableau ci-dessus) et la vitesse est fixée à 0.5. Cette valeur a été choisie pour que la réaction semble plus naturelle et éviter l'effet de mimicry direct.

### 3.2.3 Programme final

Le programme final va intégrer les deux parties expliquées plus haut. Il va donc utiliser les bibliothèques DeepFace, cv2, OpenFACS et plusieurs autres pour le traitement des données. Il contient un dictionnaire avec les six émotions appelé «expressions», le code permettant d'ouvrir la webcam de l'utilisateur et de capturer des images à partir du live pour pouvoir les analyser. La fonction «analyze» va rendre la liste des émotions détectées dans l'image, ensuite la variable «emotion» va contenir l'émotion la plus probable.

```

while True:
    ret, frame = cap.read() #reads an image from video
    try:
        result = DeepFace.analyze(frame, actions=['emotion'])
    except:
        continue
    emotion = result['dominant_emotion']

```

Figure 3: Partie du code pour détecter l'émotion

Afin d'atténuer les fortes variances dans la détection de l'émotion cible et «adoucir» la réaction de l'avatar, j'ai introduit une partie de code pour créer une liste où vont être stockées les premières cinq émotions et qui va prendre l'émotion la plus fréquente pour finalement l'envoyer à l'avatar. Une fois l'émotion envoyée, la liste va se vider pour se remplir avec les cinq prochaines émotions et ainsi de suite. Cet intervalle de cinq émotions a été choisi étant celui qui donnait les meilleurs résultats, après avoir testé plusieurs autres (10, 7 et 3). En effet, cette taille a été choisie car elle permet de rendre la transition entre les émotions plus naturelle et moins saccadée.

Le critère a été de retenir un nombre d'émotions suffisant pour pouvoir exprimer celle qui revient le plus souvent dans un intervalle qui reste proche de la simultanéité, et ainsi éviter un délai de réponse trop grand. Grâce à cette étape on évite la mimique pure et on exprime une réaction adaptée au sujet.

```

# collect 5 emotions to smooth the reaction
if (len(liste_emotions)==5):
    counts = Counter(liste_emotions)
    target = max(counts, key=counts.get)
    print(counts, target)

    #send the emotion detected to the face
    AU = expressions[target]
    #on peut changer la vitesse
    sendAUS(AU, 0.5)

    #vider la liste
    liste_emotions = []

```

Figure 4: Partie du code permettant d'éviter les fortes variations d'émotion

### **3.3 Testing**

Pour tester le programme, j'ai dû d'abord essayer différentes combinaisons d'expressions (activation des unités d'action). Je devais lancer d'un part le programme d'interface de l'avatar et d'autre part celui pour contrôler les unités d'actions et enfin choisir celles qui correspondaient le mieux. Pour m'assurer que les émotions étaient bien déterminées par deepface, j'ai dû créer une fonction pour les imprimer sur le terminal afin que je puisse voir que mon émotion exprimée était bien celle déterminée par le programme. En effet, les émotions de colère, joie, et tristesse sont reconnues assez facilement, ce qui n'est pas forcément le cas pour les émotions de peur, surprise et dégoût, comme on va le voir plus loin dans les résultats.

### **3.4 Méthode de validation**

Afin de valider la précision de ce programme, je l'ai testé sur 5 jeunes universitaires :3 femmes et 2 hommes âgés entre 21 et 27 ans. La procédure était de les installer dans un lieu calme face à l'ordinateur. D'abord ils ont du répondre à la première question du questionnaire (voir annexe A) : « Sur une échelle de 1 à 10, à quel point trouvez-vous ces émotions réalistes ? ». Ensuite, une fois le programme lancé, je leur ai demandé d'exprimer une à une les six émotions suivantes : joie, colère, peur, surprise, dégoût et tristesse. Avec des délai d'environ 20 secondes entre chaque émotion afin de retrouver une expression neutre et ne pas influencer l'expression suivante et d'avoir le temps de répondre à la questions 2 du questionnaire : « Sur une échelle de 0 à 10, ressentez-vous de l'empathie provenant de l'avatar en réaction à votre émotion ? ». La dernière question est :« Sur une échelle de 0 à 10, à quel point considérez-vous cette expérience comme une interaction sociale ? ». Tout au long de la procédure, les participants ne voyaient que la fenêtre montrant l'avatar, et pas le fenêtre de la webcam.



## 4. Résultats

### 4.1 Outcomes

Pour illustrer les outcomes obtenus j'ai lancé le programme qui affiche l'avatar et le programme principal qui gère la caméra et envoi de l'émotion. Ainsi on peut voir l'émotion détectée par l'analyse de DeepFace, la liste des cinq émotions et celle choisie pour être envoyée (la plus fréquente) et la réaction de l'agent virtuel.



Figure 5: Output pour l'expression "neutre"

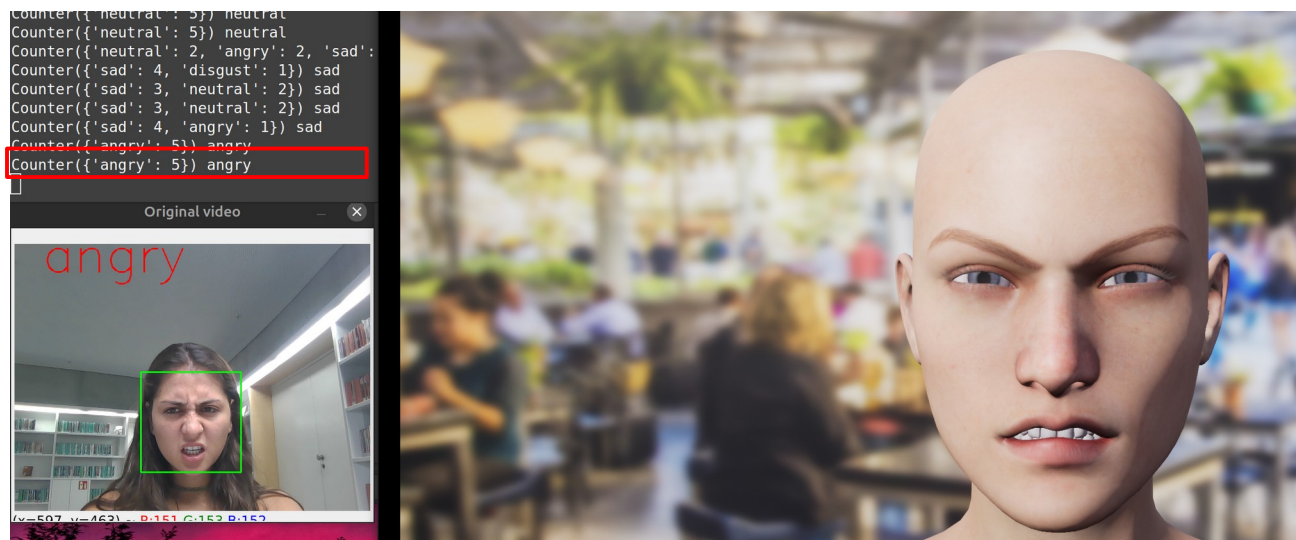


Figure 6: Output pour l'émotion "colère"



Figure 7: Output pour l'émotion "heureux"

## 4.2 Performance du logiciel

Pour mesurer la performance du logiciel je l'ai exécuté sur un ordinateur ThinkPad X1 Carbon Gen 7 de la marque Lenovo sous le système d'exploitation Linux Mint et la version de kernel 5.4.0-120-generic . Avec un CPU Intel(R) Core(TM) i7-8565U CPU @ 1.80GHz et une mémoire RAM de 16GB 2,1333MHz. Pour mesurer précisément les ressources nécessaires j'ai lancé le programme d'interface de l'avatar en parallèle de celui qui ouvre la webcam et envoi l'émotion à travers un socket. Voici la capture du résultat de la commande «htop» qui montre les ressources utilisées par les processus en exécution, entre autres.

PID	USER	PRI	NI	VIRT	RES	SHR	S	CPU%	MEM%	TIME+	Command
44765	brandyx	20	0	4651M	471M	208M	R	325.	3.0	22:57.39	/bin/python3 /home/brandyx/Desktop/BA/OpenFACS/openFACS-master/python/final.py
44806	brandyx	20	0	4651M	471M	208M	S	40.3	3.0	2:37.60	/bin/python3 /home/brandyx/Desktop/BA/OpenFACS/openFACS-master/python/final.py
44804	brandyx	20	0	4651M	471M	208M	R	35.7	3.0	2:37.22	/bin/python3 /home/brandyx/Desktop/BA/OpenFACS/openFACS-master/python/final.py
44805	brandyx	20	0	4651M	471M	208M	R	35.1	3.0	2:39.05	/bin/python3 /home/brandyx/Desktop/BA/OpenFACS/openFACS-master/python/final.py
44807	brandyx	20	0	4651M	471M	208M	R	32.5	3.0	2:36.39	/bin/python3 /home/brandyx/Desktop/BA/OpenFACS/openFACS-master/python/final.py
44803	brandyx	20	0	4651M	471M	208M	R	32.5	3.0	2:38.17	/bin/python3 /home/brandyx/Desktop/BA/OpenFACS/openFACS-master/python/final.py
44808	brandyx	20	0	4651M	471M	208M	R	30.5	3.0	2:31.81	/bin/python3 /home/brandyx/Desktop/BA/OpenFACS/openFACS-master/python/final.py
44657	brandyx	20	0	2326M	373M	170M	S	29.2	2.4	1:34.34	/home/brandyx/Desktop/BA/OpenFACS/LinuxNoEditor/ActionUnitsFace/Binaries/Linux/ActionUnitsFace
44809	brandyx	20	0	4651M	471M	208M	R	26.6	3.0	2:25.53	/bin/python3 /home/brandyx/Desktop/BA/OpenFACS/openFACS-master/python/final.py
44703	brandyx	20	0	2326M	373M	170M	S	13.0	2.4	0:38.49	/home/brandyx/Desktop/BA/OpenFACS/LinuxNoEditor/ActionUnitsFace/Binaries/Linux/ActionUnitsFace

Figure 8: Output du terminal après la commande "htop"

Comme on peut le voir sur la figure 8, l'exécution du code «final.py» nécessite entre 30 et 40% du CPU et l'interface de l'avatar «ActionUnitsFace.sh» environ 30 %, ce qui montre à quel point le programme est gourmand et explique le ralentissement important de l'ordinateur.

Pour mesurer le temps de réaction de l'avatar, j'ai chronométré le temps nécessaire entre l'expression de l'émotion par le sujet et l'expression de l'émotion par l'avatar. Après 15 mesures de toutes les émotions aléatoirement, la moyenne était de 3.4283 secondes avec un écart-type de  $\sigma = 0.29002$ .

### 4.3 Validation

Pour discuter des données récoltées durant la passation des participants j'ai effectuée des analyses exploratoires simples. Comme on peut le voir sur la figure 9, il existe une importante différence de vraisemblance entre les différentes émotions. En effet, les expressions de colère, de tristesse, de joie et de surprise de l'avatar sont considérées comme assez réalistes. Pourtant l'expression de dégoût et de peur sont à améliorer, leur reconnaissance étant moins évidente.

Émotion	Moyenne	Écart-type
Colère	8,6	1,34
Joie	8,4	1,52
Tristesse	9	0,71
Surprise	8	1,41
Peur	5,6	0,55
Dégoût	6,2	1,30

Figure 9: Analyses question 2

Émotion	Moyenne	Écart-type
Colère	7,2	1,92
Joie	8	0,71
Tristesse	8,4	1,34
Surprise	6,6	0,89
Peur	7,4	1,34
Dégoût	5,2	1,92

Figure 10: Analyses question 3

La figure 10 montre la moyenne et l'écart-type pour les réponses à la question : « Sur une échelle de 0 à 10, ressentez-vous de l'empathie provenant de l'avatar en réaction à votre émotion ? ». Dans ce cas ce sont les expressions de joie et de tristesse de l'avatar qui traduisent une plus forte empathie. Ce résultat peut être la conséquence que ces deux émotions sont une des plus facilement reconnaissables, et puisqu'elles possèdent des valences très marquées (négative pour la tristesse et positive pour la joie) elles transmettent plus aisément un sentiment d'empathie.

La dernière question était : « Sur une échelle de 0 à 10, à quel point considérez-vous cette expérience comme une interaction sociale ? ». Les réponses à cette question étaient moins satisfaisantes que les deux autres, avec une moyenne de  $\gamma = 6.6$  et un écart-type de  $\sigma = 1.4$ . On peut expliquer cette faible moyenne par la subjectivité de l'expérience vécue ainsi que de l'appréhension aux agents virtuels.

## 5. Discussion

La réalisation de projets dans ce genre est très enrichissante, d'abord d'un point de vue théorique quand il s'agit de chercher dans la littérature un soutien scientifique aux concepts étudiés : les émotions, l'empathie, l'interaction humain-machine. De plus, d'un point de vue pratique, il s'agit de mettre en commun plusieurs technologies afin de résoudre ce problème assez complexe qui est la création d'un agent virtuel empathique. Ce sujet d'étude, étant relativement récent, il est encore difficile d'arriver à un consensus sur les technologies à utiliser et les méthodes de recherche à suivre.

Pour ma part, pour ce qui est de la reconnaissance des émotions, on pourrait envisager d'utiliser d'autres méthodes, comme par exemple l'ECG, vue plus haut. Il faut tout de même rappeler que l'analyse des features d'un visage humain par les fonction de la librairie deepface reste très efficace. Une amélioration à prévoir peut être celle de l'apparence de l'avatar. En effet, il serait intéressant de le rendre plus humain et réaliste en lui mettant des cheveux et un habit. Le but ultime étant de le faire interagir avec la personne par la parole, ce qui pourrait augmenter considérablement la sensation d'interaction sociale. Mon ambition serait d'utiliser cet avatar en tant qu'assistant virtuel. Il aiderait à organiser la journée de l'utilisateur, en fonction de son planning mais aussi en fonction de son état émotionnel. Il pourrait par exemple, décider suite à une détection de tristesse, de lire une phrase motivationnelle ou bien de demander quel est la source de ce sentiment négatif. Tout ceci en montrant de l'empathie, pour atteindre un niveau de lien social plus élevé qu'avec un simple avatar.



## Références

- Baylor, A. L. (2009). Promoting motivation with virtual agents and avatars: role of visual presence and appearance. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1535), 3559-3565.
- Berthoz, A., & Jorland, G. (2004). *L'empathie*. Odile Jacob.
- Britton, J. W., Frey, L. C., Hopp, J. L., Korb, P., Koubeissi, M. Z., Lievens, W. E., ... & St Louis, E. K. (2016). Electroencephalography (EEG): an introductory text and atlas of normal and abnormal findings in adults, children, and infants.
- Darwin, C. (1890). *L'expression des émotions*. Schleicher.
- De Waal FB. Putting the altruism back into altruism: the evolution of empathy. *Annu Rev Psychol*. 2008;59:279-300. doi: 10.1146/annurev.psych.59.103006.093625. PMID: 17550343.
- Dzedzickis, A., Kaklauskas, A., & Bucinskas, V. (2020). Human emotion recognition: Review of sensors and methods. *Sensors*, 20(3), 592.
- Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of personality and social psychology*, 17(2).
- Ekman, P., & Friesen, W. V. (1978). Facial action coding system. *Environmental Psychology & Nonverbal Behavior*.
- Ekman, P. (1999). Basic emotions. *Handbook of cognition and emotion*, 98(45-60), 16.
- El Ayadi, M., Kamel, M. S., & Karray, F. (2011). Survey on speech emotion recognition: Features, classification schemes, and databases. *Pattern recognition*, 44(3), 572-587.
- Grandjean, D. & Baenziger, T. (2019). Chapitre 4. Expression vocale des émotions. Dans : David Sander éd., *Traité de psychologie des émotions* (pp. 109-155). Paris: Dunod. <https://doi.org/10.3917/dunod.sande.2019.01.0109>
- Iolanda Leite, André Pereira, Samuel Mascarenhas, Carlos Martinho, Rui Prada, Ana Paiva, The influence of empathy in human-robot relations, *International Journal of Human-Computer Studies*, Volume 71, Issue 3, 2013, Pages 250-260, ISSN 1071-5819, <https://doi.org/10.1016/j.ijhcs.2012.09.005>
- Izard, C. E. (1994). Innate and universal facial expressions: evidence from developmental and cross-cultural research.
- Jenay M. Beer, Cory-Ann Smarr, Arthur D. Fisk, Wendy A. Rogers, Younger and older users' recognition of virtual agent facial expressions, *International Journal of Human-Computer Studies*, Volume 75, 2015, Pages 1-20, ISSN 1071-5819, <https://doi.org/10.1016/j.ijhcs.2014.11.005>
- Karthick, K., & Jasmine, J. (2013). Survey of Advanced Facial Feature Tracking and Facial Expression Recognition.
- Moataz El Ayadi, Mohamed S. Kamel, Fakhri Karray, Survey on speech emotion recognition: Features, classification schemes, and databases, *Pattern Recognition*, Volume 44, Issue 3, 2011, Pages 572-587, ISSN 0031-3203, <https://doi.org/10.1016/j.patcog.2010.09.020>
- Numata, T., Sato, H., Asa, Y., Koike, T., Miyata, K., Nakagawa, E., ... & Sadato, N. (2020). Achieving affective human-virtual agent communication by enabling virtual agents to imitate positive expressions. *Scientific reports*, 10(1), 1-11.

Paper Cuculo, V., & D'Amelio, A. (2019, August). OpenFACS: an open source FACS-based 3D face animation system. In *International Conference on Image and Graphics* (pp. 232-242). Springer, Cham. [https://rd.springer.com/chapter/10.1007/978-3-030-34110-7\\_20](https://rd.springer.com/chapter/10.1007/978-3-030-34110-7_20)

Singh, G., Verma, K., Sharma, N., Kumar, A., & Mantri, A. (2020, December). Emotion recognition using deep convolutional neural network on temporal representations of physiological signals. In *2020 IEEE International Conference on Machine Learning and Applied Network Technologies (ICMLANT)* (pp. 1-6). IEEE.

S. I. Serengil and A. Ozpinar, "LightFace: A Hybrid Deep Face Recognition Framework," *2020 Innovations in Intelligent Systems and Applications Conference (ASYU)*, 2020, pp. 1-5, doi: 10.1109/ASYU50717.2020.9259802

**Annexe**  
**Questionnaire**

1. Sur une échelle de 0 à 10, à quel point trouvez-vous ces émotions réalistes ?

**Colère**



0	1	2	3	4	5	6	7	8	9	10
---	---	---	---	---	---	---	---	---	---	----

**Joie**



0	1	2	3	4	5	6	7	8	9	10
---	---	---	---	---	---	---	---	---	---	----

**Tristesse**



0	1	2	3	4	5	6	7	8	9	10
---	---	---	---	---	---	---	---	---	---	----

**Surprise**



0	1	2	3	4	5	6	7	8	9	10
---	---	---	---	---	---	---	---	---	---	----

**Peur**



0	1	2	3	4	5	6	7	8	9	10
---	---	---	---	---	---	---	---	---	---	----

**Dégoût**



0	1	2	3	4	5	6	7	8	9	10
---	---	---	---	---	---	---	---	---	---	----

2. Sur une échelle de 0 à 10, ressentez-vous de l'empathie provenant de l'avatar en réaction à votre émotion ?

Colère :

0	1	2	3	4	5	6	7	8	9	10
---	---	---	---	---	---	---	---	---	---	----

Joie :

0	1	2	3	4	5	6	7	8	9	10
---	---	---	---	---	---	---	---	---	---	----

Tristesse :

0	1	2	3	4	5	6	7	8	9	10
---	---	---	---	---	---	---	---	---	---	----

Peur :

0	1	2	3	4	5	6	7	8	9	10
---	---	---	---	---	---	---	---	---	---	----

Surprise :

0	1	2	3	4	5	6	7	8	9	10
---	---	---	---	---	---	---	---	---	---	----

Dégoût :

0	1	2	3	4	5	6	7	8	9	10
---	---	---	---	---	---	---	---	---	---	----

3. Sur une échelle de 0 à 10, à quel point considérez-vous cette expérience comme une interaction sociale ?

0	1	2	3	4	5	6	7	8	9	10
---	---	---	---	---	---	---	---	---	---	----