Personalizing facial expressions by exploring emotional mental prototypes

Sen Yan

DOCTORAT BRETAGNE LOIRE / MATHSTIC

CentraleSupélec

THÈSE DE DOCTORAT DE

CENTRALESUPÉLEC

ÉCOLE DOCTORALE N° 601
Mathématiques et Sciences et Technologies de l'Information et de la Communication
Spécialité : *Signal, Image, Vision*

Par
Sen YAN

Personalizing facial expressions

By exploring emotional mental prototypes

Thèse présentée et soutenue à CentraleSupélec Rennes, le 29/09/2023
Unité de recherche : IETR
Thèse N° : 2023CSUP0002

Rapporteurs avant soutenance :

Catherine ACHARD	Professeure, Sorbonne Université
Mohamed DAOUDI	Professeur, IMT Nord Europe

Composition du Jury :

Président :	Catherine PELACHAUD	Directrice de recherche CNRS, Sorbonne Université
Examinateurs :	Lionel PREVOST	Directeur de recherche ESIEA, ESIEA
	Catherine PELACHAUD	Directrice de recherche CNRS, Sorbonne Université
	Fan YANG	Professeure, Université de Bourgogne
Directeur de thèse :	Renaud SEGUIER	Professeur, CentraleSupélec, Rennes
Co-encadrante de thèse :	Catherine SOLADIE	Maître de Conférence, CentraleSupélec, Rennes

Acknowledgement

I would like to express my sincere gratitude and appreciation to the following individuals and organizations who have contributed to the successful completion of my thesis.

First and foremost, I am deeply grateful to my supervisors: Prof. Renaud SEGUIER and Assoc. Prof. Catherine SOLADIE, for their guidance, expertise, and unwavering support throughout the research process. Special thanks to Assoc. Prof. Catherine SOLADIE for the patient guidance and support. Quote from the previous Ph.D. student, Jingting Li: "She is the best!" [71]

I would like to extend my heartfelt thanks to the members of my thesis committee, Prof. Catherine ACHARD, Prof. Mohamed DAOUDI, Prof. Lionel PREVOST, Prof. Catherine PELACHAUD, and Prof. Fan YANG for their invaluable feedback, constructive criticism, and expert guidance. Their expertise and contributions have significantly enhanced the quality of my research.

I would also like to thank Prof. Jean-Julien AUCOUTURIER, Assoc Prof. Simon LEGLAIVE, and Director Bernard JOUGA to share their advice, insights, and expertise.

I am thankful to my lab mates: Jingting LI, Siwei WANG, Adrien LLAVE, Nam Duong DUONG, Xiang DAI, Zhigang ZHANG, Samir SADOK, Guénolé FICHE, Rafael ACCACIO NOGUEIRA, Sarra ZAIED, Corentin GUEZENOC, and Eloïse BERSON who have provided me with continuous support, insightful discussions, and encouragement during my research. Their assistance and camaraderie have made this journey more enjoyable and fulfilling. Special thanks to Adrien LLAVE for his support during the Covid-19 pandemic.

I am grateful to the university staff: Karine BERNARD, Myriam ANDRIEUX, Catherine PIEDNOIR, Grégory CADEAU, Fréderic JAN, and Nocolas MAUMY for their unfailing support and assistance.

I am deeply indebted to the over 30 participants at CentraleSupélec who participated in my experimental study. Their cooperation and willingness to share their insights and perception have been invaluable. Their contributions have added depth and richness to my research findings.

I would also like to acknowledge Randstad France and CentraleSupélec for providing

my thesis project.

Last but not least, I would like to express my heartfelt appreciation to my family: my parents and my wife. Their love and belief in my abilities have been my constant source of motivation.

To all those who have played a part, big or small, in shaping this thesis, I extend my sincerest gratitude. Thank you all.

RÉSUMÉ FRANÇAIS

Chapitre 1: Introduction

Les expressions faciales (EF) sont une forme essentielle de communication non verbale que les humains peuvent utiliser pour transmettre des informations sociales [24, 111]. En tant que forme d'expression émotionnelle, les expressions faciales ont évolué pour devenir des signaux permettant de communiquer aux autres des informations importantes sur l'état psychologique d'un individu. Les travaux de Mehrabian montrent que 55 % des messages relatifs aux sentiments et aux attitudes se trouvent dans l'expression faciale, 7 % dans les mots prononcés et le reste dans la manière dont les mots sont prononcés [81]. Physiquement, les expressions faciales sont le résultat du mouvement des différents muscles du visage. Différentes combinaisons de mouvements musculaires peuvent créer une large gamme d'expressions faciales, chaque expression faciale étant associée à l'état interne (mental) de l'individu. Par exemple, une personne qui se renfrogne peut être perçue comme étant en colère ou agressive, tandis qu'une personne qui sourit peut être perçue comme étant amicale ou avenante. Ces perceptions peuvent influencer la manière dont les gens interagissent les uns avec les autres et peuvent avoir un impact significatif sur les interactions sociales.

Arrière-plan

La motivation. Ces dernières années, les tâches basées sur la FE ont suscité un intérêt croissant. Il en existe deux principales: la reconnaissance des expressions faciales (FER) et la manipulation des expressions faciales (FEM). La première consiste à lire les expressions faciales pour interpréter les émotions humaines (haut de la figure). Inversement, la seconde vise à afficher l'état interne d'un individu en modifiant ses expressions faciales (bas de la Fig. 1). Ces tâches basées sur la FE sont omniprésentes et ont un large éventail d'applications, telles que les applications FER dans les robots sociaux [72], la détection des mensonges [38, 34] et les applications FEM dans les médias sociaux [117], les jeux vidéo [6, 104]. Dans cette thèse, j'ai également travaillé dans un contexte d'application.

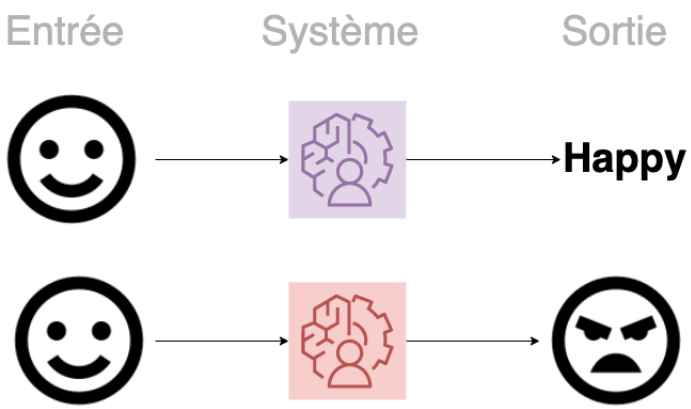


Figure 1 – Illustration des tâches basées sur la FE. En haut: reconnaissance d'expressions faciales (FER). En bas: manipulation de l'expression faciale (FEM)

Chaire Randstad. Ce travail est rattaché à une chaire industrielle (entre Randstad et CentraleSupelec). En tant qu'entreprise multinationale de ressources humaines, Randstad est spécialisée dans les services de ressources humaines pour les emplois temporaires et permanents. Généralement, le contexte de l'application concerne «l'Intelligence Artificielle pour le Recrutement».

Ma recherche se concentre sur l'un des objectifs de ce projet. Avec l'essor du recrutement en ligne, les entretiens vidéo deviennent de plus en plus courants et efficaces. Au cours du processus de recrutement, les recruteurs évaluent les compétences professionnelles du candidat (appelées «hard skills») et se concentrent également sur les compétences comportementales (appelées «soft skills»). Dans le cadre du processus de candidature en ligne, il est généralement demandé au candidat de télécharger une vidéo de présentation personnelle. Actuellement, le candidat télécharge directement une vidéo d'auto-présentation à l'intention des recruteurs (haut de la Fig. 2). Cependant, Randstad a besoin d'un nouveau cas d'utilisation. Un système, tel qu'un coach numérique, peut fournir au candidat une vidéo modifiée en traitant les expressions faciales du candidat à sa guise si le candidat autorise l'interaction avec un tel coach numérique. Ainsi, le candidat peut s'inspirer de cette vidéo pour affiner son comportement (compétences non techniques). Comme le montre le bas de la Fig. 2, ce système (le coach numérique) renvoie au candidat une vidéo modifiée en fonction de la vidéo originale téléchargée par le candidat, par exemple en remplaçant le visage original par un visage sûr de lui ou par l'expression que le candidat souhaite voir perçue. Le candidat peut remarquer les différences entre la vidéo modifiée et la vidéo originale. Par conséquent, avec l'aide du coach numérique, le candidat peut

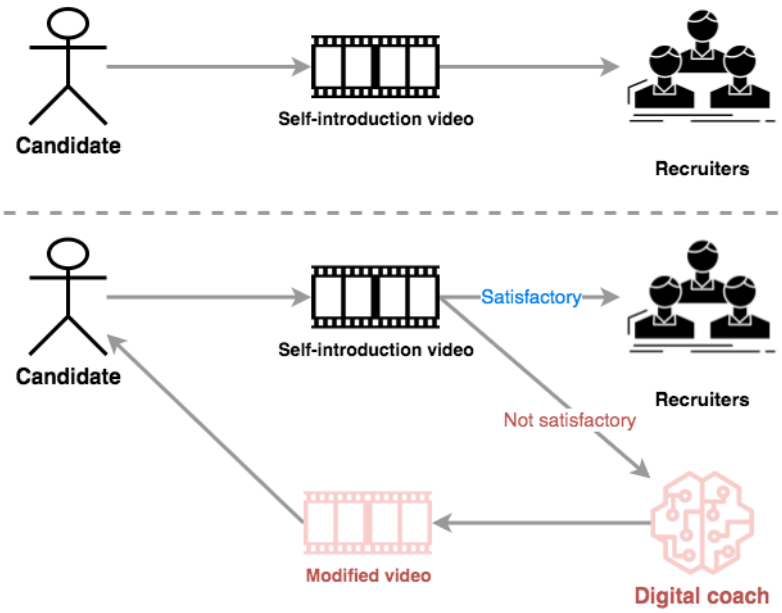


Figure 2 – Cas d'utilisation d'une application en ligne dans le cadre de cette thèse. **En haut**: actuellement, le candidat télécharge directement une vidéo de présentation personnelle à l'intention des recruteurs. **En bas**: le candidat peut s'entraîner avec l'aide du système (un coach numérique), puis télécharger la vidéo finale à sa satisfaction. Ce coach numérique effectue une tâche typique de manipulation d'expression faciale (FEM). De plus, cette tâche FEM suit la volonté du candidat pour répondre à ses besoins (vidéo satisfaisante) plutôt que de manipuler de manière automatique.

s'entraîner et affiner son comportement (en termes d'expression faciale), puis enregistrer et recharger la vidéo à sa satisfaction. La vidéo modifiée (avec des artefacts) par le système (le coach numérique) ne sera pas envoyée aux recruteurs.

En résumé, ce système exécute une tâche FEM typique: la manipulation des expressions faciales depuis l'expression faciale source (par exemple, le visage neutre) jusqu'à l'expression faciale cible qui répond au besoin du candidat (par exemple, le visage confiant ou l'expression que le candidat veut être vu). En outre, un tel système peut également être utilisé dans d'autres domaines critiques, comme les applications cliniques [40, 51, 10]. Par exemple, pour traiter les patients souffrant de troubles émotionnels, un miroir numérique peut automatiquement transformer l'expression faciale actuelle en une expression positive, telle qu'un visage confiant.

Les trois défis. L'objectif industriel susmentionné révèle les principaux défis que posent les technologies actuelles basées sur la FE.

1. **Diversité**. Comme Darwin l'a noté pour la première fois dans le livre de *The*

Expression of the Emotions in Man and Animals, les émotions sont universelles à travers les cultures et les espèces [24]. Cette hypothèse d'universalité a été soutenue par le psychologue Paul Ekman, qui a soutenu que les expressions faciales de 6 émotions (joie, tristesse, colère, dégoût, peur, surprise, dites émotions de base) ne sont pas déterminées par la culture, mais sont universelles à travers les cultures humaines [36, 39, 35]. Cependant, l'universalité des prototypes d'Ekman est aujourd'hui remise en question par un nombre croissant de psychologues [97, 63, 5]. Cela indique qu'il peut y avoir plusieurs prototypes pour une même étiquette émotionnelle.

2. **Flexibilité**. Bien que pour une émotion donnée, les prototypes d'expression faciale doivent être multiples et différents d'une culture à l'autre, on ne sait pas quel prototype d'expression faciale peut répondre aux besoins de l'utilisateur. En effet, il devrait exister une application basée sur l'expression faciale qui puisse être personnalisée. C'est-à-dire, par exemple, dans le contexte de l'application Randstad, pour une émotion donnée, l'expression faciale devrait être générée pour répondre au besoin de l'utilisateur, c'est-à-dire à la satisfaction d'un candidat à la recherche d'un emploi (l'utilisateur), telle que l'expression qu'il souhaite être perçu (le besoin).
3. **Exhaustivité**. Comme le montre la recherche en psychologie, il existe plus de 4000 étiquettes d'émotions [106]. En raison de la limitation des données labélisées importantes et fiables pour la formation, la plupart des technologies basées sur la FE ne peuvent traiter que les émotions de base d'Ekman. Les labels d'émotions non basiques, tels que la confiance en soi, ne sont pas disponibles dans les bases de données existantes. Les caractéristiques ou prototypes correspondants sont donc inconnus et ne peuvent pas être appris à partir de la base de données. Outre l'absence de données labellisées importantes, la création d'une telle base de données avec diverses étiquettes d'émotions pose de nombreux problèmes: 1) l'annotation demande beaucoup de temps et de travail et 2) certaines tâches d'étiquetage requièrent des experts formés (par exemple, des codeurs FACS certifiés [37]).

Les cinq exigences. Compte tenu des défis susmentionnés, pour répondre à des domaines plus critiques tels que l'application clinique susmentionnée et l'industrie des services comme Randstad, le système basé sur la FE devrait s'adapter à des exigences plus variées et plus fines.

1. Le système devrait être applicable à toutes les expressions qui ne se limitent pas aux émotions de base. Les expressions faciales prototypiques qui ne sont pas disponibles dans les bases de données d'apprentissage profond existantes devraient également

être prises en compte, comme les émotions complexes ou les attitudes sociales (par exemple, la confiance en soi) ainsi que des expressions plus générales (par exemple, comment voulez-vous être perçu lors de votre entretien d'embauche ?)

2. Le système doit pouvoir être contrôlé par n'importe qui, de manière précise et cohérente, sans qu'il soit nécessaire de faire appel à des experts (par exemple, des codeurs certifiés FACS, des connaissances en informatique affective).
3. Le système doit être capable de personnaliser les expressions faciales pour ses utilisateurs. Par la suite, pour être plus précis, nous appelons l'utilisateur *l'observateur* et nous appelons le personnage/sujet (dont l'expression faciale sera modifiée) *l'acteur*. Bien que dans le contexte de l'application, *les observateurs* et *les acteurs* soient les mêmes personnes, il peut s'agir de personnes différentes (c'est-à-dire d'identités différentes).
4. Le système doit tenir compte de la fatigue de l'utilisateur [58, 7]. L'ensemble du processus doit être efficace afin de minimiser l'effet de la fatigue de l'utilisateur. Nous avons fixé à 15 minutes la durée maximale de fonctionnement du système.
5. Pour une émotion, le système devrait obtenir plusieurs prototypes, ce qui le rendrait plus proche de la réalité.

La solution: les nouvelles méthodes interdisciplinaires combinant l'informatique et la psychologie

Le pipeline traditionnel d'apprentissage profond. La première étape du pipeline traditionnel d'apprentissage profond (la base de données, Fig. 3) a déjà été entravée à trois reprises. **En termes de diversité**, compte tenu de l'argument en psychologie selon lequel les prototypes d'Ekman ne sont pas universels, il devrait y avoir plusieurs prototypes pour une même émotion. Cependant, les bases de données sont généralement biaisées. En particulier, la quantité de chaque étiquette émotionnelle peut être déséquilibrée et les ethnies des sujets (acteurs) sont également déséquilibrées. La base de données biaisée augmente la difficulté d'obtenir des représentations correctes correspondant à chaque ethnie. **En termes de flexibilité**, pour un utilisateur donné (observateur), il est relativement subjectif de savoir quel prototype émotionnel peut répondre à ses besoins. La perception humaine est subjective [107]. Bien que la subjectivité soit bénéfique pour personnaliser les expressions faciales afin de répondre aux besoins de l'observateur, il peut y avoir des divergences entre la perception de l'observateur et le jugement de l'annotateur [83], de

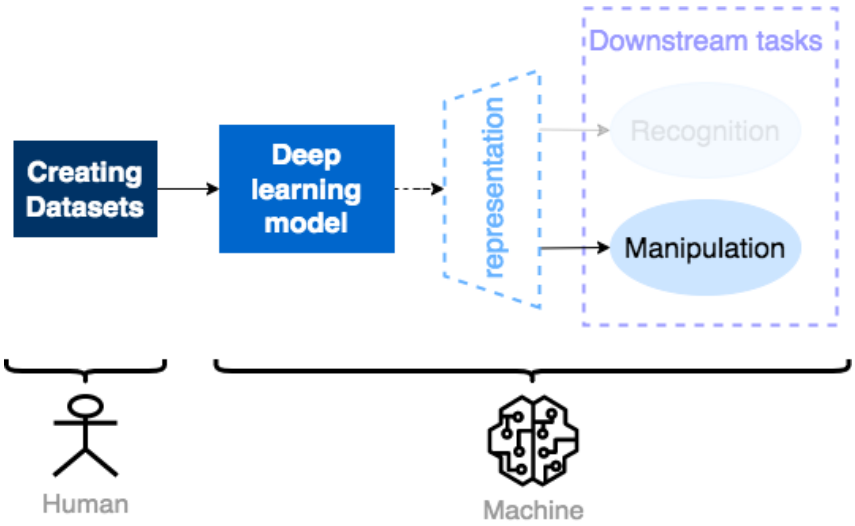


Figure 3 – Pipeline traditionnel d'apprentissage profond traitant les expressions faciales. **1) Base de données.** Création d'une base de données ou utilisation directe de bases de données existantes. **2) Modèle d'apprentissage profond.** Utilisation de données pour entraîner un modèle d'apprentissage profond. **3) Représentation.** Lorsque le modèle d'apprentissage profond est bien entraîné, les caractéristiques peuvent être extraites. **4) Tâches en aval.** Une fois les représentations extraites, elles peuvent être utilisées pour les tâches en aval. La première étape est réalisée par des humains, et les autres sont effectuées par des machines.

sorte que les expressions faciales générées sur la base de l'annotation peuvent ne pas être satisfaisantes pour l'observateur. **En termes d'exhaustivité**, la grande base de données labelisées des émotions non basiques telles que la confiance en soi n'est pas explicitement disponible. La création d'une telle base de données prend toujours beaucoup de temps et de travail, et nécessite également des experts qualifiés. Pour plus d'informations sur la base de données, voir le chapitre 2 du manuscrit.

Pourquoi ne pas penser différemment? Pouvons-nous éviter de créer une base de données aussi complexe et difficile pour obtenir la représentation des émotions? Pouvons-nous établir un nouveau système différent du pipeline d'apprentissage profond traditionnel pour relever les défis de la diversité, de la flexibilité et de l'exhaustivité tout en répondant aux exigences de l'application susmentionnée?

La réponse est oui. Inspirés par le mécanisme d'analyse des expressions faciales par les psychologues, nous proposons dans cette thèse deux approches interdisciplinaires différentes du pipeline d'apprentissage profond traditionnel. Ces approches interdisciplinaires, qui s'appliquent à la tâche FEM, relèvent les défis existants en termes de diversité, de flexibilité

et d'exhaustivité et répondent finalement à toutes les exigences susmentionnées.

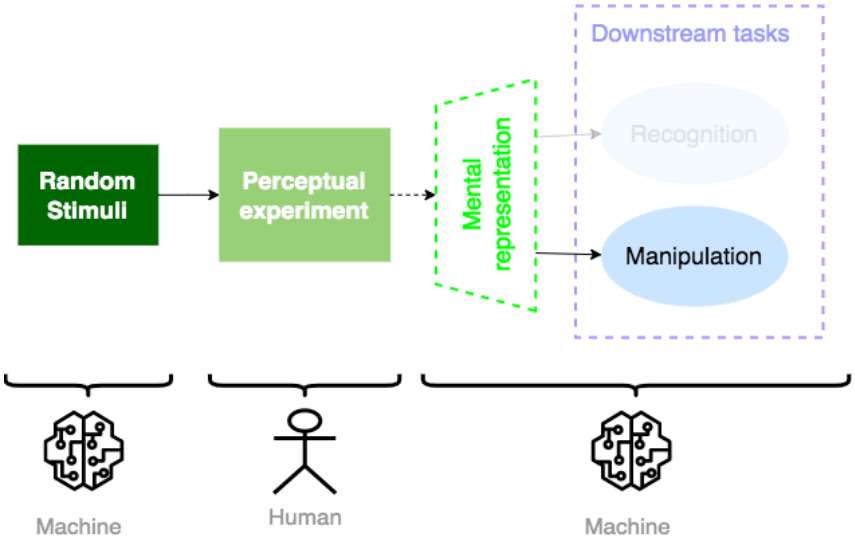


Figure 4 – **Le premier pipeline:** une nouvelle approche interdisciplinaire qui combine la corrélation inverse psychophysique (RevCor) [85, 9] de la psychologie avec les réseaux antagonistes génératifs (GAN) [94] de l'informatique.

Le premier pipeline: Système de rétro-ingénierie profonde mentale. Nous employons un état d'esprit différent de celui du pipeline d'apprentissage profond traditionnel. Nous nous inspirons du processus psychophysique de corrélation inverse (RevCor), généralement utilisé pour l'informatique affective en psychologie [85, 9]. Il s'agit d'une méthode pilotée par les données. RevCor peut être utilisé pour extraire les prototypes mentaux (ou représentations mentales) de ce à quoi une émotion donnée devrait ressembler pour un observateur (ou participant). C'est-à-dire que les prototypes mentaux ne sont pas limités aux émotions basiques. Le prototype mental d'une émotion non basique peut également être extrait. **La première exigence peut ainsi être adressée.** Comme le montre la première étape de la Fig. 4, contrairement au pipeline d'apprentissage profond traditionnel, les données créées (ci-après, appelées stimuli) sont générées de manière aléatoire par la machine. Grâce au caractère aléatoire, il est possible d'éviter les biais liés à la base de données traditionnelle, et aucune connaissance experte n'est requise. **C'est-à-dire que la deuxième exigence peut être adressée.** En s'appuyant sur le jugement subjectif de l'observateur dans l'expérience perceptive de RevCor, la sortie de l'expérience perceptive, c'est-à-dire la représentation mentale, contient la représentation qui répond au besoin de l'observateur (par exemple, à quoi doit ressembler un visage sûr de lui). Ainsi, la représentation mentale peut être utilisée pour personnaliser les expressions faciales dans la

tâche FEM en aval. **Cela permet de répondre à la troisième exigence.** Pour plus de détails sur RevCor, voir le chapitre 3 du manuscrit. En combinant la technique de la psychologie avec celle de l'informatique, le premier pipeline répond aux trois premières exigences.

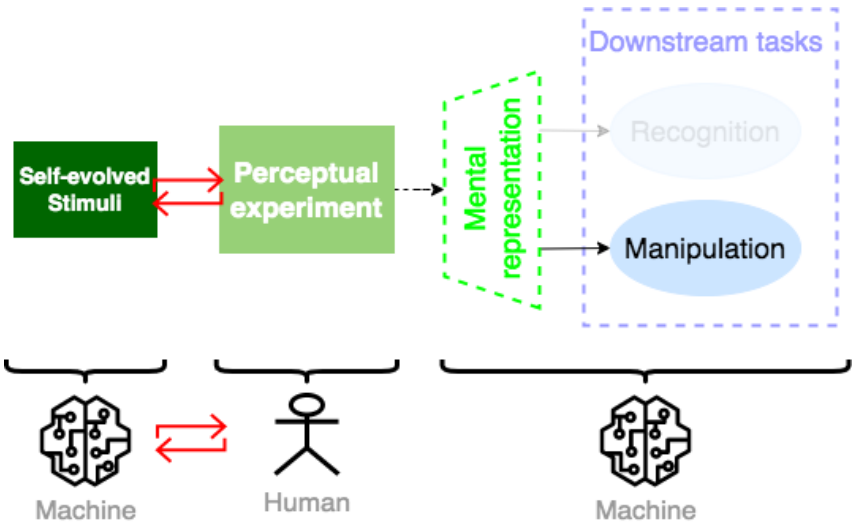


Figure 5 – **Le deuxième pipeline:** en se basant sur le premier pipeline, nous affinons les deux premières étapes en ajoutant une interaction humain-machine pour rendre l'ensemble du processus plus efficace.

Le deuxième pipeline: Algorithme génétique microbien interactif. RevCor présente deux imperfections qui devraient être résolues pour être largement appliquées dans divers scénarios, par exemple le contexte d'application de Randstad. La première imperfection correspond à la quatrième exigence. Chaque observateur doit effectuer un grand nombre d'essais générés de manière aléatoire et l'ensemble de l'expérience est conçu par des experts. En effet, les expériences qui prennent du temps fatiguent l'utilisateur, et la dépendance à l'égard des connaissances des experts (l'expertise) entrave également l'extension à d'autres domaines. La deuxième imperfection correspond à la cinquième exigence. RevCor repose sur l'hypothèse qu'il existe un, et un seul, prototype mental pour un état affectif qui existe chez un ou un groupe d'observateurs. Cette unicité peut être remise en question. Pour répondre à toutes les exigences, nous avons proposé, sur la base du premier pipeline, une méthode d'optimisation tenant compte de l'interaction homme-machine (voir Fig. 5) afin d'accélérer le pipeline, **ce qui permet de résoudre la quatrième exigence.** En outre, notre méthode d'optimisation peut fournir des solutions multiples, **ce qui répond à la cinquième exigence.**

Contributions

Le premier pipeline (MDR). Nous proposons une nouvelle approche interdisciplinaire: Système de rétro-ingénierie profonde mentale (en anglais: MDR), pour personnaliser les expressions faciales en combinant la corrélation inverse psychophysique (RevCor) [85, 9] de la psychologie avec les réseaux antagonistes génératifs (GAN) [94] de l'informatique.

Différents des GANs typiques qui peuvent manipuler l'expression faciale, combinés à RevCor, ont les avantages suivants.

1. **Exhaustivité.** On peut aborder n'importe quelle émotion ou attitude sociale, car c'est, par nature, un rôle de RevCor. Cela signifie que nous n'avons pas besoin de construire une base de données étiquetées dédiée pour chaque émotion ou attitude sociale. Un FEM contrôlé par attributs de bas niveau peut couvrir une large gamme de mouvements faciaux locaux. Il est possible de manipuler diverses expressions faciales. **Cela répond à la première exigence.**
2. **Expertise-gratuite.** Aucune connaissance experte en informatique affective ou codeur certifié FACS n'est nécessaire pour créer le prototype personnalisé, car notre approche ne nécessite que la perception de l'observateur (c'est-à-dire le jugement subjectif) plutôt que l'expertise de l'observateur. **Cela répond à la deuxième deuxième exigence.**
3. **Flexibilité.** Nous extrayons la représentation mentale de l'expression faciale désirée qui peut répondre aux besoins de l'observateur. Le prototype mental est **personnalisé** en fonction de la perception de l'observateur et ne correspond pas spécifiquement à un prototype universel appelé ainsi. **Ainsi, ce pipeline répond à la troisième exigence.**

Inversement, à la différence des approches RevCor récentes, l'utilisation des techniques FEM (telles que le GAN) permet de manipuler des visages réels (les images en 2D) plutôt que des avatars virtuels, ce qui offre une manière plus facile et intuitive d'éditer les expressions faciales. De plus, nous utilisons deux fois le même outil (GAN): une fois pour extraire le prototype mental avec RevCor et une autre fois pour la manipulation. Cela permet de garantir que la manipulation est cohérente avec la représentation mentale de l'observateur.

Enfin, pour améliorer la définition des prototypes d'expression faciale, nous introduisons **le concept d'unités d'action dominantes et complémentaires** pour décrire

précisément les prototypes d'expression faciale.

Le deuxième pipeline (IMGA). Nous avons créé une approche interdisciplinaire efficace appelée IMGA (en anglais) pour Algorithme Génétique Microbien Interactif **qui répond à toutes les exigences mentionnées**.

L'originalité de notre approche est que, sur la base du premier pipeline, nous avons intégré le processus de corrélation inversée psychophysique (RevCor) dans l'algorithme génétique interactif (IGA). Cette intégration hérite non seulement des forces du premier pipeline, mais résout également les inconvénients du premier pipeline.

Les forces héritées: Exhaustivité (lié à la première exigence), Expertise-graduite (lié à la deuxième exigence), et Flexibilité (lié à la troisième exigence).

Les inconvénients résolus. Cette méthode devient efficace et peut fournir des solutions diverses.

1. **Efficacité: grâce à la boucle de rétroaction en ligne.** Contrairement à la méthode RevCor traditionnelle qui génère de nombreuses tentatives de manière aléatoire, dans notre approche, en se basant sur la rétroaction de l'observateur, les tentatives mises à jour automatiquement peuvent contenir des informations plus précieuses (plus proches des prototypes mentaux des observateurs). **Cela est lié à la quatrième exigence.**

2. **Efficacité: par accélération.** La façon de générer les tentatives pour les calculs des prototypes mentaux est intelligente. De plus, nous adoptons l'algorithme génétique microbien (MGA) [55] comme module GA au sein d'IGA pour accélérer encore la convergence du système. En effet, le GA avec un mécanisme élitiste (comme le MGA) peut converger plus rapidement que celui sans mécanisme élitiste [67]. **Cela est également lié à la quatrième exigence.**

3. **Diversité.** En bénéficiant de l'algorithme génétique, pour une émotion, cette méthode peut fournir plusieurs prototypes mentaux à chaque observateur. Cela signifie plusieurs solutions (c'est-à-dire des prototypes) même pour un observateur. **Cela est lié à la cinquième exigence.**

Une autre originalité de notre approche concerne la création d'une telle pipeline. Contrairement à l'algorithme génétique traditionnel qui nécessite d'acquérir les valeurs de fitness de tous les individus, nous avons ajouté un module d'évaluation de population qui évalue la qualité de l'ensemble de la population avec un nombre limité d'essais. De plus, nous avons ajouté un automate à contraintes à trois états pour augmenter progressivement le nombre

d'unités d'action faciale (AUs) [37] activées pour chaque visage et déterminer la fin du processus.

Chapitre 2: État de l'art

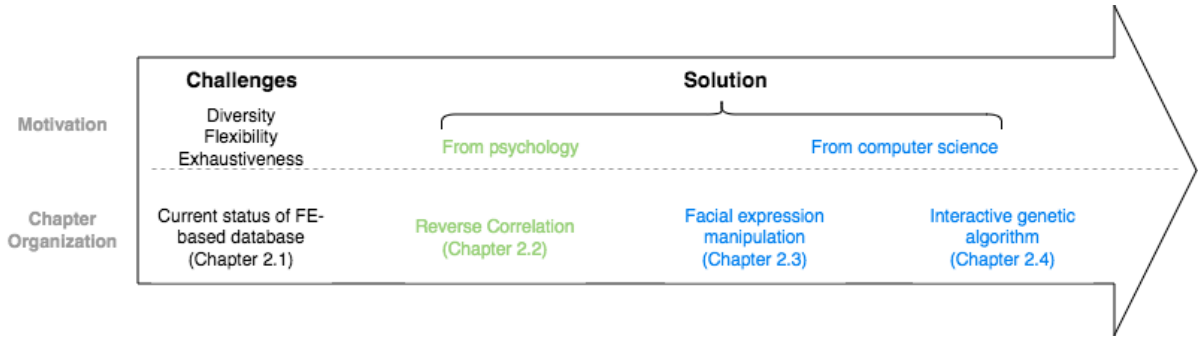


Figure 6 – Organisation du Chapter 2.

Dans ce chapitre, comme le montre la Fig. 6, nous commençons par présenter l'état actuel des bases de données basées sur la FE (dans la Section 2.1). Nous discutons des bases de données **en termes de diversité, de flexibilité et d'exhaustivité**. Pour répondre aux trois défis, nous proposons une autre façon de penser en combinant une idée issue de la psychologie et des idées issues de l'informatique. Ainsi, nous décrivons les travaux relatifs à la manipulation des expressions faciales (FEM), et dans la Section 2.3, nous présentons le processus psychophysique de corrélation inverse (RevCor). Ces deux techniques sont liées au premier pipeline (MDR). Toutefois, la combinaison de ces techniques issues de deux disciplines ne peut répondre qu'aux trois premières exigences (voir la Section 1.1.3). L'optimisation est nécessaire pour répondre à toutes les exigences. Dans la section 2.4, nous présentons l'algorithme génétique interactif (IGA) qui peut optimiser un système avec la participation de l'homme. La technique de l'IGA est liée à notre deuxième pipeline (IMGA).

Les bases de données basées sur les expressions faciales

Nous listons 14 bases de données et les résumons en fonction de 7 caractéristiques: la taille de la base de données (Taille), le nombre de sujets (Sujets), l'état de la collection (Collection), l'âge moyen (Âge), le ratio de genre (Genre), le ratio d'ethnicité (Ethnicité),

Table 1 – Le lien entre les caractéristiques des bases de données et les trois défis. Car. = Caractéristique; D = Diversité; F = Flexibilité; E = Exhaustivité; Collect = état de la collection.

Car.	Taille	Sujets	Collect	Âge	Genre	Ethnicité	Annotateur	Labels
Défi	D	D	D	D	D	D	F	D, F, E

le type d'annotateurs (Annotateur), et le type de labels (Labels). Le lien entre les caractéristiques des bases de données et les trois défis (diversité, flexibilité, et exhaustivité) est montré dans le Tableau 1.

En conclusion, nous résumons deux inconvénients des bases de données existantes. Le premier inconvénient est que la plupart des bases de données sont biaisées. Le deuxième inconvénient est que la création d'une base de données de FE étiquetés repose toujours sur des connaissances d'experts. C'est pourquoi la création d'une base de données prend toujours beaucoup de temps et nécessite une main-d'œuvre importante.

En ce qui concerne les perspectives, la démographie des sujets (âge, genre et ethnicité) doit être **diverse et équilibrée**. Pour les étiquettes de visage, il est possible d'utiliser **plusieurs labels** afin de réduire la subjectivité du jugement humain et d'utiliser **des attributs objectifs** de bas niveau (tels que les unités d'action, AU [37]) pour décrire un visage.

La manupulation de l'expression faciale

La manipulation d'expressions faciales (FEM) est une tâche typique en aval basée sur la FE. Sur la base d'une image de visage en entrée et de paramètres de contrôle (c'est-à-dire d'instructions de manipulation), la FEM peut générer une nouvelle image de visage en sortie. En fonction des paramètres de contrôle de la manipulation, la FEM peut être divisée en deux catégories: la manipulation à haut niveau d'attributs et la manipulation à bas niveau d'attributs.

En général, les attributs qui peuvent être manipulés sont limités par la disponibilité d'attributs de haut niveau dans la base de données d'apprentissage. Bien que le contrôle des attributs de haut niveau puisse générer une variété de visages très réalistes, ces approches ne peuvent en principe pas générer d'expressions faciales arbitraires et fines. **La manipulation d'attributs de bas niveau permet de générer diverses expressions faciales avec un contrôle relativement fin**. En outre, les attributs de bas niveau, tels que les unités d'action, sont relativement **objectifs** (ils utilisent les informations

anatomiques pour décrire les expressions faciales) par rapport aux attributs de haut niveau, tels que les étiquettes d'émotion et les valeurs de valence-arousal.

En ce qui concerne le défi de la diversité et de la flexibilité, la manipulation des attributs de bas niveau est une bonne solution. C'est pourquoi nous avons choisi GANimation [94] pour synthétiser les expressions faciales en contrôlant les unités d'action. **Cependant, en ce qui concerne le défi de l'exhaustivité,** en raison du manque de bases de données d'émotions non basiques, la technique FEM ne peut toujours pas synthétiser des émotions non basiques. Pour atténuer la dépendance à l'égard de la base de données et parvenir à l'expertise-graduite (c'est-à-dire sans avoir besoin de connaissances spécialisées), nous avons mélangé la GANimation avec la corrélation inverse, c'est-à-dire un concept issu de la psychologie.

Le processus psychophysique de corrélation inverse

Dans cette section, nous décrivons les travaux utilisant la corrélation inverse pour le calcul affectif. Le processus peut être divisé en trois étapes: étape 1) la génération de stimuli, étape 2) l'expérience perceptive, et étape 3) le calcul de la représentation mentale. Nous avons détaillé le fonctionnement de chaque étape et présenté les travaux représentatifs. En outre, nous avons comparé ces travaux représentatifs en fonction du type de stimuli, du paradigme, du nombre d'états affectifs, du nombre d'essais nécessaires (pour chaque observateur et pour chaque observateur dans chaque état affectif) et du nombre de prototypes mentaux pouvant être extraits de chaque observateur.

Sur la base de ces comparaisons, nous avons conçu le premier pipeline (Fig. 4) qui est différent du pipeline traditionnel d'apprentissage profond (Fig. 3): étape 1) utilisation d'un outil FEM (c'est-à-dire GANimation [94]) pour générer aléatoirement des stimuli, **il y a une implication de la machine**; étape 2) en adoptant des expériences perceptives (utilisant le paradigme 2-AFC), et en proposant une nouvelle façon de calculer les représentations mentales (unité d'action dominante et unités d'action complémentaires), **il y a une implication de l'homme**; étape 3) utilisation de la représentation mentale pour personnaliser les expressions faciales par le même outil FEM, **il y a une implication de la machine**.

Globalement, le premier pipeline hérite des points forts de RevCor: **Exhaustivité (liée à la première exigence), Expertise-graduite (liée à la deuxième exigence), et Flexibilité (liée à la troisième exigence)**. Cependant, ce pipeline hérite également des inconvénients de RevCor.

— **Efficacité (liée à la quatrième exigence)**: les essais sont générés aléatoirement, donc moins efficaces.
— **Diversité (liée à la cinquième exigence)**: pour un observateur, un seul prototype mental peut être extrait, alors que les prototypes devraient être multiples.

Pour générer des essais pour RevCor de manière efficace et apporter différents prototypes mentaux, nous nous inspirons de l'algorithme génétique interactif (IGA en anglais).

L'algorithme génétique interactif

En intégrant l'AGI, nous affinons le premier pipeline, ce qui nous permet de répondre à toutes les cinq exigences susmentionnées et de relever tous les trois défis. L'AGI étant une variante des algorithmes génétiques, nous présentons brièvement l'algorithme génétique traditionnel avant d'introduire l'AGI utilisé pour le calcul affectif. Ensuite, nous présentons les travaux connexes utilisant l'AGI et proposons le deuxième pipeline (IMGA) en intégrant l'AGI pour optimiser le premier pipeline. Par rapport aux travaux connexes, nous remarquons que notre IMGA est plus intelligent: 1) il utilise une boucle de rétroaction en ligne pour générer des essais de manière efficace, 2) il ajoute un module d'évaluation de la population pour surveiller la convergence du système et déterminer quand le système peut s'arrêter, et 3) il ajoute un automate pour contrôler la FEM.

Dans l'ensemble, le deuxième pipeline répond aux cinq exigences. Les trois premières exigences sont héritées du premier pipeline (c'est-à-dire **Exhaustivité, Expertise-graduite, et Flexibilité**, voir Section 1.1.3). Les deux dernières exigences (c'est-à-dire les inconvénients hérités du premier pipeline: **Efficacité et Diversité**) sont traitées:

— **Efficacité: grâce à la boucle de rétroaction en ligne.** Contrairement à la méthode traditionnelle RevCor qui génère massivement des essais au hasard, dans notre approche, en fonction de la rétroaction de l'observateur, les essais mis à jour automatiquement peuvent contenir des informations plus précieuses (plus proches des prototypes mentaux des observateurs). Ceci est lié à *la quatrième exigence.*
— **Efficacité: par accélération.** La méthode de génération d'essais pour les calculs de prototypes mentaux est intelligente. De plus, nous avons adopté l'algorithme génétique microbien (MGA) [55] comme module GA dans IGA pour accélérer davantage la convergence du système. En effet, le GA avec un mécanisme élitiste (comme MGA) peut converger plus rapidement que celui sans mécanisme élitiste [67]. Ceci est également lié à *la quatrième exigence.*
— **Diversité.** Bénéficiant d'IGA, pour une émotion donnée, ce pipeline peut fournir

plusieurs prototypes mentaux à chaque observateur. Cela signifie plusieurs solutions (c'est-à-dire prototypes) même pour un observateur. Ceci est lié à *la cinquième exigence.*

Chapitre 3 Le premier pipeline: Système de rétro-ingénierie profonde mentale (en anglais: MDR)

Inspirés par le mécanisme d'analyse des expressions faciales par les psychologues, nous proposons le premier pipeline (MDR). Il s'agit d'une nouvelle approche interdisciplinaire qui combine la récente technique d'apprentissage profond de l'informatique, c'est-à-dire le GAN [45], avec la corrélation inverse psychophysique (RevCor), une technique récemment apparue en psychologie. Les informations supplémentaires sur ce pipeline sont disponibles sur https://yansen0508.github.io/emotional-prototype/.

Dans ce chapitre, nous introduisons **la méthodologie** du premier pipeline MDR:

1. comment générer des stimuli;
2. comment concevoir l'expérience perceptive;
3. comment calculer la représentation mentale;
4. comment personnaliser l'expression faciale.

Puis nous détaillons **la partie expérimentation**:

1. le cadre de l'expérimentation;
2. illustrer le calcul sur l'AU dominante et les AUs complémentaires et valider ce concept;
3. présenter les prototypes personnalisés et valider les prototypes.

De plus, nous avons mené **deux évaluations subjectives**:

1. l'évaluation par des observateurs (c'est-à-dire les participants qui ont effectué des expériences perceptives);
2. l'évaluation par des non-observateurs.

Nous avons finalement discuté de **l'efficacité de convergence** de ce processus. Cette discussion est également la motivation pour le deuxième pipeline: optimiser le premier pipeline par améliorer l'efficacité et le diversité.

Chapitre 4 Le deuxième pipeline: Algorithme génétique microbien interactif (en anglais: IMGA)

Dans ce chapitre, nous présentons le deuxième pipeline(IMGA). L'objectif du deuxième pipeline est d'affiner le premier pipeline (c'est-à-dire MDR) répondant ainsi à toutes les exigences. Plus en détail, une telle approche interdisciplinaire intègre le processus de corrélation inverse psychologique (RevCor) dans un algorithme génétique interactif (IGA). Cette approche non seulement hérite des atouts du premier pipeline, mais résout également les inconvénients du premier pipeline. Les informations supplementaires sur ce pipeline sont disponibles sur https://yansen0508.github.io/Interactive-Microbial-Genetic-Algorithm/.

Pour **la méthodologie**, semblable à l'algorithme génétique traditionnel (GA), il s'agit d'un processus itératif qui répète quatre étapes: population (initialisation et mise à jour), sélection, croisement et mutation. L'ensemble du système, notamment l'interaction entre l'humain (observateur) et la machine (système GA), est détaillé par une démonstration vidéo (voir le lien ci-dessus).

Pour **la partie expérimentation**, nous présentons: 1) le détail de l'implémentation, 2) le protocole d'expérimentation, et 3) les résultats illustrant l'évolution de la population et montrant le fonctionnement de cet algorithme.

Ensuite, **nous évaluons quantitativement** les prototypes représentatifs sous deux angles: les unités d'action et les prototypes. Nous présentons également **le processus d'évaluation subjective** dans deux buts: 1) pour valider que nos prototypes représentatifs peuvent refléter les prototypes mentaux des observateurs, et 2) pour comparer subjectivement avec l'état de l'art. Enfin, **nous discutons de l'efficacité** en comparant notre approche avec les travaux connexes.

Chapitre 5: Conclusion et perspective

Dans ce chapitre, nous résumons les défis récents dans les tâches basées sur FE: **Diversité, Flexibilité, et Exhaustivité** et nos solutions avec les contributions correspondantes: **Exhaustivité** liée à la première exigence, **Expertise-graduite** liée à la deuxième exigence, **Flexibilité** liée à la troisième exigence, **Efficacité** liée à la quatrième exigence, et **Diversité** liée à la cinquième exigence.

Nous présentons **les perspectives de notre méthode et de cette thèse sous des aspects plus généraux mais différents**. D'abord, nous présentons les perspectives de

notre approche sous 3 aspects: l'amélioration de notre approche, la recherche en psychologie, et l'application en informatique. Ensuite, nous présentons les perspectives de cette thèse sous des aspects plus généraux mais différents: en termes de construction d'une base de données, en termes d'humain et de machine, et enfin repenser l'approche interdisciplinaire.

Table of Contents

Chapter 1

Introduction

1.1 Background

Facial expressions (FE) are an essential form of nonverbal communication that humans can use to convey social information [24, 111]. As a form of emotional expressions, facial expressions evolved as signals that communicated important information about an individual's psychological state to others. The work of Mehrabian shows that 55% of messages pertaining to feelings and attitudes is in facial expression, 7% of which is in the words that are spoken, the rest of which are the way that the words are said [81]. Physically, facial expressions are the results of the movement of various facial muscles. Different combinations of muscle movements can create a wide range of facial expressions, each facial expression is associated with the individual's internal (mental) state. For example, a person who is scowling may be perceived as angry or aggressive, while a person who is smiling may be perceived as friendly or approachable. These perceptions can influence how people interact with each other, and they can have a significant impact on social interactions.

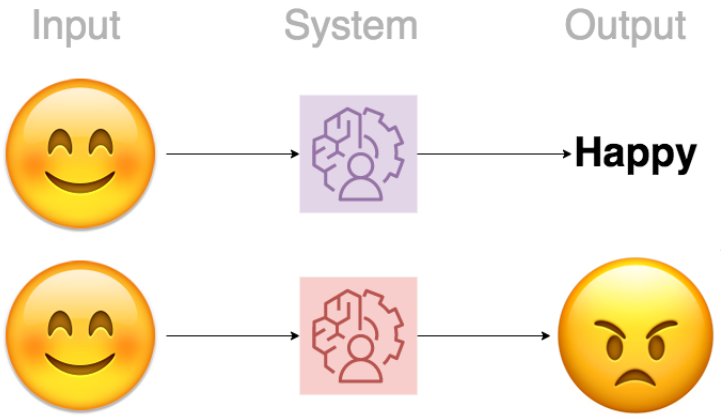


Figure 1.1 – Illustration of FE-based tasks. Top: facial expression recognition (FER). Bottom: facial expression manipulation (FEM)

1.1.1 Motivation

In recent years, there has been a growing interest in FE-based tasks. There are two major FE-based tasks: facial expression recognition (FER) and facial expression manipulation (FEM). The first one is reading facial expressions to interpret human emotions (top of Fig. 1.1). Conversely, the second one aims at displaying an individual's internal state through editing facial expressions (bottom of Fig. 1.1). These FE-based tasks are ubiquitous and have a wide range of applications, such as FER applications in social robots [72], lie detection [38, 34] and FEM applications in social media [117], video games [6, 104]. In this thesis, I also worked in an application context.

Randstad Chair. It is a French Industrial Chair project (fr: Chaire Industrielle) created by Randstad France and CentraleSupélec. As a multinational human resource consulting firm, Randstad specializes in human resource services for temporary and permanent jobs. Generally, the application context is about "Artificial Intelligence for Recruitment".

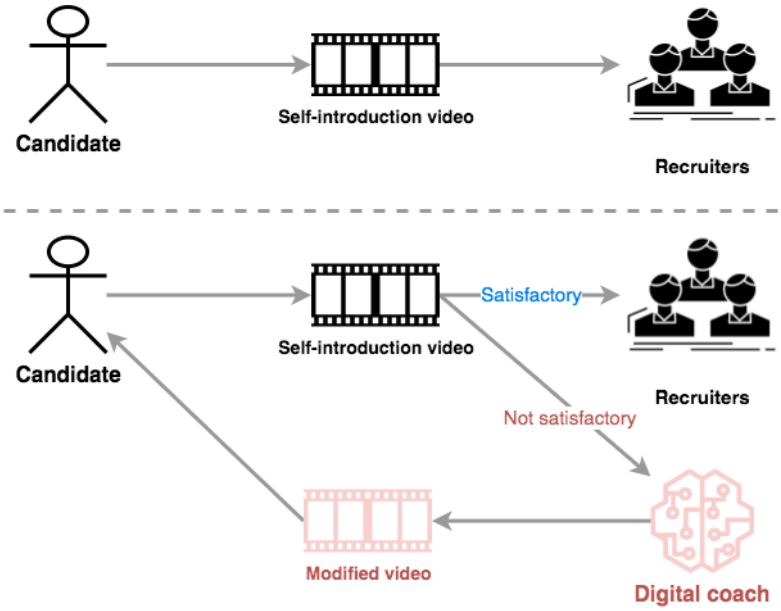


Figure 1.2 – Online application use case involved in this thesis. **Top**: currently, the candidate directly uploads a self-introduction video to the recruiters. **Bottom**: the candidate can get practice with the help of the system (i.e., digital coach) and then upload the final video to his satisfaction. This digital coach performs a typical facial expression manipulation (FEM) task. Moreover, this FEM task follows the candidate's will to meet the need of the candidate (satisfactory video) rather than manipulating in an automatic manner.

My research focuses on one of the objectives of this project. With the rise of online recruiting today, video interviews are becoming more common and efficient. During the recruiting process, recruiters will evaluate the candidate's job-related skills (so-called hard skills) and also focus on behavioral competencies (so-called soft skills). In the online application process, the candidate is usually required to upload a self-introduction video. Currently, the candidate directly uploads a self-introduction video to the recruiters (top of Fig. 1.2). However, a new use case is required by Randstad. A system, such as a digital coach, can provide the candidate with a modified video by processing the candidate's facial expressions to his will if the candidate allows interaction with such a digital coach. Thus, the candidate can get inspired by this video to refine their behavior (soft skills). As shown at the bottom of Fig. 1.2, this system (i.e., digital coach) returns a modified video to the candidate according to the original video uploaded by the candidate, such as modifying the original face with a self-confident face, or the expression that the candidate wants to be perceived. The candidate can notice the differences between the modified video and the original one. Therefore, with the help of the digital coach, the candidate can get practice and refine his behavior (in terms of facial expression) and then record and re-upload the video to his satisfaction. The modified video (with artifacts) by the system (i.e., the digital coach) will not be sent to the recruiters.

In summary, this system performs a typical FEM task: manipulating facial expressions from the source facial expression (e.g., the neutral face) to the target facial expression that meets the need of the candidate (e.g., the self-confident face or the expression that the candidate wants to be seen). Furthermore, such a system can also be used in other critical domains, e.g., clinical application [40, 51, 10]. For instance, in order to treat patients with emotional disorders, a digital mirror can automatically transfer the current facial expression into a positive expression such as a self-confident face.

1.1.2 Existing challenges

The aforementioned industrial objective reveals the existing major challenges of current FE-based technologies.

Diversity. As first noted by Darwin in the book of *The Expression of the Emotions in Man and Animals*, emotions are universal across cultures and species [24]. This universality hypothesis was supported by the psychologist Paul Ekman, who argued that facial expressions of 6 emotions (i.e., happy, sad, angry, disgust, fear, surprise, so-called basic emotions) are not culturally determined, but are universal across human cultures [36, 39,

35]. However, the universality of Ekman's prototypes is now being challenged by a growing number of psychologists [97, 63, 5]. This indicates that there can be multiple prototypes for one emotional label.

Flexibility. Although for a given emotion, facial expression prototypes should be multiple and different across cultures, it is unknown which facial expression prototype can meet the need of the user. Indeed, there should be an FE-based application that can be personalized. That is to say, for instance, in the application context of Randstad, for a given emotion, the facial expression should be generated to meet the need of the user, i.e., to a job-searching candidate's (the user) satisfaction such as the expression that he wants to be perceived (the need).

Exhaustiveness. As research in psychology covers, there are more than 4000 labels of emotions [106]. Due to the limitation of large and reliable labeled data for training, most FE-based technologies can only deal with Ekman's basic emotions. Non-basic emotion labels, such as self-confidence, are unavailable in existing databases. Thus the corresponding features or prototypes are unknown and can not be learned from the database. In addition to the lack of large labeled data, creating such a database with various emotion labels comes with many concerns: 1) time-consuming and labor-intensive for the annotation and 2) requiring trained experts (e.g., certified FACS coders [37]) for some labeling tasks.

1.1.3 Requirements

Considering the aforementioned challenges, to address more critical domains such as the aforementioned clinical application and the service industry like Randstad, the FE-based system should adapt to more various and fine-grained requirements. We summarize the requirements as follows.

1. The system should be applicable to any expressions that are not limited to basic emotions. The prototypical facial expressions that are not available in existing deep-learning databases should also be considered such as complex emotions or social attitudes (e.g., self-confidence) as well as more general expressions (e.g., how do you want to be seen during your job interview?).
2. The system should be controllable by anyone in a precise and consistent manner without the need for expert knowledge (e.g., FACS-certified coders, knowledge in affective computing).
3. The system should be capable of personalizing facial expressions for its users. There-

after, to be more precise, we call the user *the observer* and we call the character/subject (whose facial expression will be changed) *the actor*. Although in the application context, *the observers* and *the actors* are the same people, they can be different people (i.e., different identities).

4. The system should consider user fatigue [58, 7]. The entire process should be efficient in order to minimize the effect of user fatigue. We set 15 minutes as the maximum running time of the system.
5. For one emotion, the system should obtain multiple prototypes, thus being closer to reality.

1.2 Solution: new interdisciplinary approaches combining computer science with psychology

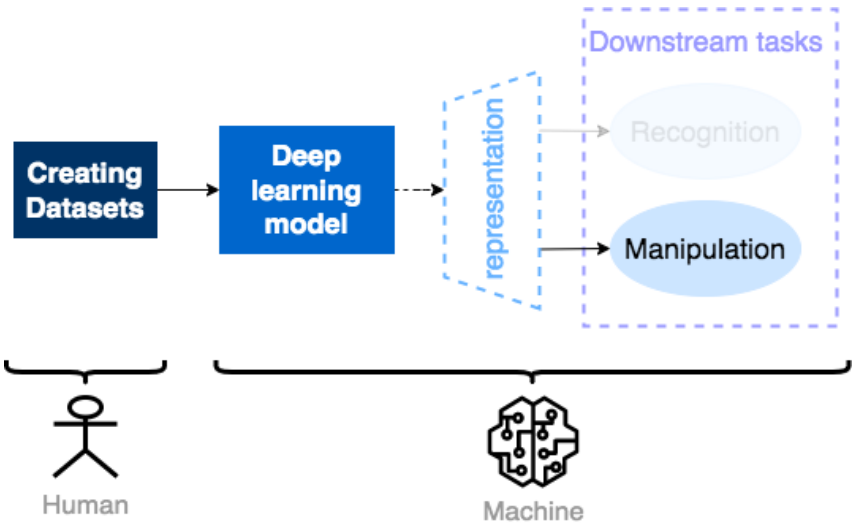


Figure 1.3 – Traditional deep learning pipeline dealing with facial expressions. The first step is achieved by humans, and the others are done by machines. Note that in this thesis, we focus on the facial expression manipulation task (FEM). Facial expression recognition is not the major topic of this thesis, even though some of the concepts presented in this thesis can certainly be applied to FER (the application of FER will be discussed in Chapter 5 Conclusion and perspective).

Traditionally, the recent FE-based deep learning techniques can be summarized into 4 steps (see Fig. 1.3).

1. Database. Establishing a database or directly using existing databases.
2. Deep learning model. Using data to train a deep learning model.
3. Representation. When the deep learning model is well-trained, the features can be extracted.
4. Downstream tasks. Once the representations can be extracted, they can be used for the downstream tasks. Usually, FE-based downstream tasks are facial expression recognition (FER) [1] and facial expression manipulation (FEM).

However, the first step of the traditional deep learning pipeline (i.e., Database) has already been hindered at three-fold. **In terms of diversity**, considering the argument in psychology that Ekman's prototypes are not universal, there should be multiple prototypes for one emotion. However, the databases are usually biased. Notably, the amount of each emotional label can be unbalanced and the ethnicities of the subjects (actors) are also unbalanced. The biased database increases the difficulty to obtain proper representations corresponding to each ethnicity. **In terms of flexibility**, for a given user (observer), it is relatively subjective which emotional prototype can meet his needs. Human perception is subjective [107]. Although subjectivity is beneficial for personalizing facial expressions to meet the need of the observer, there may be discrepancies between the observer's perception and the annotator's judgment [83], thus the facial expressions generated based on the annotation may not be satisfactory to the observer. **In terms of exhaustiveness**, the large labeled database of non-basic emotions such as self-confidence is not explicitly available. Creating such a database is always time-consuming and labor-intensive, and also requires trained experts.

Why not think in a different way? Using the traditional deep learning pipeline likewise confirms the challenges in terms of diversity, flexibility, and exhaustiveness. Can we avoid creating such a complex and challenging database to obtain the representation of emotions? Can we establish a novel system differing from the traditional deep learning pipeline to address the challenges of diversity, flexibility, and exhaustiveness (in Section 1.1.2) and also meet the aforementioned application requirements (in Section 1.1.3)? The answer is yes. Inspired by the mechanism of how psychologists analyze facial expressions, in this thesis, we propose two interdisciplinary approaches differing from the traditional deep learning pipeline. These interdisciplinary approaches that apply to the FEM task, address the existing challenges in terms of diversity, flexibility, and exhaustiveness and finally fulfill all the aforementioned requirements.

1. Note that the facial expression recognition task is not the research objective of this thesis.

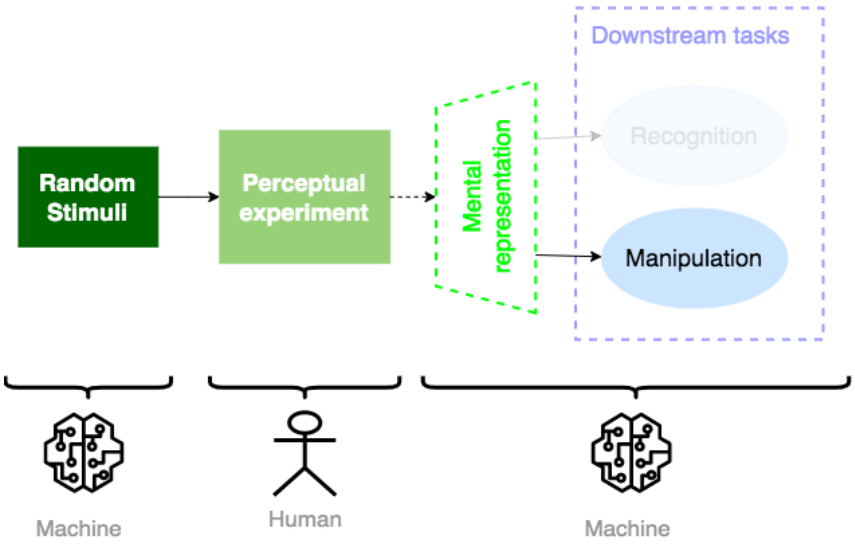


Figure 1.4 – **The first proposed pipeline:** a novel interdisciplinary approach that combines the psychophysical reverse correlation (RevCor) [85, 9] from psychology with Generative Adversarial Networks (GANs) [45] from computer science. Note that facial expression recognition is not the major topic of this thesis.

The first pipeline. We employ a different mindset from the traditional deep learning pipeline. We get inspired by the psychophysical reverse correlation (RevCor) process, typically employed for affective computing in psychology [85, 9]. It is a data-driven method. RevCor can be used to extract the mental prototypes (or called mental representations) of what a given emotion should look like for an observer (or called participant). That is to say, mental prototypes are not limited to basic emotions. The mental prototype of a non-basic emotion can also be extracted. Thus the 1st requirement mentioned in Section 1.1.3 can be addressed. As shown in the first step of Fig. 1.4, differing from the traditional deep learning pipeline, the created data (hereafter, called stimuli) is randomly generated by the machine. Due to the randomness, it can avoid bias derived from the traditional database, and no expert knowledge is required. That is to say, the 2nd requirement mentioned in Section 1.1.3 can be addressed. Leveraging the subjective judgment of the observer in the perceptual experiment of RevCor, the output of the perceptual experiment, i.e., mental representation, contains the representation that meets the need of the observer (e.g., what a self-confident face should look like). Thus the mental representation can be used to personalize facial expressions in the downstream FEM task. This can fulfill the 3rd requirement mentioned in Section 1.1.3. More details about RevCor can be found in Chapter 2. Overall combining the technique from psychology with the technique from

computer science, the first proposed pipeline meets the first three requirements (see Section 1.1.3).

The second pipeline. RevCor has 2 imperfections that should be solved to be largely applied in various scenarios, e.g., the application context of Randstad. The first imperfection corresponds to the 4th requirement (see Section 1.1.3). Each observer is required to perform a large number of randomly generated trials and the entire experiment is designed by experts. Indeed, time-consuming experiments lead to user fatigue, and the reliance on expert knowledge (i.e., expertise) also hinders expansion to other areas. The second imperfection corresponds to the 5th requirement (see Section 1.1.3). RevCor is based on the assumption that there is one, and only one, mental prototype for one affective state that exists in one or a group of observers. This unicity can be questioned. To meet all the requirements, based on the first pipeline, we proposed an optimization method considering human-machine interaction (see Fig. 1.5) to accelerate the pipeline thus solving the 4th requirement. Moreover, our optimization method can provide multiple solutions thus fulfilling the 5th requirement.

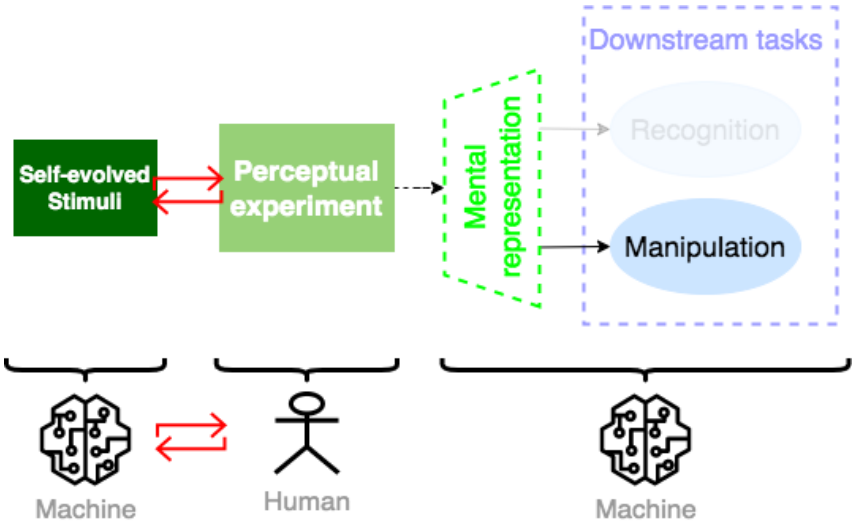


Figure 1.5 – **The second proposed pipeline:** based on the first pipeline, we refine the first two steps by adding human-machine interaction to make the entire process more efficient.

1.3 Contributions

1.3.1 The first pipeline: Mental Deep Reverse-Engineering System (MDR)

We propose a novel interdisciplinary approach: Mental Deep Reverse-Engineering System (MDR), to personalize facial expressions by combining the psychophysical reverse correlation (RevCor) [85, 9] from psychology with Generative Adversarial Networks (GANs) [45] from computer science. Indeed, as an intermediate step of the first pipeline, we extract the mental representation of the desired facial expression that can meet the need of the observer. The mental prototype is **personalized** based on the perception of the observer and does not especially fit any so-called universal prototype. Thus this pipeline meets *Requirement #3 (see Section 1.1.3).*

Differing from typical GANs that can manipulate facial expression, combined with RevCor has the following strengths.

1. **Exhaustiveness.** One can address any emotion or social attitude, for it is, by nature, a role of RevCor. This means we do not need to build a dedicated labeled database for each emotion or social attitude. A low-level-attribute-controlled FEM can cover a wide range of local facial movements. It can be possible to manipulate various facial expressions. This meets *Requirement #1 (see Section 1.1.3).*

2. **Expertise-free.** No expert knowledge in affective computing or certified FACS coder is needed to create the personalized prototype since our approach only requires the observer's perception (i.e., subjective judgment) rather than the observer's expertise. This meets *Requirement #2 (see Section 1.1.3).*

Conversely, differing from recent RevCor approaches, using FEM techniques (such as GANs) allows **manipulating real faces** (2D pictures) rather than virtual avatars, which provides **an easier and more intuitive way** to edit facial expressions. Moreover, we use the same tool twice (GAN, for instance): once for extracting the mental prototype with RevCor and another time for the manipulation. This can ensure that the manipulation is **consistent** with the mental representation of the observer.

Finally, to enhance the definition of facial expression prototypes, we introduce the concept of **dominant and complementary action units** to precisely describe facial expression prototypes.

1.3.2 The second pipeline: Interactive Microbial Genetic Algorithm (IMGA)

We created an efficient interdisciplinary approach called IMGA for Interactive Microbial Genetic Algorithm that meets all the mentioned requirements.

The originality of our approach is that, based on the first pipeline, we integrated the psychophysical reverse correlation (RevCor) process into the interactive genetic algorithm (IGA). This integration not only inherits the strengths from the first pipeline but also solves the drawbacks of the first pipeline.

— **The inherited strengths.**

1. **Exhaustiveness.** The categories of emotions are not limited to those provided in existing deep-learning databases. This is related to *Requirement #1 (see Section 1.1.3).*
2. **Expertise-free.** Our approach only requires the observer's perception (i.e., judgment on intuition) rather than the observer's expertise (e.g., no expert knowledge in affective computing, psychology, or certified FACS coders [37]). This is related to *Requirement #2 (see Section 1.1.3).*
3. **Flexibility.** Everyone can use this pipeline to extract their own mental prototypes of a given emotion. Thus facial expressions can be personalized. This is related to *Requirement #3 (see Section 1.1.3).*

— **The solved drawbacks.** This pipeline becomes efficient and can bring diverse solutions.

1. **Efficiency: by the online feedback loop.** Unlike the traditional RevCor that generates massive trials randomly, in our approach, based on the observer's feedback, automatically updated trials can contain more valuable information (closer to the mental prototypes of observers). This is related to *Requirement #4 (see Section 1.1.3).*
2. **Efficiency: by acceleration.** The way of generating trials for mental prototype computations is intelligent. Moreover, we adopt the microbial genetic algorithm (MGA) [55] as the GA module within IGA to further accelerate the convergence of the system. Indeed, the GA with an elitist mechanism (like MGA) can converge faster than that without an elitist mechanism [67]. This is also related to *Requirement #4 (see Section 1.1.3).*

3. **Diversity.** Benefiting from the genetic algorithm, for one emotion, this pipeline can provide multiple mental prototypes to each observer. That is to say multiple solutions (i.e., prototypes) even for one observer. This is related to *Requirement #5 (see Section 1.1.3).*

Another originality of our approach concerns the creation of such a pipeline. Differing from the traditional genetic algorithm that needs to acquire the fitness values of all individuals, we added a population evaluation module that evaluates the quality of the entire population with limited trials. In addition, we added a three-state constraint automaton to gradually increase the number of activated facial action units (AUs) [37] for each face and determine the process's termination.

1.4 Thesis organization

The thesis is organized as follows.

Chapter 2 introduces the state of the arts to address the existing challenges (Diversity, Flexibility, and Exhaustiveness) and meet the presented 5 requirements.

— This chapter first introduces the current status of facial expression-based databases, since the limitations of the database induce the existing challenges.

— In order to solve the existing challenges, we outline the psychophysical reverse correlation process (RevCor) which is the concept that we get inspired from another discipline.

— Then, we give an overview of facial expression manipulation (FEM). This is one of the major FE-based downstream tasks and is part of the RevCor process (generating stimuli).

— We present an overview of the interactive genetic algorithm (IGA). By integrating IGA into RevCor, the optimized pipeline becomes more efficient so that all 5 requirements can be met.

Chapter 3 introduces our first pipeline using a novel interdisciplinary (computer science & psychology) approach. This chapter is divided into 3 parts:

— We detail each part of this pipeline. For this pipeline, we choose FEM as the downstream task.

— We present the corresponding results.

— We employ subjective evaluations to prove the validity of this approach and discuss the convergence efficiency.

Chapter 4 presents the second pipeline that optimizes the previous approach and fulfills all the aforementioned requirements. This chapter is organized as follows:

— We first detail the framework of this approach.

— Then we present the experiment setting and the corresponding results.

— We next employ quantitative and subjective evaluations to prove the validity of this approach and compare our approach with related works in terms of efficiency.

Chapter 5 summarizes the contributions of this thesis and presents comprehensive perspectives.

Chapter 2

State of the art

In recent years, facial expression-based (FE-based) techniques have flooded our daily lives. Different applications have emerged in many domains. However, the limitation of large labeled databases can still be a significant obstacle, resulting in the following challenges: diversity, flexibility, and exhaustiveness (mentioned in Chapter 1). In this chapter, as shown in Fig. 2.1, we first introduce the current status of FE-based databases in Section 2.1. To address the existing challenges of **Diversity**, **Flexibility**, and **Exhaustiveness**, we propose another way of thinking by combining an idea from psychology and ideas from computer science. Thus, in Section 2.2, we outline the related works on facial expression manipulation (FEM), and in Section 2.3, we overview the psychophysical reverse correlation process (RevCor). These two techniques are related to our first pipeline aforementioned in Section 1.2. However, combining these techniques from two disciplines can only meet the first three requirements (see Section 1.1.3). Optimization is necessary to fulfill all the requirements. In Section 2.4, we overview the interactive genetic algorithm (IGA) that can optimize a system with human involvement. The IGA technique is related to our second pipeline.

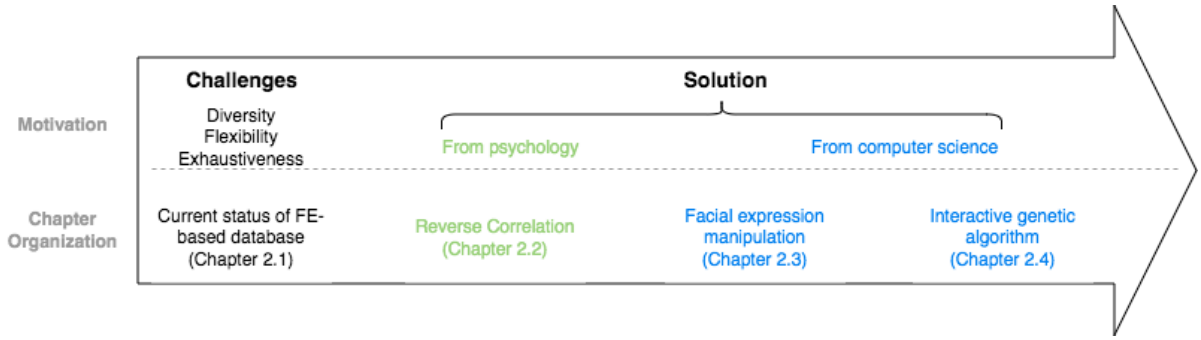


Figure 2.1 – Organization of Chapter 2.

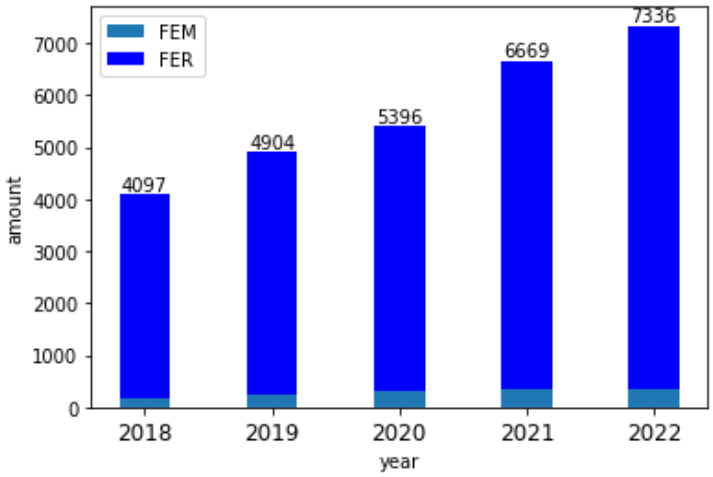


Figure 2.2 – Recent publications from 2018 to 2022 on downstream tasks (in total): facial expression recognition (FER) and facial expression manipulation (FEM). We search on Google Scholar with the keywords facial expression recognition for FER and facial expression manipulation/editing/synthesis/transfer for FEM.

2.1 Facial expression-based databases

Data is an inevitable topic in deep learning. In general, good data is a prerequisite for good results. Fig. 2.2 shows that in the recent 5 years, there is a stable increase in publications on the FE-based downstream tasks: facial expression recognition (FER) and facial expression manipulation (FEM). Although FE-based databases are primarily employed for FER [1], they can also be used for FEM. Indeed, the existing drawbacks of the database influence the downstream tasks. This section includes the FE-based databases widely used (at least 500 citations) for facial-expression downstream tasks. These databases are first presented and then discussed in terms of diversity, flexibility, and exhaustiveness.

2.1.1 Introduction of FE-based databases

We summarize the publicly available databases that are widely used (at least 500 citations) for facial-expression downstream tasks since 1998. These databases are sorted according to the published year. For a purpose of clarity, databases listed in this thesis are given by their abbreviations.

JAFFE [78]. Published in 1998, the Japanese Female Facial Expression (JAFFE) database is a laboratory-controlled image database containing 213 samples of posed facial

1. as a downstream task, FER is not the main topic of this thesis.

expressions from 10 Japanese females. There are few examples per subject and expression. Each subject has three or four images with each of six basic emotions (happy, sad, angry, fear, disgust, and surprise) and one image with a neutral face.

CK+ [77]. First published in 2000 as CK [65], then updated in 2010, the Extended Cohn-Kanade (CK+) database is the most widely used laboratory-controlled database for evaluating FER systems. CK+ contains 593 video sequences from 123 subjects. The sequences vary in duration from 10 to 60 frames and show a shift from a neutral expression to the peak expression. Among these videos, 327 sequences from 118 subjects are labeled with seven basic expression labels (the six basic emotions plus contempt) based on the Facial Action Coding System (FACS) [37].

MMI [88, 112]. First published in 2005, the MMI database is laboratory controlled and contains 326 sequences from 32 subjects. A total of 213 sequences are labeled with six basic emotions. Unlike CK+, the sequence in MMI starts with a neutral expression and reaches the peak expression near the middle, then returns to the neutral face.

BU-3DFE [124]. Published in 2006, the Binghamton University 3D Facial Expression (BU-3DFE) database contains 606 facial expression sequences recorded from 100 subjects. For each subject, six basic emotions are elicited with different intensities. This database is typically used for multi-view 3D facial expression analysis.

Multi-PIE [50]. Published in 2010, the CMU Multi-PIE database contains 755,370 images from 337 subjects under 15 viewpoints and 19 illumination conditions. Each facial image is labeled with one of six expressions: disgust, neutral, scream, smile, squint, and surprise. This database is typically used for multi-view facial expression analysis.

RaFD [70]. Published in 2010, the Radboud Faces Database (RaFD) [2] is laboratory-controlled containing a total of 1,608 images from 67 subjects with three different gaze directions, i.e. front, left, and right. Each sample is labeled with one of eight emotions: the seven basic emotions (including contempt) plus neutral.

Oulu-CASIA [130]. Published in 2011, the Oulu-CASIA database contains 2,880 image sequences collected from 80 subjects labeled with six basic emotions. Each of the videos is captured using two imaging systems: near-infrared (NIR) or visible light (VIS), under three different illumination conditions. Similar to CK+, the first frame is neutral and the last frame has the peak expression.

DISFA [80]. First published in 2013, updated in 2016 [79], the Denver Intensity of Spontaneous Facial Action (DISFA) Database. This database contains stereo videos of 27

2. updated information can be found: https://rafd.socsci.ru.nl/RaFD2/RaFD?p=overview

adult subjects (12 females and 15 males) with different ethnicities. The intensity of AUs (0-5 scale) for all video frames is manually scored by two FACS experts. The database also includes 66 facial landmark points of each image.

FER2013 [47]. Published in 2013, the FER2013 database was presented during the ICML 2013 Challenges in Representation Learning. FER2013 is a large and unconstrained database automatically collected by the Google image search API. FER2013 contains 28,709 training images, 3,589 validation images, and 3,589 test images with six basic emotions and neutral. Unlike the previous databases, this database is an in-the-wild database.

AFEW [27]. Published in 2013, the Acted Facial Expressions in the Wild (AFEW) database was first established and introduced in 2011 [26] and has been wildly known in the annual Emotion Recognition In The Wild Challenge (EmotiW) since 2013 [3]. As an in-the-wild database, AFEW contains video clips collected from different movies with spontaneous expressions, various head poses, occlusions, and illuminations. Samples are labeled with the six basic emotions plus neutral. The annotations have been continuously updated, and reality TV show data have been continuously added.

EmotioNet [41]. Published in 2016, EmotioNet is a large-scale database with one million facial expression images collected from the Internet. A total of 950,000 images were annotated by the automatic action unit (AU) detection model and the remaining 25,000 images were manually annotated with AUs. This in-the-wild database also provides basic and compound emotion annotations.

RAF-DB [73]. Published in 2017, the Real-world Affective Face Database (RAF-DB) is a real-world database containing 29,672 highly diverse facial images downloaded from the Internet. With crowd-sourcing and reliable estimation, seven basic and eleven compound emotion labels are provided. Specifically, 15,339 images from the basic emotion set are divided into two groups (12,271 training samples and 3,068 testing samples) for evaluation.

AffectNet [83]. Published in 2017, AffectNet contains more than one million images from the Internet that were obtained by querying different search engines using emotion-related tags. This database provides facial expressions in two different emotion models (categorical model and dimensional model), of which 450,000 images are manually annotated by seven basic expressions.

CMU-MOSEI [127]. Published in 2018, CMU Multimodal Opinion Sentiment and Emotion Intensity (CMU-MOSEI) dataset is the largest dataset of multimodal sentiment

3. updated with the EmotiW challenges 2013 - 2020: https://sites.google.com/view/emotiw2020/.

Table 2.1 – Widely used FE-based databases in terms of data, number of subjects, and collection condition. Collect = Collection condition; n/a = not available.

Database	Data	Subject	Collect
JAFFE [78]	213 images	10	lab
CK+ [77]	593 videos	123	lab
MMI [88, 112]	2903 videos	75	lab
BU-3DFE [124]	2500 images	100	lab
Multi-PIE [50]	755,370 images	377	lab
RaFD [70]	1608 images	67	lab
Oulu-CASIA [130]	2,880 videos	80	lab
DISFA [80]	130,000 images	27	lab
FER2013 [47]	35,887 images	n/a	in-the-wild
AFEW [27]	54 movies	330	in-the-wild
EmotioNet [41]	1,000,000 images	n/a	in-the-wild
RAF-DB [73]	29,672 images	n/a	in-the-wild
AffectNet [83]	450,000 images	n/a	in-the-wild
CMU-MOSEI [127]	23500 videos	1000	in-the-wild

analysis and emotion recognition to date. The dataset contains more than 23,500 sentence utterance videos from more than 1000 online YouTube speakers. The dataset is gender balanced. All the sentences utterance are randomly chosen from various topics and monologue videos. The annotated emotions only include the six basic emotions.

Table 2.1, 2.2, and 2.3 provide an overview of these databases, including data, number of subjects, range of ages, gender ratio, ethnicity, collection condition, elicitation method, face labels, annotators.

In the following subsections, We explain how the current status of databases leads to the three existing challenges (in Chapter 1): diversity, flexibility, and exhaustiveness. The summary is shown in Table 2.4.

2.1.2 Diversity

Biased databases limit the diversity of facial expression-based (FE-based) techniques. We discuss it in three aspects: the collection condition (listed in Table 2.1), available demographics (including the range of ages, gender ratio, and ethnicity, listed in Table 2.2), and face labels (including emotional labels and low-level attributes, available in Table 2.3).

Table 2.2 – Widely used FE-based databases in terms of age range, gender, and ethnicity. yo = years old; ad = adults; ch = children; avg = average age; EA = Euro-American; AA = African American; SA = South American; A = Asian; E = European; H = Hispanic; n/a = not available.

Database	**Age**	**Gender**	**Ethnicity**
JAFFE [78]	n/a	100% female	100% Japanese
CK+ [77]	18 - 50 yo	69% female	81% EA, 13% AA, 6% other
MMI [88, 112]	19 - 62 yo	44% female	E, SA, A
BU-3DFE [124]	18 - 70 yo	56% female	7 ethnicities
Multi-PIE [50]	avg: 27.9 yo	30.3% female	60% EA, 30% A, 3% AA, 2% other
RaFD [70]	57 ad,10 ch	37% female	73% Dutch, 27% Moroccan
Oulu-CASIA [130]	23 - 58 yo	26.2% female	62.5% Finnish, 37.5% Chinese
DISFA [80]	18 - 50 yo	44% female	11% A, 78% EA, 7% H, 4% AA
FER2013 [47]	n/a	n/a	n/a
AFEW [27]	1 - 70 yo	n/a	n/a
EmotioNet [41]	n/a	n/a	n/a
RAF-DB [73]	0 - 70 yo	52% female	77% Caucasian, 8% AA, 15% A
AffectNet [83]	n/a	n/a	n/a
CMU-MOSEI [127]	n/a	47% female	n/a

Table 2.3 – Widely used FE-based databases in terms of face labels and annotators. basic = basic emotions; compound = compound emotions; lmk = landmarks.

Database	**Labels**	**Annotator**
JAFFE [78]	6 basic	Human
CK+ [77]	7 basic & 30 AUs	FACS coder
MMI [88, 112]	6 basic & 31 AUs	FACS coder, human
BU-3DFE [124]	6 basic[3]& 3D lmk	Expert
Multi-PIE [50]	6 expressions[2]& 68 lmk	n/a
RaFD [70]	7 basic & 17 AUs	FACS coder
Oulu-CASIA [130]	6 basic	Human
DISFA [80]	12 AUs[1], 66 lmk	FACS coder
FER2013 [47]	6 basic	Human
AFEW [27]	6 basic	n/a
EmotioNet [41]	23 basic and compound & 17 AUs	Machine, human
RAF-DB [73]	18 basic and compound & 37 lmk	Expert
AffectNet [83]	7 basic & valence-arousal	Expert
CMU-MOSEI [127]	6 basic	Expert

[1] with 6 levels of intensity

[2] smile, surprised, squint, disgust, scream, neutral

[3] with 4 levels of intensity

Char.	Data	Subject	Collect	Age	Gender	Ethnicity	Annotator	Labels
Chllg.	D	D	D	D	D	D	F	D, F, E

Table 2.4 – The link between the characteristics of databases and the existing FE-based challenges. Char. = Characteristic; Chllg. = Challenge; D = Diversity; F = Flexibility; E = Exhaustiveness; Data = Database size; Collect = Collection condition.

Collection condition: laboratory-controlled vs in-the-wild.

Since 1998, the size of FE-based databases is increasing: from hundreds of megabytes to hundreds of gigabytes and from about two hundred images to one million images. Categorizing the listed databases by the collection condition, **in terms of database size**, the laboratory-controlled databases (from JAFFE to DISFA) generally contain fewer data than the in-the-wild databases (from FER2013 to CMU-MOSEI). As shown in Table 2.1, **in terms of the number of subjects**, the laboratory-controlled databases have limited subjects from 10 to 377. whereas with the increase in database size, there are usually much more subjects from most in-the-wild databases.

In terms of diversity, although the data collected in the wild increases the diversity of the database (e.g., more subjects with different poses, illumination conditions, with or without occlusion, etc.), some recording information (or called metadata), such as the number of subjects (summarized in Table 2.1) and the demographics (summarized in Table 2.2), is not available. It might be difficult to track (or annotate) the precise demographic information for in-the-wild databases, for instance, the fictional characters of Harry Potter collected in AFEW [27] database, and quote from CMU-MOSEI [127] "accurate identification requires quadratic manual annotations, which is infeasible for (a) high number of speakers." Conversely, the laboratory-controlled databases contain less diverse data but more detailed recording information than the in-the-wild databases. For the present, a deep-learning model contains hundreds of thousands to hundreds of billions of parameters, e.g., GPT-3 [11] already has 175 billion parameters. **Less available data is not enough to train such a big deep-learning model and less diverse data prevents the AI model from extracting diverse emotional prototypes.**

Demographics: age, gender, and ethnicity.

According to the available information, we observe that **the demographics in most databases are not diverse and remain biased.** Although databases become larger, the demographics of subjects remain biased and less diverse.

— **Range of ages.** Except that RaFD [70] collected face images from 10 children, most laboratory-controlled databases only contain adults' faces.
— **Gender ratio.** Based on the available demographics, MMI [88, 112], BU-3DFE [124], DISFA [80], RAF-DB [73], and CMU-MOSEI [127] have a fair male-female ratio (between 44% and 56%). Others are either not available or rather unbalanced.
— **Ethnicity.** Most subjects are European or Euro-American, whereas subjects of other ethnicities are much fewer.

As aforementioned the challenge of diversity in Chapter 1, a growing number of psychologists refute the universality of emotions. That is to say, emotional prototypes can be diverse across age, gender, and ethnicity. However, a database containing unbalanced and less diverse subjects may lead to the AI-learned prototypes being biased toward a particular group of people (e.g., Caucasian male adults). This can limit the diversity of the facial expression manipulation (FEM) task. Indeed, the generated facial expression might be more western rather than eastern, or more like an adult and less like a child. Moreover, a database that is large enough but lacks detailed recording information, such as age, gender, and ethnicity (see in-the-wild databases in Table 2.2), can also prevent the AI model from extracting diverse emotional prototypes.

Face labels

In terms of emotional labels. In terms of the listed emotions, the frequency of each emotional label is **not relatively balanced**. Most databases have more "happy" faces. For instance, as a laboratory-controlled database, in CK+ [77], of the 327 videos annotated by manual FACS coders as an emotion, there are 69 and 83 videos that are annotated by "happy" and "surprise". However, there are 18 and 25 videos that are annotated by "contempt" and "fear". As an in-the-wild database, FER2013 [47] contains only 547 "disgust" but 8989 "happiness". For an example of valence-arousal values [96, 52], most images in AffectNet [83] are "happy" and "neutral". That is why there are thousands of samples (in red and orange) in the center and the right middle (positive valence and small positive arousal) of the circumplex, whereas there are less than 32 samples (in blue) in a fairly large area in the circumplex, even no sample (in white) in some areas (see Fig.4 of [83]).

In terms of low-level-attribute labels. The unbalanced emotion labels can lead to **some of the low-level attributes having very few labels**, such as some action units [37] (AUs). For an example of AUs, CK+ [77] contained 30 AUs manually labeled by FACS coders. The AU that appears the most is AU25 (lips part) with 287 times. However,

there are 8 AUs that appear less than 10 times, with AU28 (lip suck), AU29 (jaw thrust), and AU34 (cheek puff) occurring only once.

Overall, on one hand, the unbalanced face labels may make it more difficult for the model to well learn the features of the minority emotions. On the other hand, too few low-level-attribute labels are impossible to train a deep-learning model. That is why there are 46 main action units but the existing tool can only manipulate about 16 AUs. This may lead to poor performance: fewer optional texture combinations for synthesizing facial expressions.

2.1.3 Flexibility

The challenge of flexibility is due to a series of factors. Here, we explain the factors in detail first through the annotators and then through the face labels.

Annotators

All the databases are labeled. The annotators can be classified into three categories. We first list these three categories, then discuss each category.

— Human. Non-expert-judgment by the subjects is employed to annotate data, such as MMI [88]. In Oulu-CASIA [130], the subjects directly imitate the facial expression example to automatically annotate data. JAFFE [78] and FER2013 [47] are rated by human annotators (expertise: unknown).

— Expert. Trained FACS coders (CK+ [77], MMI [88, 112], RaFD [70], DISFA [80]), psychologists (BU-3DFE [124], AffectNet [83]), or trained annotators (with a one-hour tutorial on psychology, RAF-DB [73]) are hired to annotate the facial expressions.

— Machine. EmotioNet [41] employed the machine-learning method [4] to automatically annotate AU and emotions.

Human-observer judgment is subjective. Especially in crowd-sourcing, the labels can vary considerably among annotators without expert knowledge [83]. The possible reasons are as follows. 1) User fatigue should be considered for human annotators. 2) Expert knowledge may be required to classify some facial expressions and annotate low-level attributes such as AUs. 3) The cross-culture hypothesis in psychology might be the reason, i.e., human annotators with different cultural backgrounds may lead to different emotional annotations [61, 62].

Differing from non-expert annotators, in some databases, annotators are experts (usually psychologists or trained in psychological knowledge) thus relying on expert knowledge to increase the reliability of the annotation. Yet, to validate the emotion corresponding to a facial expression, **even expert annotators have different judgments**. Quote from AffectNet [83] "*the annotators (experts) highly agreed on the Happy and No Face categories, and the highest disagreement occurred in the None category.*" and quote from BU-3DFE [124] "*the most likely confused expressions were sad-fear and disgust-angry even for experts.*"

Machine labeling can quickly annotate a massive amount of data without human involvement. Thus, user fatigue does not need to be considered. Moreover, the entire process follows a fixed annotation mechanism. For a given data, the machine will always bring the same annotation rather than different annotations that may be brought by human or expert annotators. **However, machine labeling is still based on Ekman's hypothesis of universality.** Indeed, for EmotioNet [41], Openface [4] was employed to automatically detect AUs, whereas validating the corresponding emotions still follows the *emotional FACS rules* of Ekman (or called prototypes of Ekman) [4] [37].

Overall, **in terms of flexibility**, on the one hand, non-expert annotators are too subjective and might lead to different annotations. Even experts can make different judgments. On the other hand, due to adherence to Ekman's hypothesis of universality [37], the annotating process can be biased. Indeed, most emotion validation processes done by humans or machines are based on the prototypes of Ekman. Therefore, personalizing facial expressions using the existing databases is not feasible since the generated facial expressions only meet Ekman's emotional prototypes but may not be suitable for the observer (i.e., the user).

4. Note that although FACS (Facial Action Coding System) is proposed by Ekman, the argument in psychology about the universality of emotion focuses on emotional FACS rules (i.e., the prototypes of Ekman) rather than FACS itself [97, 63, 5, 61, 62]. In detail, the *emotional FACS rules* are used to interpret the corresponding emotion based on the activated AUs (i.e., a decision-making process). These rules are based on Ekman's hypothesis of universality. For instance, based on *emotional FACS rules*, the combination of AU6 and AU12 is classified as happiness. However, FACS is an objective coding system. It only focuses on facial muscle movements (i.e., encoding a face by action units). That is to say, it is independent of interpretation. For instance, based on FACS, an annotator (e.g., FACS coder) annotates a face as AU6 (activated) and AU12 (activated), but inferring which emotion category this AU combination belongs to is not included in FACS coding.

Face labels

In terms of low-level attributes, it can be a good choice to use objective low-level attributes to personalize facial expressions. To reduce the subjectivity of data labeling and increase the reliability of the annotation, a generally accepted way is to employ FACS coding (i.e., action units, AUs) to annotate human facial expressions. The main reason is that AUs are relatively objective descriptors (containing anatomical information) and independent of interpretation. They can be used for FEM task (e.g., synthesizing facial expressions by activating several AUs.) [128]. No prior knowledge involves in the manipulation process and the manipulation can be more flexible.

To sum up, FACS (not *emotional FACS rules*[5]) is a good and objective way to annotate facial expressions. Indeed, **AUs** (a low-level attribute) only represent facial muscle movements. They are very suitable to describe the richness of spontaneous facial behavior since hundreds of anatomically possible facial expressions can be represented by a combination of only a few AUs. Similar to FACS-coded AUs, **facial 2D/3D landmarks** are another low-level attribute that can objectively describe facial expressions.

In terms of emotion labels, using emotion labels to personalize facial expressions is not feasible. Indeed, the emotional labels (e.g., basic emotions, and compound emotions) are biased or relatively subjective compared with low-level attributes.

— **Biased**. As we discussed in Section 2.1.3, annotating the emotion labels by the machine usually follows the prototypes of Ekman. However, the prototypes of Ekman are not universal (mentioned Section 1.1.2). That is to say, for a given emotion, the prototypes can be different between different people. If the prototype of the observer (i.e., user) is different from the prototype of Ekman, the generated facial expression can not follow the observer's will and can not meet the need of the observer.

— **Subjective.** Non-expert annotators' judgments are subjective, even experts can make different judgments. For some facial expressions, there can be disagreement regarding the annotation. The controversial emotion labels can cause the deep learning model to inaccurately learn the corresponding features. Thus for the FEM task, the generated facial expression can be controversial.

5. hereafter, uniformly expressed as *prototypes of Ekman*

2.1.4 Exhaustiveness

Most listed databases focused on basic facial expressions, i.e., happiness, sadness, anger, fear, surprise, disgust, and contempt. Only EmotioNet [41] and RAF-DB [73] contain compound emotions [32], i.e., combinations of two basic emotions. Indeed, data labeling becomes a challenge once we move beyond basic emotions. **The reliance on expertise limits the variety of available labels.** The non-basic emotion labels are not explicitly available. Even though some facial expressions are FACS-coded and included in the databases (e.g., a part of EmotioNet [41]), they can not be assigned an emotional label. Indeed, **it is much more difficult to annotate complex emotions, even for the annotators with pre-studied knowledge.** For example, the EmoReact dataset [6] [86] was annotated by a group of selected annotators (balanced male-female ratio, with pre-studied knowledge on affective computing). Based on the agreement levels (see [54] for computation details) for 17 affective states, 2 basic emotions (happiness and fear) and 4 complex emotions (curiosity, uncertainty, excitement, and frustration) were categorized as *moderate agreement*; only 2 basic emotions (surprise and disgust) and valence were categorized as *substantial agreement*; whereas 8 of 17 affective states had poor agreement levels (i.e., invalid) including 2 basic emotions (anger and sadness), Neutral and 5 complex emotions (exploration, confusion, anxiety, attentiveness, and embarrassment).

Note that as research in psychology covers, there are already more than 4000 emotional labels [106]. In the real world, it is not enough to deal with only basic emotions and very few non-basic emotions, such as compound emotions.

2.1.5 Conclusion

Limitation

Data is the root of deep learning. Nowadays, a deep-learning model contains hundreds of thousands to hundreds of billions of parameters (e.g., GPT-3 [11]: 175 billion). That is to say, the amount of training data should reach at least a similar magnitude. As the current trend of databases, through keyword searches on the internet, sufficient in-the-wild data can be collected. However, the drawbacks of both laboratory-controlled

6. A newly collected multimodal emotion dataset of children between the ages of four and fourteen years old. This dataset is annotated for 17 affective states, including six basic emotions (happiness, sadness, surprise, fear, disgust, and anger), neutral, valence, and nine complex emotions (curiosity, uncertainty, excitement, attentiveness, exploration, confusion, anxiety, embarrassment, and frustration).

and in-the-wild databases are the triggers for the existing FE-based challenges: diversity, flexibility, and exhaustiveness.

The first drawback is that most databases are biased. Unbalanced subject demographics (including the range of ages, gender ratio, and ethnicity) and unbalanced face labels (including emotional labels and low-level attributes) lead the emotional prototypes toward a specific group of people and contain fewer various prototypes for some emotions. Hence, this reduces the **diversity** of databases. This also influences the **flexibility** and makes it impossible to personalize facial expressions by a specific observer to a specific audience.

The second drawback is that creating a labeled FE-based database always relies on expertise. After collecting raw data, expert knowledge is always required for annotations. Whether the annotators are experts or non-experts, their decisions regarding emotional labels (such as basic emotions) can be different due to human factors (e.g., user fatigue, and subjective judgment). Generally, experts' annotations are more reliable than those of non-expert. Moreover, in order to ensure the reliability of the annotations, all the organizers of the listed databases hired several annotators to independently annotate the same set of data. **This also indicates that creating a database is always time-consuming and labor-intensive.** In addition, relying on expertise limits the **exhaustiveness** of databases. Most databases can only deal with basic and compound emotions. Considering the refutation of the universality of prototypical expressions of emotions in psychology, we speculate that annotating non-basic emotions might be more challenging since their prototypes vary across age, gender, and culture. So that might be why labels for non-basic emotions are always not explicitly available.

Perspectives

In terms of subject demographic. For different people, their emotional prototypes can be different. That is why subject demographic is an important characteristic of an FE-based database. Although it can be easier to create a laboratory-controlled database with comprehensive and diverse subjects, for in-the-wild databases, tracking the demographic information of subjects is still challenging and needs to be solved.

In terms of face labels. A possible way is to use multiple labels in order to reduce the subjectivity of human judgment and to represent comprehensive attributes of affect displays. Face labels should contain not only emotional labels but also low-level attributes. However, **Which kind of low-level facial attributes should be provided?** Low-level

attributes should adopt relatively objective descriptors, such as face landmarks and action units. These low-level attributes only provide geometric and anatomical information, which is more objective. Providing objective low-level attributes can increase the reliability of the database since the face data has an additional objective description rather than just an emotion label (which may be controversial between different annotators). Moreover, controlling low-level attributes allows for more flexibility in generating a wider variety of facial expressions than controlling emotion labels. Furthermore, the reliability of annotating low-level attributes can be guaranteed by providing enough training (to annotators) on annotating schemes such as FACS [37] and AAM [20].

Our proposition

As shown in Fig.1.3, the goal in creating an appropriate database containing non-basic emotions (e.g., self-confidence) is to train a deep-learning model and then derive the corresponding emotional representations to solve the downstream task: FEM (e.g., automatically transferring the neutral face to self-confidence). As far as databases are concerned, it is still challenging 1) to create such a database that considers diversity, flexibility, and exhaustiveness, and then 2) to use this database to train a deep-learning model dealing with the application requirements of Randstad (in Chapter 1).

Why not think in a different way? Can we avoid creating such a complex and challenging database to obtain the representation of emotions? Can we establish a system to address the challenges of diversity, flexibility, and exhaustiveness (in Section 1.1.2) and also meet the aforementioned application requirements (in Section 1.1.3)? Inspired by the mechanism of how psychologists analyze facial expressions, we propose a novel interdisciplinary approach that is presented in two pipelines that we called Mental Deep Reverse-engineering system (i.e., the first pipeline, abbreviation: MDR), and Interactive Microbial Genetic Algorithm (i.e., the second pipeline, abbreviation: IMGA). These pipelines combine the concept of psychology and the concept of computer science. In the following sections, we outline the related technique inspired by computer science, i.e., facial expression manipulation (FEM, in Section 2.2) and interactive genetic algorithm (IGA, in Section 2.4), and the related technique inspired by psychology, i.e., reverse correlation (RevCor, in Section 2.3).

2.2 Facial expression manipulation

Facial expression manipulation (FEM) is a typical FE-based downstream task. Based on an input face image and control parameters (i.e., instruction for manipulation), FEM is able to generate a new face image as an output. According to the control parameters for manipulation, FEM can be divided into two categories: high-level-attribute manipulation and low-level-attribute manipulation.

2.2.1 High-level-attribute manipulation

Generative Adversarial Networks (GANs) [45] have achieved a series of impressive results in image-to-image translation tasks on real faces [60, 131]. Most GANs for FEM tasks generate faces by controlling the high-level attributes defined by the training database. For instance, as one of the representative FEM models, StyleGAN [66] provided a state-of-the-art architecture to generate high-resolution and more realistic faces. Trained from the proposed FFHQ dataset, this approach can manipulate so-called "styles" of the face, i.e., the identity-related features such as pose, general hairstyle, general face shape, eyes open/closed, and color scheme. Although the results are spectacular (high resolution and photo-realistic), StyleGAN focuses on editing the identity-related features (i.e., styles) rather than facial expressions of emotion. Unlike StyleGAN, StarGAN [17] can edit facial expressions by transferring from one emotion category to another. StarGAN is trained from the Radboud Faces Database (RaFD) [70]. As aforementioned, due to the limitation of the database, most FEM models like StarGAN can only deal with the six basic emotions of Ekman. However, the reality is that there are already more than 4000 emotional labels [106].

Overall, the attributes that can be manipulated are limited by the availability of high-level attributes in the training database. While controlling high-level attributes can generate a variety of highly realistic faces, these approaches cannot in principle generate arbitrary and fine-grained facial expressions.

2.2.2 Low-level-attribute manipulation

Instead of manipulating high-level attributes, other methods edit relatively low-level attributes, such as geometric landmarks or action units, to generate facial expressions. As a representative task of manipulating low-level attributes, face reenactment aims at

animating the facial expression of a target video using the video of a source actor [13]. Most face reenactment approaches are based on facial landmarks. For example, the Face2Face model [109] and the dual-generator-based approach [59] employed 3D landmarks to encode head pose, face shape, and facial expression. The approaches such as ReenactGAN [118] and FReeNet [129] employed 2D facial landmarks. However, for the purpose of synthesizing various facial expressions, face reenactment obviously requires that the source video must include all possible facial expressions.

Other than face reenactment, G2-GAN [100] employed facial geometry as controllable parameters to synthesize the basic facial expressions. Yet, the range of expressions that can be synthesized remains limited. Departing from this approach, GANimation [94] can generate "anatomically-aware" face animation by taking a list of action units (AUs) as input. AUs [37] are defined by the contraction or relaxation of one or some muscles, and can be used in combination to construct facial expressions. Thus, GANimation can manipulate facial expressions with relatively fine control. Moreover, AUs are relatively objective compared with high-level attributes such as emotion labels and valence-arousal values. Controlling AUs (by activation or deactivation) can be more intuitive and flexible for synthesizing various facial expressions. Nevertheless, controlling the relatively objective low-level attribute (e.g., an AU vector) to synthesize a given facial expression still requires human expertise (e.g., FACS-certified coder).

2.2.3 Discussion and conclusion

In terms of the existing challenge of diversity. For high-level attribute manipulation such as emotion categories, the output for each emotion is always fixed. That is to say, for each emotion, the FEM technique follows the same and the unique emotional prototype (usually, the prototype of Ekman) to synthesize facial expressions. However, for each emotion, there should be multiple prototypes and should be different across age, gender, and culture (see Chapter 1). For low-level attribute manipulation, more facial expressions (not limited to expressions that only correspond to emotions) can be generated by controlling the low-level attribute. As aforementioned in Section 2.1.3, the low-level attributes for the FEM task should be relatively **objective**, such as AUs and facial landmarks.

In terms of the existing challenge of flexibility. No matter how spectacular (high resolution, photo-realistic) the generated results are, the output facial expression can not be personalized. Indeed, expert knowledge is required for low-level attribute manipulation.

For instance, which AUs are related to eyebrows? Which combination of AUs can generate a self-confident face? To the best of our knowledge, there is always an expert process required for low-level attribute manipulation, e.g., FACS for AUs and *emotional FACS rules* for emotion. Especially, the manipulation should be applied to different subjects (or called actors), and different audiences, thus meeting the needs of different observers (i.e., users).

In terms of the existing challenge of exhaustiveness. FEM technique still can not synthesize non-basic emotions. The possible reasons are as follows: 1) for high-level attribute manipulation, the emotion labels are usually limited to the basic emotions of Ekman; 2) for low-level attribute manipulation, the corresponding representation of non-basic emotion is unknown.

In this thesis, **we chose GANimation [94] as a tool to synthesize facial expressions by controlling AUs**. As we have discussed in Section 2.1, this low-level attribute is relatively **objective** and only covers anatomical information about the face, i.e., facial muscle movements. Various facial expressions can be generated by activating only a few AUs. To alleviate the reliance on the database and achieve expertise-free (i.e., without the need for expert knowledge), we mixed GANimation with reverse correlation, i.e., a concept from psychology.

2.3 Psychophysical reverse correlation process

Here, we outline the psychophysical reverse correlation (RevCor) process and discuss how to combine the two disciplines (FEM from computer science and RevCor from psychology) to tackle the existing challenges.

2.3.1 Reverse correlation process for affective computing

The reverse correlation process (RevCor) is a powerful data-driven method widely used in the field of psychology. RevCor involves presenting a series of stimuli (e.g., visual or audio) to observers (or participants) and then asking them to report their perception of the stimuli. Based on observers' judgments of the large quantity of randomly-varied stimuli, RevCor is able to reverse-engineer what perceptual (or called mental) representations are most strongly associated with their judgments [9]. This can help researchers to identify the neural mechanisms and processing strategies involved in perception. In detail, this

process can be divided into three steps.

— **Stimuli generation.** Relying on the existing tool [7] such as [126, 63, 14, 12, 93], visual or audio stimuli can be randomly generated by modifying the corresponding control parameters. These control parameters are associated with prosodic features such as pitch, duration, and loudness [12, 93], or visual features such as action units of virtual avatars [126, 63, 14]. Even some stimuli were generated by adding noise to the original data such as Gaussian white noise used in the bubble method [48].

— **Perceptual experiment.** The perceptual experiment is designed by experts. It can be divided into hundreds or thousands of trials. In each trial, the participant (or called observer) perceives the given stimuli (or stimulus) usually by watching or listening, then answers the question provided by experts. The questions follow different paradigms: 2-AFC (2-alternative forced choice) and n-AFC (n-alternative forced choice, n is an integer greater than 2). In these works [12, 93, 49], the 2-AFC paradigm is employed. Each trial consists of two stimuli. Participants were asked to *choose the stimulus* that best reflected their mental representation (see Fig. 2.3 left). However for the n-AFC paradigm employed by these works [126, 63, 14, 48], in each trial, only one stimulus was presented. Participants were asked to *choose the answer*, i.e., one option out of n that best reflected their mental representation, and they were usually also asked to indicate the intensity or probability of the corresponding option (see Fig. 2.3 right).

— **Mental representation computation.** Based on the participants' responses, it is possible to determine which features/parameters are significantly related to participants' perceptions. The corresponding mental representation can then be obtained. Applied in audio stimuli, the idea of [12, 93] to compute mental representation was the subtraction of two opposite options, for instance, the mean pitch contour of the voices classified as "trustworthy" minus that classified as "non-trustworthy" in [12], and the mean pitch contour of "interrogative" minus that of "declarative" in [93]. Applied in visual stimuli, due to the paradigm employed in [126, 63, 14, 48] (i.e., n-AFC) was different from that employed in audio (i.e., 2-AFC in [12, 93]), the way to compute mental representation is thereby different.

7. In terms of this step, it belongs to RevCor process. That is why we use green in the first step of our pipelines shown in Fig. 1.4 and 1.4 (indicating that this step comes from psychology). In terms of the corresponding technique, it can be achieved by the computer science technique, such as FEM. That is why we use blue to paint the first step of Fig. 2.4 and 2.6 (indicating that this step is achieved by computer science technique).

These works [126, 63, 14] usually computed the Pearson correlation on the labeled stimuli. That is to say, the label is the response of the participants to each randomly generated stimulus, such as the emotion and the intensity of this emotion (see n-AFC in Fig. 2.3).

Finally, after analyzing the computed mental representation, psychologists draw their conclusions.

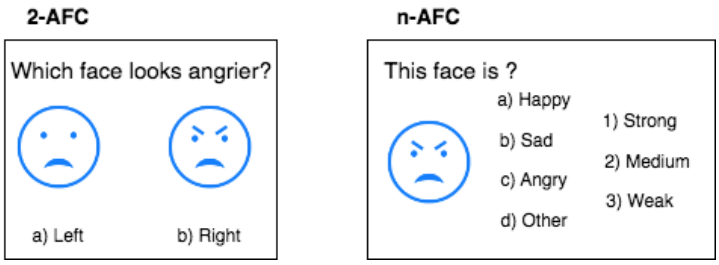


Figure 2.3 – Examples of left: 2-AFC and right: n-AFC (where n=4).

Table 2.5 – Related works using reverse correlation process for affective computing [126, 63, 14, 93, 12, 49]. We list in the first column: the stimuli category (denoted by stimuli), the reverse correlation paradigm (denoted by paradigm), the number of affective states (denoted by states), the number of trials performed by one observer for all affective states (denoted by trials/obs), the number of trials performed by one observer for one affective state (denoted by trials/obs/state) and the number of mental prototypes for one observer (denoted by proto/obs).

	stimuli	paradigm	states	trials/obs	trials/obs/state	proto/obs
Yu[126]	face	7-AFC	6	2400	n/a	single
Jack[63]	face	7-AFC	6	4800	n/a	single
Chen[14]	face	3-AFC	2	3600	n/a	single
Ponsot[93]	speech	2-AFC	2	n/a	∼700	single
Burred[12]	speech	2-AFC	1	n/a	700	single
Goupil[49]	speech	2-AFC	2	n/a	880	single

Generally, RevCor is widely employed to study the perception of faces [48, 63, 126, 14], speech [93, 49, 12] and bodies [64, 74]. Note that the works [64, 74] use RevCor to understand how humans identify gender via bodies, and the work [48] focuses more on identity, gender, with/without expression via faces. These works are far from the research on affective states.

In Table 2.5, we summarize the related works using RevCor for affective computing. The works [93, 49, 12] focus on the audio modality, and the works [126, 63, 14] focus on

the visual modality. Due to different paradigms, we detail the trials into two sub-classes: the trials performed by one observer for all affective states (corresponding to n-AFC) and the trials performed by one observer for one affective state (corresponding to 2-AFC). The difference between the two sub-classes can be explained in Fig. 2.3. For 2-AFC, one trial corresponds to one affective state (i.e., angry, displayed on the question); and for n-AFC where n=4, one trial corresponds to three affective states (i.e., happy, sad, and angry, displayed on the options). Note that the representative works highly related to this thesis are [126, 63, 14]. They focused on FE-based affective computing using RevCor. We detail them in the following paragraphs.

Yu et al. [126] proposed a tool to synthesize arbitrary facial expressions on virtual avatars. The synthesis was controlled by AUs on 3D avatars. The organizers hired 8 Western Caucasian observers and instructed them to categorize 2400 randomly generated animations as one of the six basic emotions. According to RevCor, they modeled the mental representation of each basic emotion for each observer. By evaluating the mental representations, the authors validated that the generated facial expressions genuinely conveyed the intended message.

Jack et al. [63] randomly generated 4800 trials. Each trial consists of one dynamic facial animation created by the 3D morphing tool of [126]. Each of the 15 Western Caucasian and the 15 East Asian observers was asked to categorize the animations into six basic emotion categories. The authors then used reverse correlation to extract one mental representation of each emotion for each cultural group and conclude that these representations were, in fact, not culturally universal. This work [63] can in principle produce control parameters for its generative model, i.e., generate a 3D synthetic face that maximizes the probability that a given observer judges it representative of one of the tested emotion categories.

Chen et al. [14] modeled the mental representations of dynamic facial expressions of pain and pleasure in 40 observers in each of two cultures (Western, and East Asian) using RevCor. Each observer completed 3600 trials, resulting in a set of facial animations for pain and pleasure. After analyzing the mental representations of these two non-basic facial expressions, the authors concluded that these two non-basic facial expressions were physically and perceptually distinct in each culture, thus proving that these two different and intense facial expressions were not virtually indistinguishable.

Considering the practicality (i.e., integrating into such an application context of Randstad), all these works [63, 126, 14] manipulated facial expressions on virtual avatars. However, we need to manipulate the facial expressions of any actors in real photographs.

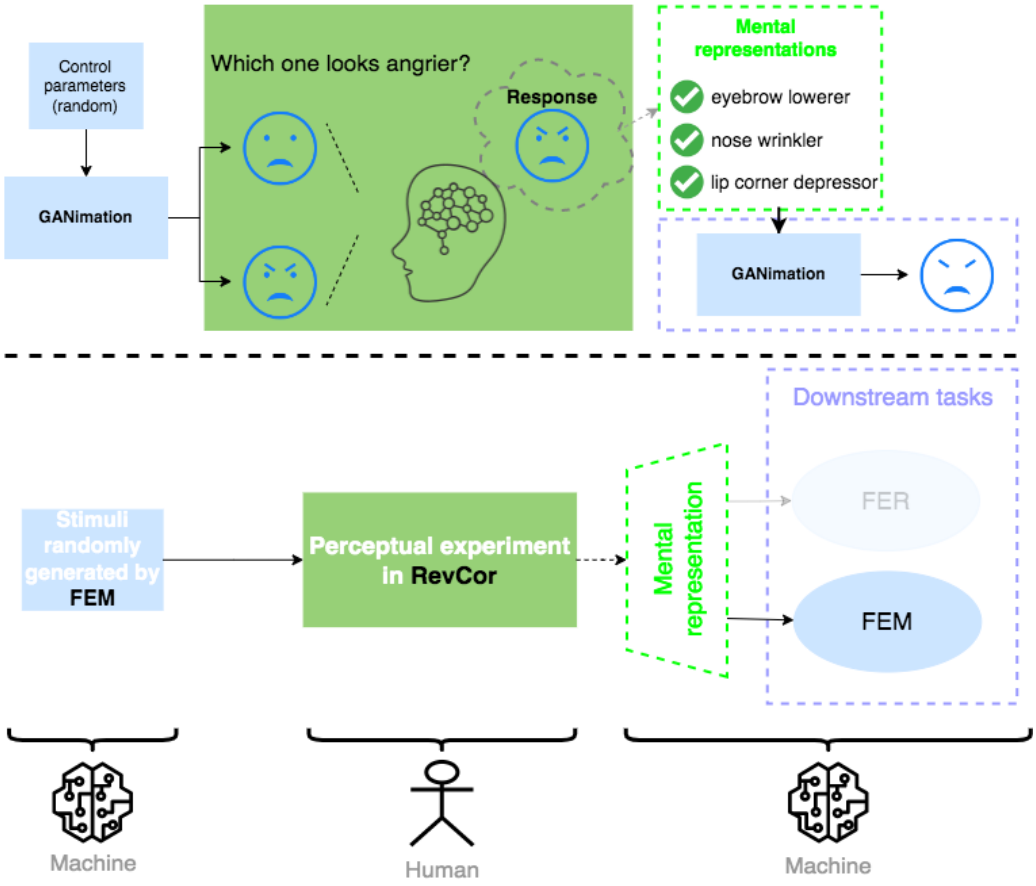


Figure 2.4 – Inspired by the reverse correlation process (RevCor) from psychology and the facial expression manipulation technique (FEM) from computer science, we come up with our first pipeline. **Top**: detail of the first pipeline. **Bottom**: the general pipeline (aforementioned in Chapter 1). We chose GANimation [94] as the FEM tool to generate stimuli. Stimuli were employed in RevCor. After the typical RevCor steps: perceptual experiment and mental representation computation, we employed the mental representation as control parameters only for the downstream task: FEM. Note that, to be consistent, the FEM technique used in this pipeline (for stimuli generation and the downstream task) is the same.

2.3.2 Discussion and conclusion

RevCor can reveal neural mechanisms and processing strategies during the perception of facial expressions. The mental representation (hereafter, mental prototype) can be regarded as the exclusive (or personalized, which can meet the need of the observer) prototype of the observer. Thus, the facial expression of a given emotion can be personalized. Mental prototypes can be employed as the personalized control parameters for the

downstream FEM task. That is to say, like using GANimation [94], we can activate the AUs corresponding to the mental prototype to generate the personalized facial expression. Moreover, by asking appropriate questions to the participants (e.g., "which face looks more self-confident?"), the mental prototype of any affective state (not only basic or non-basic emotions but also more general affective states such as the expression that the candidate wants to be seen in an interview) can be computed.

Hence, by integrating RevCor into FEM, we propose **the first pipeline** shown in Fig. 2.4: a GANimation-based model to personalize facial expressions controlled by RevCor. **For the first step of our pipeline, we choose GANimation [94] as a FEM tool.** Note that it can be replaced by other FEM tools if appropriate. **For the downstream task (i.e., the last step of our pipeline), we also choose FEM.** There are two reasons. 1) considering the fact (i.e., exhaustiveness) that non-basic emotions are not explicitly available in the existing database, there is no baseline for the FEM task (generating such non-basic emotions), and the FER task (recognizing such non-basic emotions). It will be difficult to evaluate the results and validate our approach. However, the generated facial expressions (via FEM task) of a non-basic emotion can be evaluated subjectively. 2) considering the application context of Randstad, FEM is required. That is why for the downstream task, we choose FEM rather than FER. To sum up, in the first step and the last step of our pipeline, we both chose FEM. **In order to be consistent, we employ the same FEM module twice**: 1) to generate stimuli for RevCor and 2) to generate personalized facial expressions controlled by the mental prototypes obtained from RevCor.

For the perceptual experiment, we chose the 2-AFC paradigm. There are two reasons. 1) considering the complexity of non-basic emotions [86, 22], the randomly generated stimuli can not frequently correspond to the target non-basic emotion. The observer will face the following situations. For the n-AFC paradigm (see Fig. 2.3), there can be a very large number of trials where the observer chooses the option "Other", i.e., not corresponding to the target non-basic emotion. Thus, **for n-AFC, a large number of trials can be not valuable.** 2) in the 2-AFC paradigm, comparing a pair of stimuli can obtain **relative information rather than absolute information** collected in the n-AFC paradigm based on annotating a single stimulus by a list of options for the n-AFC paradigm.

For the mental representation computation, we propose a statistical way to compute mental representation. Although we choose 2-AFC, we can not mimic the same way as the works [93, 12, 49] using 2-AFC in audio modality to compute mental

representations (see Table 2.5). The reason is that in audio modality, it is possible and meaningful subtracting two opposite options such as the pitch contour of the "trustworthy" and the "non-trustworthy" voices. However, the subtraction of the action unit is meaningless, e.g., it is unexplainable to assign the value "-1" to *AU12 lip corner puller*. **Thus, we propose a concept of dominant and complementary AUs to describe facial expressions.**

Overall, our first pipeline inherits the strengths of RevCor.

— **Flexibility.** Our approach can flexibly personalize facial expressions to fit the expectations of any observers.

— **Exhaustiveness.** Our approach allows subjective judgments on any emotion, including those not available in existing deep-learning databases.

— **Expertise-free.** Expert knowledge is not required. The only requirement is the observer's perception (i.e., judgment on intuition).

However, this pipeline also inherits the drawbacks of RevCor.

— **Efficiency.** It is less efficient. In the conventional RevCor process (see Table 2.5), hundreds or thousands of randomly generated trials are required to compute mental representation [48, 63, 126, 14, 12, 93]. Indeed, the study on speech intonation has already indicated that for some of their tasks, they could have reached the same precision with fewer trials (figure 6 in [12]). That is to say, some of the randomly generated trials are not necessary.

— **Diversity.** Through perceptual experiments, these studies provided one, and only one, mental prototype for one or a group of observers. However, there can be multiple mental prototypes for one or a group of observers. That is to say, the challenge of diversity remains unresolved.

To generate trials for RevCor in an efficient way and bring various mental prototypes, we get inspired by Interactive Genetic Algorithm (IGA).

2.4 Interactive Genetic Algorithm

In this section, we introduce Interactive Genetic Algorithm (IGA). By integrating this algorithm, we refine the first pipeline thus fulfilling all the aforementioned requirements and tackling all the challenges (see Chapter 1). Since IGA is a variant form of Genetic Algorithms, before introducing IGA used for affective computing, we briefly introduce the

traditional Genetic Algorithm (GA).

2.4.1 What is Genetic algorithm?

Like Darwin's theory of the survival of the fittest in nature, Genetic Algorithm (GA) is a well-known heuristic algorithm that mimics the biological evolution process [82]. The basic elements of GA are chromosome representation, fitness evaluation, and biological-inspired operators such as selection, crossover, and mutation [67]. Typically, the chromosomes are usually in binary string format. They are considered as the points in the solution space and are processed using genetic operators by iteratively updating their population. The fitness function is used to assign a value to each chromosome in the population. In the selection operator, some chromosomes are selected for further processing (e.g., crossover and mutation) based on their fitness values. In the crossover operator, some bits of chromosome pairs are exchanged to create offspring. In the mutation operator, some bits of chromosomes will be randomly flipped. Fig. 2.5 illustrates 1) Left: the traditional

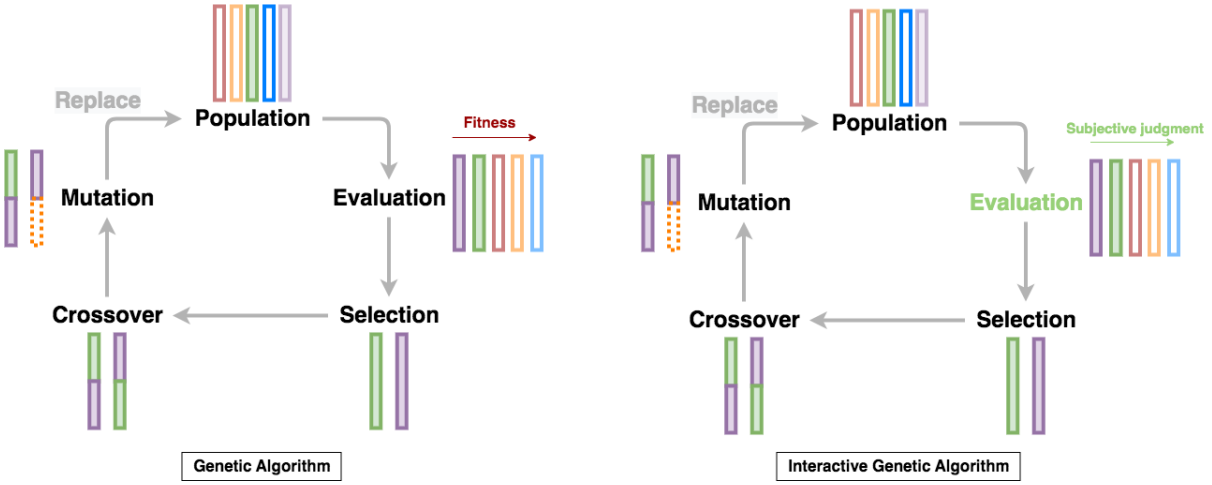


Figure 2.5 – **Left**: Demonstration of traditional Genetic Algorithm (GA) process. **Right**: Demonstration of Interactive Genetic Algorithm (IGA). The obvious difference between GA and IGA is that the fitness function of IGA is the subjective judgment of humans. There is no objective or mathematical fitness function in IGA.

GA process and 2) Right: the Interactive Genetic Algorithm (IGA). Differing from the traditional GA that uses a mathematical fitness function to evaluate chromosomes, IGA employs human judgments to evaluate chromosomes.

2.4.2 Interactive genetic algorithm for affective computing

As a typical Interactive Evolutionary Computation (IEC), Interactive Genetic Algorithm (IGA) optimizes the target system to fit one's preference based on subjective judgments [107]. Like the reverse correlation process, IGA requires subjective judgments and is capable of carrying several solutions. This technique is widely used in several domains: geology [91], design [68, 116, 101], image processing, such as image retrieval [16, 69] and 3D facial animations [53]. To the best of our knowledge, before the publication of our second pipeline in 2022, there is no literature using IGA to manipulate facial expressions for affective computing.

In 2022, after the submission of our work [121], another paper [8] was published using IGA in facial expression manipulation for experimental psychology. We will describe it below and then indicate the main difference between this work and our second pipeline.

This work [8] personalized facial expressions via virtual 3D avatars. In each iteration of the GA (called generation), the participant selects from the population a number of facial expressions most similar to some internalized target. Among an unconstrained number of selections, one elite face is selected as the best, and the faces also matched the target emotion are also selected (unconstrained number). There is no further fitness ranking of the remaining selected samples. The elite is guaranteed in the next generation. Two mechanisms for non-elite gene propagation are averaging, and the tandem of crossover and mutation. On one hand, the mean of two or more blend shape vectors of avatars is propagated to the next generation. On the other hand, crossover and mutation involve the substitution of randomly selected weights of one chromosome by those of another ("crossover") and the subsequent assignment of new random values to a fixed number of arbitrary genes in the chromosome ("mutation"). To maintain diversity and avoid premature convergence, the population at each iteration is boosted by 40% insertion of novel samples completely uncorrelated to prior user selections. After 10 generations, an elite face is selected and determined as the personalized facial expression.

Overall, [8] has four drawbacks compared to ours [121] (i.e., the second pipeline that we will present in Section 2.4.3.). 1) this work [8] provides a solution according to the drawback of the efficiency of RevCor, whereas **the drawback of the diversity is still unsolved**. Although GA can provide multiple solution, in the last iteration of this work, only the elite face (i.e., the only one solution) is selected as the personalized facial expressions. 2) considering the challenge of exhaustiveness, **the non-basic emotion is not addressed.** The prototypes of non-basic emotion or called complex emotion will be more varied (non-

universal) and more difficult to be determined [61, 14, 86]. 3) similar to the traditional RevCor, the number of experimental trials (empirically 10 iterations, i.e., called trials [8] in this work, for a population of 10 individuals) is still **fixed**. This might be less flexible and obscure the optimal solution even when the participant only performed 10 trials. 4) in addition, processing control parameters of **the virtual avatar** (i.e., blend shape vectors) by averaging or crossover-and-mutation may activate too many blend shapes and **increase visual artifacts**.

2.4.3 Discussion and conclusion

As a closed-loop optimization algorithm, IGA can be adaptive to successive ratings by human observers. Indeed, like RevCor, IGA considers subjective judgments, but contrary to RevCor, it can offer multiple solutions. Hence, the reverse correlation process (RevCor) can be embedded into an IGA system to generate more user-preferred trials that are also more correlated to target tasks, thus reducing the workload of human observers and obtaining multiple mental prototypes. Such a system will address the aforementioned challenges and fulfill the aforementioned requirements (in Chapter 1). Moreover, such a system will largely reduce the time of human involvement and have more intelligent and flexible controls. That is to say, unlike the related work [8], 1) the system should reasonably determine the termination of the process rather than setting a fixed number of iterations and also 2) reduce visual artifacts during facial expression editing.

For these purposes, we propose **the second pipeline, Interactive Microbial Genetic Algorithm (IMGA)**, that refines the first pipeline by integrating IGA. Differing from the traditional genetic algorithm that needs to acquire the fitness values of all individuals, we added a population evaluation module that evaluates the quality of the entire population of each generation (i.e., iteration) thus monitoring the convergence of the system. In addition, we added a three-state constraint automaton to gradually increase the number of activated facial action units (AUs) [37] for each face and determine the process's termination. That is to say the population evaluation module and the three-state constraint automaton in our pipeline avoid the aforementioned two drawbacks (controlling the termination of the system and reducing artifacts) that appeared in the related work [8].

To fulfill all the requirements and address all the challenges aforementioned in Chapter 1,

8. In this work [8], one iteration (generation) contains only one trial. However, in our pipeline, one iteration contains 10 trials. For more information, please see Chapter 4.

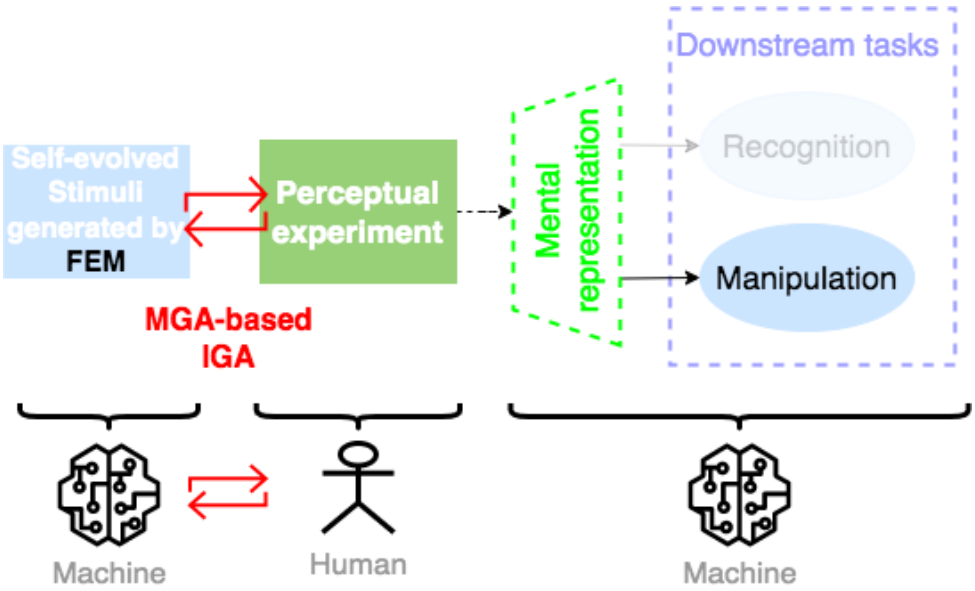


Figure 2.6 – We propose the second pipeline: an optimization process that embedded RevCor into an MGA-based interactive genetic algorithm (IGA). This pipeline not only inherits the strengths from the first pipeline (i.e., flexibility, exhaustiveness, and expertise-free) but also solves the drawbacks of the first pipeline (i.e., efficiency, diversity).

such an optimization process that embedded RevCor into an interactive genetic algorithm (IGA), not only inherits the strengths from the first pipeline but also solves the drawbacks of the first pipeline (see Fig. 2.6).

— **The inherited strengths.**

1. **Flexibility.** Everyone can use our approach to extract their own mental prototypes of a given emotion. Thus facial expressions can be personalized.
2. **Exhaustiveness.** The categories of emotions are not limited to those provided in existing deep-learning databases.
3. **Expertise-free.** Our approach only requires the observer's perception (i.e., judgment on intuition) rather than the observer's expertise (e.g., no expert knowledge in affective computing, psychology, or certified FACS coders [37]).

— **The solved drawbacks.** This pipeline becomes efficient and can bring diverse solutions.

1. **Efficiency: by the online feedback loop.** Unlike the traditional RevCor that generates massive trials randomly, in our approach, based on the observer's feedback, automatically updated trials can contain more valuable information

(closer to the mental prototypes of observers).

2. **Efficiency: by acceleration.** The way of generating trials for mental prototype computations is intelligent. Moreover, we adopt the microbial genetic algorithm (MGA) [55] as the GA module within IGA to further accelerate the convergence of the system. Indeed, the GA with an elitist mechanism (like MGA) can converge faster than that without an elitist mechanism [67].
3. **Diversity.** Benefiting from IGA, for one emotion, this pipeline can provide multiple mental prototypes to each observer. That is to say multiple solutions (i.e., prototypes) even for one observer.

In the next two chapters, we introduce in detail the first pipeline and the second pipeline.

Chapter 3

The first pipeline: Mental Deep Reverse-Engineering System (MDR)

Inspired by the mechanism of how psychologists analyze facial expressions, we propose our first pipeline: Mental Deep Reverse-Engineering System (MDR). As shown in Fig. 3.1, it is a novel interdisciplinary approach that combines the recent deep learning technique from computer science, i.e., Generative Adversarial Network [45], with psychophysical reverse correlation (RevCor), a recently emerging technique from psychology.

This approach can meet the first three requirements aforementioned in Chapter 1. By extracting the mental prototype of the desired facial expression from the observer, this approach can personalize facial expressions (i.e., specific to the observer), thus meeting **the Requirement #3 (mentioned in Section 1.1.3).** The personalized facial expressions are not limited to the basic emotions of Ekman. They can be complex emotions or social attitudes such as self-confidence. This can meet **the Requirement #1 (mentioned in Section 1.1.3).** The entire process is expertise-free. That is to say, no experts are required such as FACS coders and psychologists. Subjective judgment (intuition/perception of observers) is all we need. This can meet **the Requirement #2 (mentioned in Section 1.1.3).**

As explained in Chapter 2, we use GANimation[94] as the tool for facial expression manipulation task (FEM) that can flexibly generate a wide variety of facial expressions by controlling the objective low-level attribute, i.e., action unit. Moreover, we use the same FEM tool (GANimation) twice: once for extracting the mental prototype during the RevCor process and another time for manipulating facial expressions based on the mental prototype. This can ensure that the manipulation is consistent with the mental prototype of the observer. To enhance the definition of facial expression prototypes, we introduce the concept of dominant and complementary action units to precisely describe

facial expression prototypes.

The organization of this chapter is as follows. We introduce the methodology of MDR in Section 3.1, then detail the experiments in Section 3.2, and next evaluate the results and discuss the convergence efficiency in Section 3.3, and draw conclusions in Section 3.4. All the information about this pipeline can be found at https://yansen0508.github.io/emotional-prototype/. The code is available at https://github.com/yansen0508/Mental-Deep-Reverse-Engineering.

3.1 Methodology

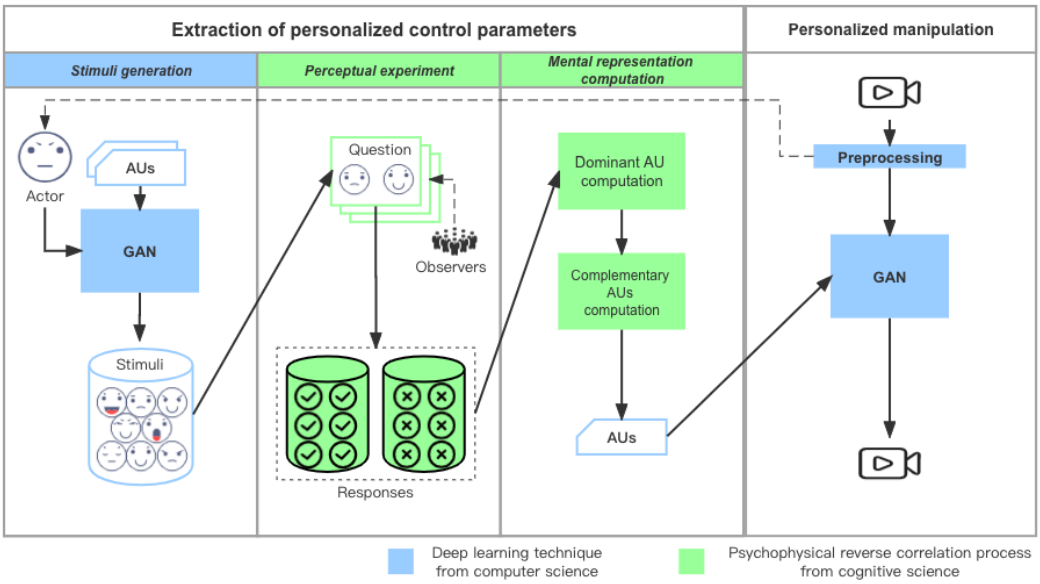


Figure 3.1 – **Framework of our approach to personalize facial expressions.** We combine the recent deep learning technique, i.e., Generative Adversarial Network (highlighted in blue), with psychophysical reverse correlation, a recently emerging technique from psychology (highlighted in green). We employ the same GAN to extract personalized control parameters (i.e., mental representation) and to personalize facial expressions of any emotion, including those not available in existing deep-learning databases. The only requirement of our approach is the observer's perception rather than expertise (such as certified FACS coders [37]). We also introduce the concept of dominant and complementary action units to describe facial expressions.

Our approach is composed of four successive steps (in Fig. 3.1). In the first step (in Section 3.1.1), based on the real face of an actor, a generative model (denoted by GAN) is

applied to synthesize a large number of arbitrary facial expressions (i.e., stimuli for reverse correlation process). Then (in Section 3.1.2), an observer performs a perceptual experiment of RevCor in which the inputs are the generated stimuli. Next (in Section 3.1.3), from all the answers of the observer, we compute the dominant AU and the complementary AUs and then construct the mental representation (i.e., personalized control parameters). Finally (in Section 3.1.4), according to the mental representation, we employ the same generative model (i.e., GAN) to generate the personalized facial expression that meets the observer's expectation.

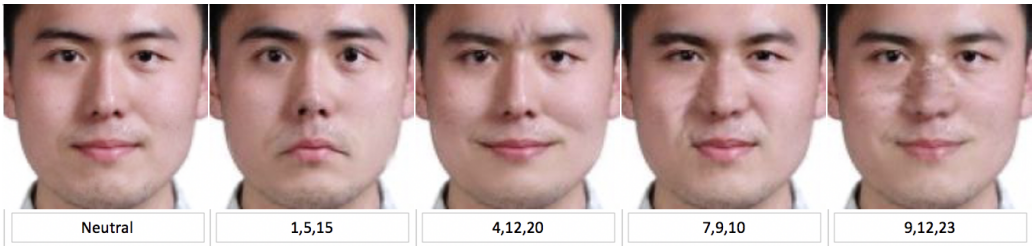


Figure 3.2 – **Examples of stimuli for reverse correlation process.** Based on a single neutral face (on the left), we randomly activate 3 AUs to generate stimuli presented to observers. Since each stimulus is randomly generated, the facial expression does not have to correspond to any emotional state, such as the third stimulus with the activation of AU4 (brow lowerer), AU12 (lip corner puller) and AU20 (lip stretcher).

3.1.1 Stimuli generation

To generate input stimuli (random facial expressions) for RevCor, we can employ any tool that can control low-level attributes. Here, we choose GANimation [94] controlled by facial action units (AUs) [37] to synthesize random facial expressions (i.e., stimuli for RevCor). In this step, GANimation (thereafter, G) takes as input an image of the actor's face S (e.g. captured with an emotionally neutral expression) and a n-dimensional binary vector v of AUs to create a deformed face (i.e., stimulus) $I{=}G(S, v)$.

GANimation is capable of manipulating $n = 16$ AUs [1]. We define as $v{=}[\lambda_1, ..., \lambda_n]$, the binary AU vector where each component λ_i represents the activation ($\lambda_i = 1$) or deactivation ($\lambda_i = 0$) of $AU_{\mu[i]}$ (the i_{th} element in the AU list μ). For instance, $\lambda_3 = 1$ represents that AU4 (brow lowerer) is activated, $\lambda_9 = 0$ represents that AU12 (lip corner puller) is deactivated. See the literature [37] for a complete list of AUs.

1. from the list $\mu{=}\{1, 2, 4, 5, 6, 7, 9, 10, 12, 14, 15, 17, 20, 23, 25, 26\}$

(a) AU1, AU5, AU6, AU7, AU9 (b) AU10, AU12, AU15, AU23, AU25

Figure 3.3 – Visual artifacts generated by GANimation [94]. Left: AU1, AU5, AU6, AU7, AU9. Right: AU10, AU12, AU15, AU23, AU25.

While GANimation can simultaneously activate AUs, activating too many AUs typically will create visual artifacts (see Fig. 3.3 and Appendix Fig. 3). Therefore, we generate stimuli by activating 3 AUs: $\forall v, \sum_{i=1}^{16} \lambda_i = 3$. Combining 3 out of $n = 16$ AUs, there can be $C_{16}^{3} = 560$ possible AU vectors $\mathcal{V} = \{v_1, v_2, ..., v_{560}\}$, where C is the mathematical combination function. We note $\Phi = \{I{=}G(S, v) \,|\, \forall v \in \mathcal{V}\}$ the set of all the possible stimuli (that have 3 AUs activated) generated by GANimation based on face S. Fig. 3.2 shows some examples of randomly generated stimuli by GANimation [94].

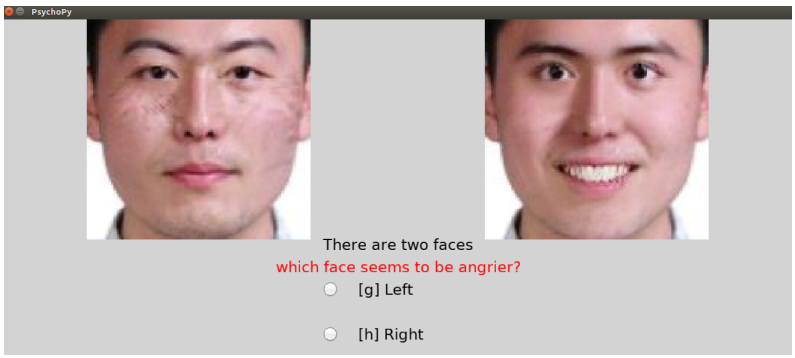
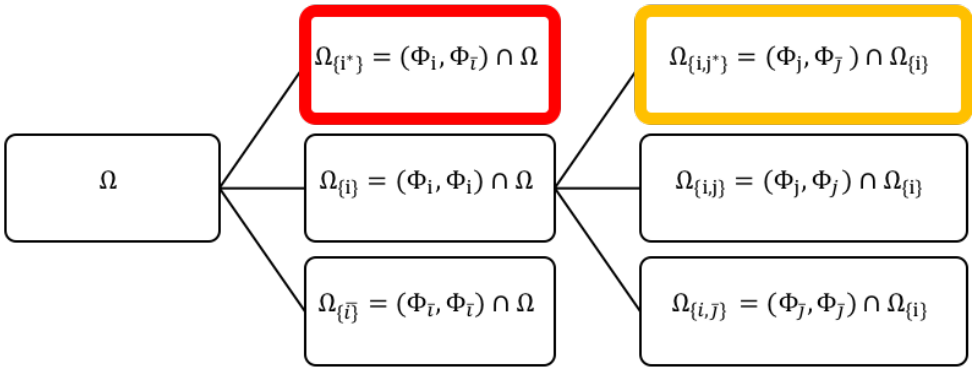


Figure 3.4 – Interface used in the perceptual experiment. It is implemented by PsychoPy.

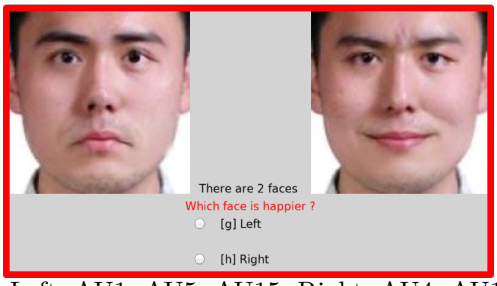
3.1.2 Perceptual experiment

The second step of our approach is the perceptual experiment. In each trial of the perceptual experiment, a pair of randomly generated stimuli is presented to the observer. The observer is asked to choose which stimulus of the given pair best corresponds to the target expression (e.g., "There are two faces, which face seems to be angrier?", shown in Fig. 3.4). For each perceptual experiment, observers perform m trials and each pair of randomly generated stimuli is displayed only once. Note that there are $C_{560}^{2} \approx 1.56 \times 10^{5}$

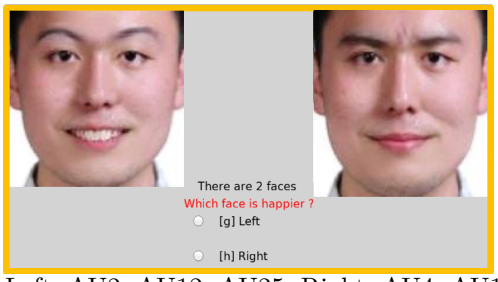
possible combinations if we randomly select a pair of 3 AU-activated stimuli from the set Φ. We define the set of all m trials (in one perceptual experiment, $m \ll 1.56 \times 10^5$) for a given observer as Ω, and we use $s_I \in \{0, 1\}$ to annotate that the observer selected ($s_I = 1$) or did not select ($s_I = 0$) the stimulus I. Hence, the set of all selected stimuli from the set of all the trials in one perceptual experiment Ω is defined as $Z_\Omega = \{I | s_I = 1\}$, where $|Z_\Omega| = m$. Note that $|.|$ represents the cardinality of the set.



(a) Subsets of trials for dominant and complementary AU computations.



(b) Left: AU1, AU5, AU15. Right: AU4, AU12, AU20.



(c) Left: AU2, AU12, AU25. Right: AU4, AU12, AU20.

Figure 3.5 – 3.5(a): Subsets of stimuli for dominant and complementary AUs computations. Dominant AU computation: see the middle column highlighted in red. Complementary AUs computation: see the right column highlighted in yellow. Based on 3.5(a), we list two examples for the dominant and complementary AUs computations of happiness (correspondingly highlighted in red and yellow), where $i = 12$ (AU12: lip corner puller) and $j = 25$ (AU25: lips part). 3.5(b): The trial from the set $\Omega_{\{12^*\}} = (\Phi_{12}, \Phi_{\overline{12}}) \cap \Omega$ will be used to verify if AU12 is the dominant AU of happiness. 3.5(c): With the premise that AU12 is the dominant AU of happiness, the trial from the set $\Omega_{\{12,25^*\}} = (\Phi_{25}, \Phi_{\overline{25}}) \cap \Omega_{\{12\}}$ will be used to verify if AU25 is the complementary AU of happiness.

3.1.3 Mental representation computation

We define the set Φ_i of all stimuli in which AU_i is activated: $\Phi_i = \{I{=}G(S,v)|\forall v \in \mathcal{V}, \lambda_k{=}1, \mu[k] = i\}$, and $\Phi_{\bar{i}}$ the set of all stimuli in which AU_i is deactivated: $\Phi_{\bar{i}} = \{I{=}G(S,v)\,|\,\forall v{\in}\mathcal{V}, \lambda_k{=}0, \mu[k] = i\}$. Thus, for each AU_i, the perceptual experiment Ω can be divided into three subsets shown in Fig. 3.5(a):

— $\Omega_{\{i^*\}}$: the subset of trials in which one of the paired stimuli has AU_i activated and another one has AU_i deactivated.

— $\Omega_{\{i\}}$: the subset of trials in which both stimuli have AU_i activated.

— $\Omega_{\{\bar{i}\}}$: the subset of trials in which both stimuli have AU_i deactivated.

Dominant action unit computation

We first define $Z_{\Omega_{\{i^*\}}}$ the set of stimuli selected in subset $\Omega_{\{i^*\}}$ (Fig. 3.5(a)-red). We then count $P(i|\Omega_{\{i^*\}})$ the proportion of the selected stimuli that have AU_i activated in the subset $\Omega_{\{i^*\}}$, i.e. how likely an activated AU_i is to drive the observer's perception. Note that $|.|$ represents the cardinality of the set:

$$P(i|\Omega_{\{i^*\}}) = \frac{|Z_{\Omega_{\{i^*\}}} \cap \Phi_i|}{|Z_{\Omega_{\{i^*\}}}|} \tag{3.1}$$

Finally, we can determine the action unit AU_i with the largest proportion $P(i|\Omega_{\{i^*\}})$ as the dominant action unit denoted by AU_d. d is the subscript number of dominant AU.

$$d = \arg\max_i P(i|\Omega_{\{i^*\}}) \tag{3.2}$$

Complementary action units computation

We define $\Omega_{\{d\}}$ as the subset of trials where both stimuli have dominant AU_d activated. We continue to divide subset $\Omega_{\{d\}}$ into three subsets according to the activation status of other AUs, as shown in Fig. 3.5(a). Knowing that $\forall AU_j \neq AU_d$, we compute $P(j|\Omega_{\{d,j^*\}})$ the proportion of selected stimuli in subset $\Omega_{\{d,j^*\}}$ that have AU_j activated (Fig. 3.5(a)-yellow), i.e. how likely the addition of AU_j to dominant AU_d is to drive the observer's perception:

$$P(j|\Omega_{\{d,j^*\}}) = \frac{|Z_{\Omega_{\{d,j^*\}}} \cap \Phi_j|}{|Z_{\Omega_{\{d,j^*\}}}|} \tag{3.3}$$

In practice, we limit the number of complementary action units $c \in \mathcal{C}$ by introducing a

threshold T_q (to separate complementary AUs and non-complementary AUs):

$$\mathcal{C} = \{j | P(j|\Omega_{\{d,j^*\}}) \geq T_q\} \tag{3.4}$$

The output of this step is the mental representation (i.e., personalized control parameters for facial expression manipulation) which is a n-dimensional binary AU vector v_m for the observer. Within v_m, both dominant AU (AU_d) and complementary AUs ($\{AU_c, \forall c \in \mathcal{C}\}$) are activated: $v_m = \{\lambda_i = 1 | \forall i \in [1, 16], c \in \mathcal{C} : \{\mu[i] = d\} \cup \{\mu[i] = c\}\}$.

3.1.4 Personalized manipulation

Once the mental representation v_m of the observer is extracted, we apply the personalized manipulation on each frame with $I_i = \{G(S_i, v_m)\}$. To be consistent with the mental representation and the final manipulation, we employ the same tool (i.e., GANimation [94]) for the stimuli generation and the personalized manipulation. To make the video compatible with GANimation (especially with the dimension of the face S_i), we crop, align and resize the face S_i of the actor in each frame.

3.2 Experiment results

We list the implementation details and detail the experimental protocol in Section 3.2.1. For the results, we adopt an example from an observer to illustrate and discuss the dominant and complementary AUs computation in Section 3.2.2, then list and discuss all personalized prototypes and the corresponding manipulations in Section 3.2.3.

3.2.1 Experiment settings

Implementation

GANimation model. We choose GANimation [94] as the tool to generate facial expressions by editing the relatively objective low-level attribute, i.e., action units (AUs) [37]. We use the code of GANimation released by its authors. All settings are unchanged. The input image and the output image are $148px \times 148px$. To crop, align, and resize the face, we employ OpenFace [4].

Mental representation computation. As aforementioned, we determine a dominant AU for one emotion as the action unit that dominantly drives the observer's perception.

For the complementary AUs, we need to determine which AUs combined with the dominant AU have a significant effect on driving the observer's perception. Therefore, we need to set a relatively high threshold T_q to eliminate most AUs with less significant proportions $P(j|\Omega_{\{d,j^*\}})$. Indeed, $T_q = 50\%$ corresponds to the situation in which, among each pair of AU_d-activated stimuli, the observer selects as many AU_j-activated stimuli as AU_j-deactivated stimuli. This means that AU_j carries no information content for this experimental task. Thus the threshold T_q should be higher than 50%. Considering that state-of-the-art prototypes [37, 126] have 2 to 5 AUs activated, we manually set the threshold T_q to 80% for happiness, sadness and anger and 70% for self-confidence.

Experimental protocol

Observers. Four observers (one female) participated in the perceptual experiment, all relatively young (mean=27.7yo) adults of three cultural groups: Brazil (1), China (2), and France (1), respectively denoted by observers #1 to #4. Only one observer has experience in affective computing, and nobody is a certified coder in Facial Action Coding System [37] or a psychologist. Each observer signed informed consent, and the experimental data were anonymous.

Perceptual experiment. The perceptual experiment aims to illustrate that our approach can personalize the facial expressions of a given emotion, even though this emotion is not available in existing deep-learning databases. Note that the purpose of the perceptual experiment is not to give extensive results or to discuss facial expression prototypes. We chose three basic emotions (happiness, sadness, and anger) that existed in deep-learning databases and one non-basic emotion (confidence) that is not explicitly available in existing deep-learning databases. Each of the four observers participated in four different experimental tasks to extract his/her mental representation of happiness, sadness, anger, and confidence. Considering the related works using reverse correlation [63, 14, 93, 12], the average number of trials for the perceptual experiment (of one emotion) varies from 700 to 1800. For each experimental task, we decided that observers performed $m = 840$ trials. The question was fixed and unique, e.g., "Which of these two faces looks happier?" The order of the four experimental tasks was counterbalanced among observers, and all experimental tasks used the same actor's photograph. It took about 40 to 60 minutes for one observer to complete a task (about 16 hours for all the perceptual experiments). The time interval between experimental tasks was set to half a day. All experiments were conducted in a quiet room in the lab, using a custom computer graphic interface

implemented by PsychoPy (Fig. 3.4).

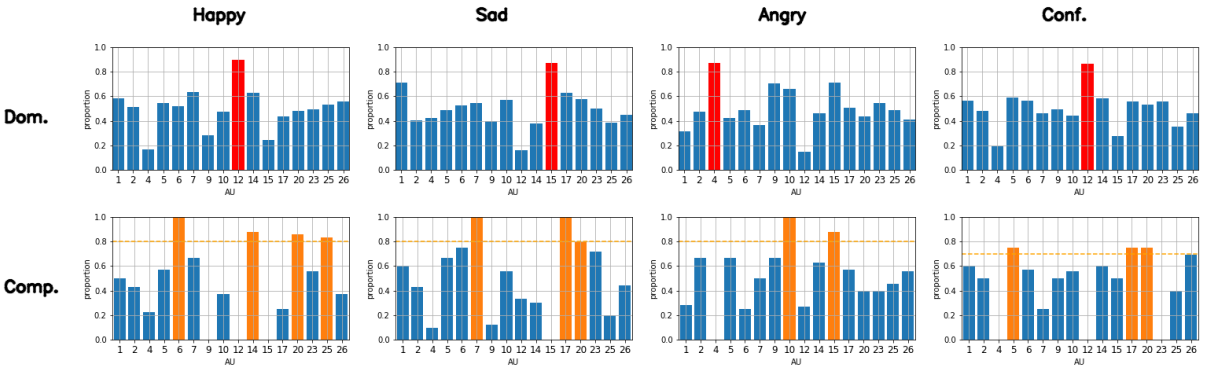


Figure 3.6 – Dominant and complementary AUs computation from observer #2. Each chart column lists the dominant AU computation (denoted by "Dom.") and the complementary AUs computation (denoted by "Comp.") of each emotion. In each chart, the proportion for each AU is computed based on the corresponding subset of trials. We highlight the dominant AU in red and the complementary AUs in yellow. The thresholds for the complementary AUs computation are marked by yellow dashed lines.

3.2.2 Results: dominant and complementary AUs computation

Fig. 3.6 details dominant and complementary AUs computations of each emotion (happy, sad, angry, and confident) from observer #2. See Appendix Fig. 4 from other observers. As mentioned in Mental representation computation, the proportion of each AU is computed based on the corresponding subset of trials. The dominant AU computation considers the subset $\Omega_{\{i^*\}}$ for a given AU_i (see Fig. 3.5(a)-red, and Eq. 3.1 and 3.2). The complementary AUs computation considers the subset $\Omega_{\{d,j^*\}}$ for a given non-dominant AU_j (see Fig. 3.5(a)-yellow, and Eq. 3.3 and 3.4). Since $\forall AU_j \neq AU_d$, the proportion for AU_d (the dominant AU) always equals zero in the chart for complementary AUs computation.

The concept of dominant and complementary AUs contains more information about emotional prototypes than a list of activated AUs in the universal prototypes [37]. Here are our observations.

The dominant AU drives the observer's perception. As defined in Mental representation computation, the dominant AU is the AU_i with the largest proportion $P(i|\Omega_{\{i^*\}})$, where $i \in \mu$. As shown in the first row of Fig. 3.6, the corresponding proportions exceed 80%. This means the observer has a significant probability to choose the facial

expression that has the dominant AU activated.

We can observe the dependency between the dominant AU and the complementary AUs. The charts on the first row of Fig. 3.6 illustrate that the corresponding complementary AUs have much lower proportions than that of the dominant AU. This indicates that a single complementary AU can not drive the observer's perception as much as the dominant AU. For the complementary AU computation shown on the second row of Fig. 3.6, when the dominant AU is activated, the facial expressions that have the complementary AUs activated have a significant probability of being selected by the observer. This means that complementary AUs can drive the observer's perception only in combination with the dominant AU.

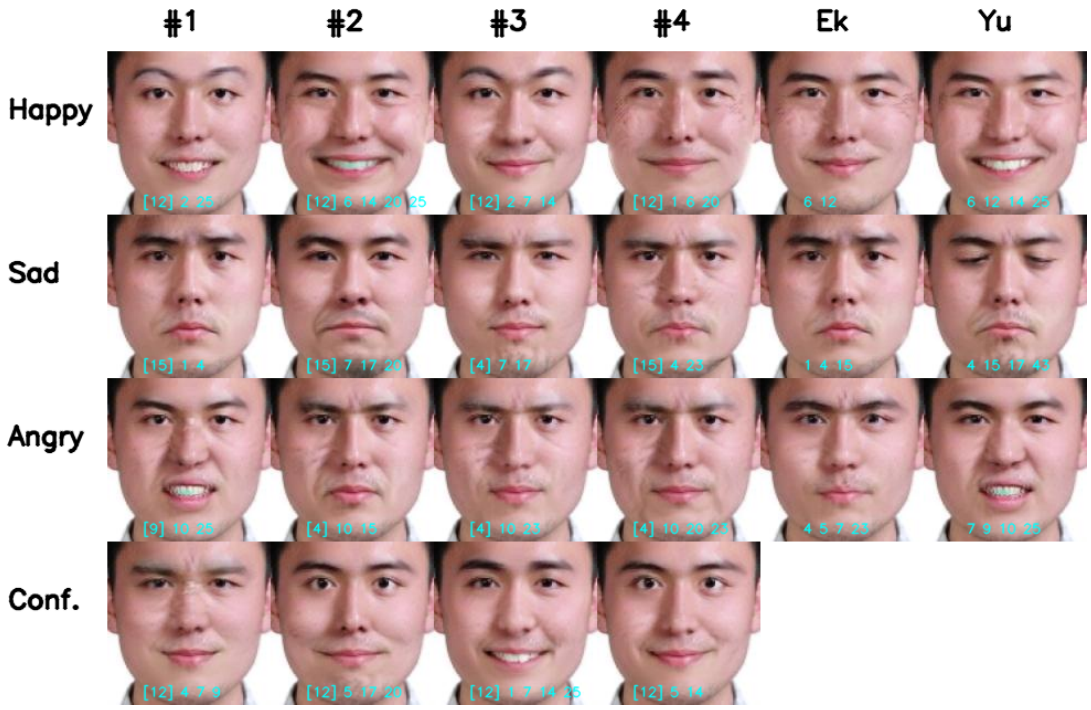


Figure 3.7 – Personalized prototypes (from observers "#1" to "#4") and state-of-the-art prototypes (denoted by "Ek" and "Yu") [37, 126] of happiness, sadness, anger, and confidence. For each personalized prototype, we detail the dominant AU in square brackets; the others are complementary AUs. For the state-of-the-art prototypes, we list the activated AUs. All facial expressions are reconstructed by the same GAN and the same actor. Note that GANimation [94] can not edit AU16 (lower lip depressor). We replace AU16 with AU25 (lips part) to reconstruct the prototype of anger from "Yu".

3.2.3 Results: personalized prototypes

We reconstruct the personalized prototypes of each observer by activating the dominant and complementary AUs. Fig. 3.7 shows the personalized prototype of each observer, as well as state-of-the-art prototypes from the literature [37, 126] (denoted by "Ek" and "Yu"). The corresponding activated AUs are listed at the bottom of the faces. For the personalized prototypes, we highlight the dominant AU in square brackets; the others are the complementary AUs. Note that there is not any published database of confidence, and there is no state-of-the-art prototype for confidence.

According to the results in Fig. 3.7, we observe that **the personalized prototypes are compatible with state-of-art prototypes**. All 12 prototypes of basic emotions generally convey expressions similar to that of Ekman [37] and Yu [126]. All dominant AUs: AU12 for happiness, AU4 or AU15 for sadness, and AU4 or AU9 for anger, can be found in state-of-the-art prototypes.

The prototypes are personalized. In each perceptual experiment task, although observers were asked the same questions, all observers acquired subtly different mental representations and, especially, different complementary AUs. With the exception of observer #1-sad and Ekman [37], all synthesized facial expressions also differed from the state-of-the-art prototypes by at least one AU. For instance, observer #2-happy is the same as Yu [126], plus the addition of AU20 (lip stretcher), resulting in a wider smile.

Our manipulations can be extended to the emotions that are not available in existing databases, such as confidence. We had no comparison prototypes for confidence. Although all confidence manipulations had the same dominant AU12 as happiness, the expressions remained different from any of the listed prototypes of happiness, notably because of the involvement of AU4 (brow lowerer), AU5 (upper lid raiser), AU9 (nose wrinkler) and AU17 (chin raiser).

3.3 Evaluations and discussion

Here, we evaluate the personalized prototypes to prove the validity of our approach. That is to say, our approach can meet the raised requirements. In Section 3.3.1, we conduct a subjective evaluation with the people who participated in the perceptual experiment, i.e., observers. The purpose of this evaluation is to assess the satisfaction of each observer with the prototypes they have created. In Section 3.3.2, we conduct another subjective evaluation by a diverse group of people who are not observers (so-called non-observers).

Table 3.1 – Subjective evaluation by observers. Each observer rated their personalized prototypes. The mean opinion score for each emotion is shown in the last column. Observers were satisfied with their personalized prototypes.

emo. \ obs.	#1	#2	#3	#4	mean
Happy	4	5	5	4	4.5
Sad	5	5	4	4	4.5
Angry	5	4	4	4	4.25
Conf.	4	5	4	5	4.5

Based on the perception of non-observers, we aim to quantify the acceptance of the personalized prototypes and compare them with the state-of-art prototypes [37, 126]. Note that the purpose of this pipeline is not to the extensive discussion of prototypes, such as their impact on affective states across different cultures, but only to validate the effectiveness of the procedure.

3.3.1 Subjective Evaluation by observers

To know the satisfaction of observers with their own personalized facial expression prototypes, we employed the Mean Opinion Score, which is a very popular indicator of perceived media quality [103]. We asked the four observers to rate their personalized facial prototypes from 1 to 5 representing bad satisfaction to excellent satisfaction. In Table 3.1, we listed all the scores rated by observers (denoted by #1, #2, #3, and #4) and presented the mean opinion score for each emotion (happy, sad, angry and confident) in the last column.

All observers rated their personalized facial prototypes with "4" and "5", meaning that each observer was quite satisfied with his/her personalized facial prototypes. **This indicates that the personalized prototype can reflect the observer's mental image and well answer the question in the perceptual experiment.**

3.3.2 Subjective Evaluation by non-observers

To quantify the acceptance of the personalized prototypes, we added the state-of-art prototypes of Ekman and Yu [37, 126] as the baseline of this evaluation process. We asked 217 anonymous participants from Amazon Mechanical Turk (AMT) to rank the prototypes listed in each row in Fig. 3.7 and we analyzed their answers by the Schulze method [98].

Subjective evaluation protocol

In more detail, each participant performed 4 ranking tasks. Each task corresponds to one of the four emotions. In each ranking task, all prototypes (6 listed prototypes for happy, sad, and angry, 4 listed prototypes for confident as in Fig. 3.7) were presented in shuffled order. Participants were asked to rank these faces from the happiest / saddest / angriest / most confident to the least.

To analyze the rankings from all the AMT participants, we applied the following two steps. 1) We first counted for each possible pair of prototypes how many participants preferred one of the prototypes over the other. 2) We then employed the Schulze voting method [98] to compute the preferences between each pair of prototypes and to derive the final ranking of these prototypes. Indeed, in the first step, according to the ranking, some preferences can be cyclic (similar to the game rock paper scissors, where each hand shape wins against one opponent and loses to another one). Thus, we can not directly quantify the acceptance between these prototypes. That's the reason why we employ the Schulze method to compute the preferences in the second step. Finally, we compared personalized prototypes with state-of-the-art prototypes. The voting results can be found in Appendix Fig. 5. For more details about the Schulze voting method, please see the literature [98].

Results: subjective evaluation

In Table 3.2, we present the preferences between each pair of prototypes computed by the Schulze method and the final rankings. Due to cyclic preferences, such as "#2", "#4", and "Yu." for sadness, the sum of the paired preferences is not always equal to 100%. For instance, in Table 3.2(b), the sum of the preference for "#4" over "Yu." (57%) and the preference for "Yu." over "#4" (51%) is 108%. Considering state-of-the-art prototypes as the baselines and for these 217 participants, our observations are as follows.

— **The low-ranking personalized prototypes are about equally preferred to at least one of the state-of-the-art prototypes.** "#3", "#4", and "Ek." in Table 3.2(a) and 3.2(c), and "#2", "#3", and "Yu." in Table 3.2(b) are low-ranking (ranked in the last three). The paired preferences between them are around 50%. For instance, in Table 3.2(a), 54% of the participants preferred "#4" to "Ek.", and 46% of the participants preferred "Ek." to "#4". That is to say, these low-ranking personalized prototypes are about equally preferred to the state-of-the-art prototype ("Ek." or "Yu."). This also validates that our approach can generate personalized

Table 3.2 – Preferences between each pair of prototypes computed by the Schulze method and the final rankings. The personalized prototypes of the corresponding observers are denoted by "#1" to "#4". "Ek." and "Yu." refer to state-of-the-art prototypes [37, 126]. Since the sadness prototypes of observer #1 and "Ek." are identical, we merged their preference data and denoted them by "#1/Ek.". For each pair of prototypes, we highlight the larger preferences in bold. For instance, for happiness, 84% of the participants preferred "#2" to "#1", whereas the preference for "#1" over "#2" is 16%.

(a) happiness

For \ Over	#1	#2	#3	#4	Ek.	Yu.	ranking
#1	-	16%	**75%**	**70%**	**82%**	**63%**	2
#2	**84%**	-	**82%**	**87%**	**94%**	**87%**	1
#3	25%	18%	-	**54%**	**52%**	24%	4
#4	30%	13%	46%	-	46%	25%	6
Ek.	18%	6%	48%	**54%**	-	15%	5
Yu.	37%	13%	**76%**	**75%**	**85%**	-	3

(b) sadness

For \ Over	#2	#3	#4	#1/Ek.	Yu.	ranking
#2	-	**52%**	51%	48%	51%	4
#3	48%	-	42%	38%	46%	5
#4	**52%**	**58%**	-	49%	**57%**	2
#1/Ek.	**52%**	**62%**	**51%**	-	**64%**	1
Yu.	**52%**	**54%**	51%	36%	-	3

(c) anger

For \ Over	#1	#2	#3	#4	Ek.	Yu.	ranking
#1	-	**70%**	**76%**	**78%**	**76%**	**84%**	1
#2	30%	-	**78%**	**72%**	**73%**	45%	3
#3	24%	22%	-	**52%**	**51%**	33%	4
#4	22%	28%	48%	-	**54%**	36%	5
Ek.	24%	27%	49%	46%	-	34%	6
Yu.	16%	**55%**	**67%**	**64%**	**66%**	-	2

(d) self-confidence

For \ Over	#1	#2	#3	#4	ranking
#1	-	46%	38%	46%	4
#2	**54%**	-	26%	44%	3
#3	**62%**	**74%**	-	**76%**	1
#4	**54%**	**56%**	24%	-	2

prototypes.

— **Emotional prototypes are not universal.** As shown in Table 3.2, the prototypes are not universally preferred among participants. Although in Table 3.2(a) and 3.2(c), the top-ranking prototypes are much preferred over the others ("#2" of happiness and "#1" of anger), most preferences are far from 100% (and 0%). Especially in Table 3.2(b), most prototypes of sadness (including state-of-the-art prototypes) are about equally preferred among the hired participants. Indeed, most preferences are close to 50% which is quite far from 100%. Even there are cyclic preferences. Hence, there can be many prototypes of one emotion.

To sum up, our approach generated personalized prototypes differing from each other and from state-of-the-art prototypes. According to the ranking from 217 participants, some personalized prototypes are close to state-of-the-art prototypes (even "#1" and "Ek." of sadness are identical) and others are even much preferred to state-of-the-art prototypes. This suggests that the prototypes of one emotion are not unique among different people. They are supposed to be diverse.

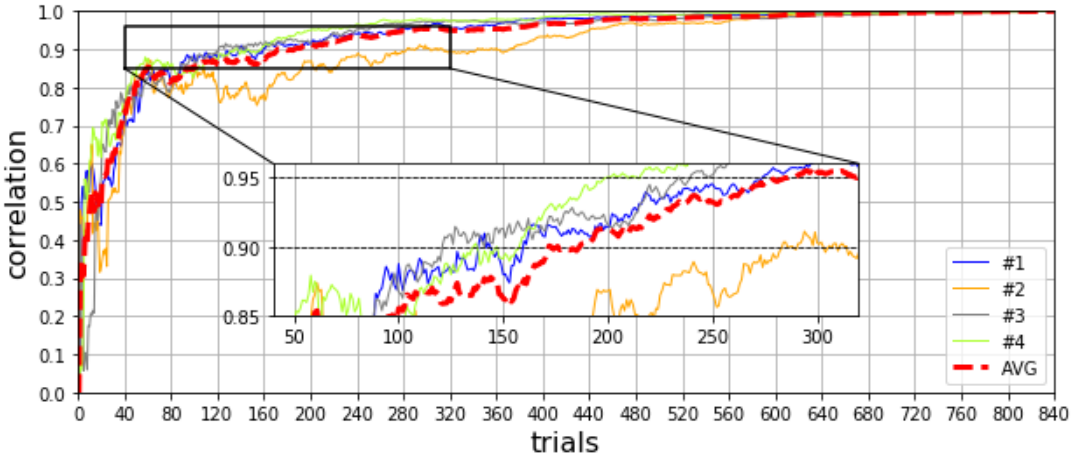
3.3.3 Discussion of convergence efficiency

We discuss here the convergence efficiency of our approach by monitoring the convergence of 1) dominant AU computation and 2) complementary AUs computation as we increase the number of trials used in the reverse correlation procedure.

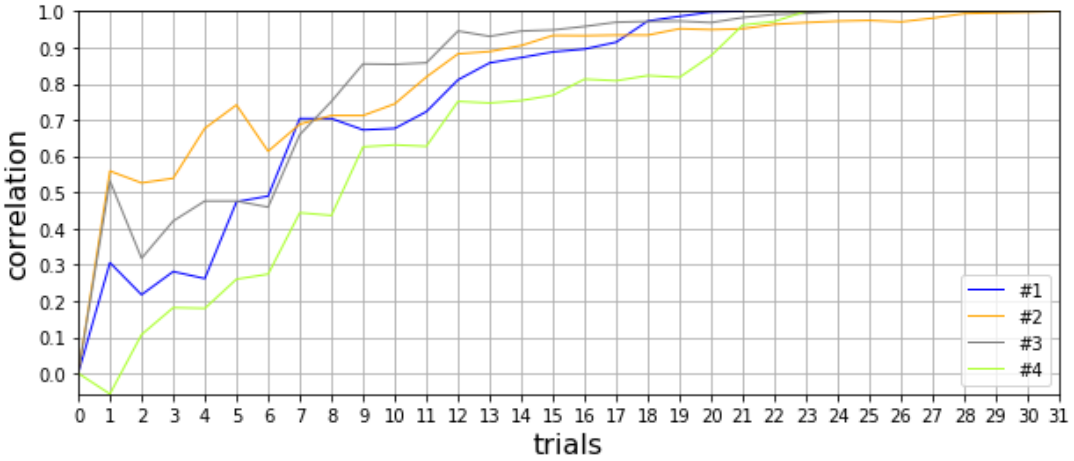
To do so, we compute 1) the correlation between the histogram for dominant AU computation (in Fig. 3.6) using the first n trials from Ω (i.e., the entire perceptual experiment), and the final histogram of dominant AU computation; 2) the correlation between the histogram of complementary AUs computation (in Fig. 3.6) using the first n trials from $\Omega_{\{d\}}$ (i.e., the subset of trials in which all the stimuli have the dominant AU activated), and the final histogram of complementary AUs computation.

Fig. 3.8 shows the convergence of the dominant AU computation and the complementary AUs computation from the perceptual experiment of confidence. Similar converging curves from the perceptual experiments of happiness, sadness, and anger can be found in Appendix Fig. 6. These curves reflect the typical reverse-correlation convergence (referring to Fig.6 of the related work [12]).

The dominant AU can be determined with only a 12-minute experiment. For the convergence of the dominant AU computation (in Fig. 3.8(a)), while all the 840 trials are considered, it takes less than 170 trials to reach a correlation of 0.9. As our



(a) Dominant AU computation



(b) Complementary AUs computation

Figure 3.8 – Example from the perceptual experiment of confidence to monitor the convergence of our approach. 3.8(a): Correlation between the result of dominant AU computation after the first n trials (x-axis) and the result after 840 trials. The average correlation for all observers is marked by the red dashed line. 3.8(b): Correlation between the result of complementary AUs computation after the first n trials (x-axis) in the corresponding subset and the result using all trials in the corresponding subset.

approach just takes the AU with maximum proportion to determine the dominant AU (see Eq. 3.2), performing 170 trials is enough. That is to say, only $170/840 \approx 20\%$ of the trials are necessary (equivalent to 12 minutes if the entire perceptual experiment needs 60 minutes).

Only a few data are used for complementary AUs computation. For each

observer, only a small subset of trials (i.e., $\Omega_{\{d\}}$) are, in effect, used to estimate complementary AUs. For instance, Fig. 3.8(b) illustrates that the largest subset, which is from observer #2, only includes 31 trials, and these trials are distributed throughout the entire perceptual experiment.

A prototype could have been determined in about 20 minutes. Although in our approach, the number of trials is set based on related works [63, 14, 93, 12], it appears unnecessary to randomly generate as many as 840 trials to determine dominant and complementary AUs for an observer. In fact, the duration of the perceptual experiment can be largely reduced. If our approach only randomly generates the first 170 trials to determine the dominant AU and then generates another 100 trials only from $\Omega_{\{d\}}$ to determine the complementary AUs, it will be less than 20 minutes (270 trials, instead of 840 trials) to obtain the mental prototype.

3.4 Conclusion

In conclusion, we proposed a novel interdisciplinary approach as the first pipeline (MDR) to personalize facial expressions by combining the reverse correlation from psychology with the facial expression manipulation technique from computer science. Our approach can personalize facial expressions (i.e., **Requirement #3** mentioned in Chapter 1) that are not limited to basic emotions (i.e., **Requirement #1** mentioned in Chapter 1) without the need for expertise (i.e., **Requirement #2** mentioned in Chapter 1). We choose the tool that controls the objective low-level attribute, i.e., action units, for the FEM task. We use this tool twice (at the beginning and at the end of our pipeline) to ensure that the manipulation is consistent with the mental prototype of the observer. Moreover, we introduce the concept of dominant and complementary action units to precisely describe facial expression prototypes.

The first limitation of our approach comes from the tool (GAN or any other type of low-level-attribute manipulation tool) for personalizing prototypes. Indeed, the choice of the tool can limit the number and the type of low-level attributes that could be manipulated. For instance, GANimation only focuses on AUs and does not consider other attributes, such as gaze direction [1]. Such attributes should also be integrated into the reverse correlation process. Moreover, among AUs, some AUs are not provided. For instance, GANimation is incapable of editing AU16 (lower lip depressor). Although it is not the goal of this paper, the authenticity of the face textures can be improved.

The common limitation of the related works and our approach is that all the stimuli are unimodal. Multimodal stimuli (e.g., video and audio) should be employed to enrich affective computing studies in the future.

Another limitation comes from the reverse correlation process. Performing 840 trials (about 40 to 60 minutes) for the perceptual experiment is time-consuming. As mentioned in Section 3.3.3, the generation of trials can be more efficient, and the duration can be greatly reduced. If user fatigue can be solved, potential applications can be imagined. For instance in the context of job searching like Randstad, once such a tool like ours extracts the prototypes of confidence of a candidate, this tool can be applied in the online interview by automatically transferring the candidate's face to the confident face.

We present the "milestone of this thesis" (in Fig. 3.9) as the mental prototypes generated by the first pipeline, compared with state-of-the-art prototypes. This figure will be updated when we employ the second pipeline. In the next chapter, we introduce our second pipeline Interactive Microbial Genetic Algorithm (IMGA) refining the first pipeline and fulfilling all the requirements mentioned in Chapter 1.

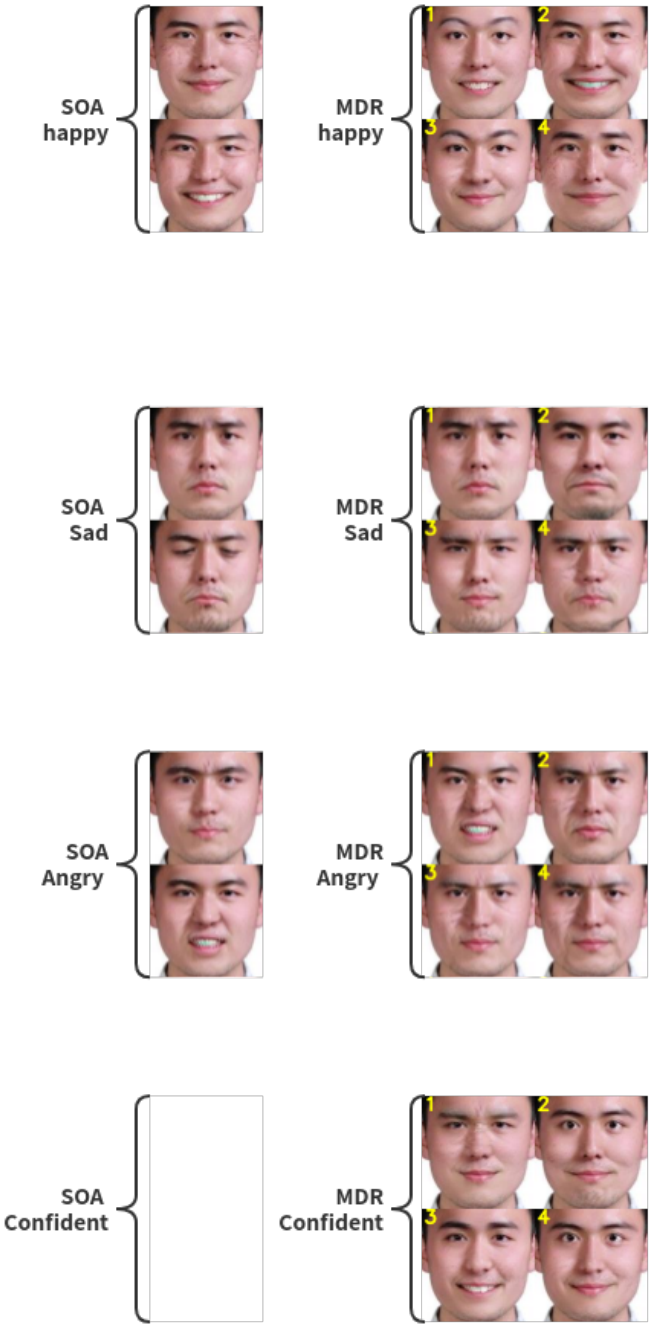


Figure 3.9 – Milestone of this thesis. **Left**: the state-of-the-art prototypes of Ekman [37] and Yu [126]. **Right**: the prototypes extracted by the first pipeline (MDR). *This figure will be updated in the next chapter.*

Chapter 4

The second pipeline: Interactive Microbial Genetic Algorithm (IMGA)

In this Chapter, we introduce our second pipeline: Interactive Microbial Genetic Algorithm (IMGA). The objective of the second pipeline is to refine the first pipeline (i.e., MDR system) thus fulfilling all the requirements mentioned in Section 1.1.3. In more detail, such an interdisciplinary approach integrates the psychological reverse correlation process (RevCor) into an interactive genetic algorithm (IGA). This approach not only inherits the strengths from the first pipeline but also solves the drawbacks of the first pipeline.

— The inherited strengths:

1. **Exhaustiveness, i.e., Requirement #1.** The categories of emotions are not limited to those provided in existing deep-learning databases. IMGA can extract the mental prototypes of a broader range of facial expressions (not only basic emotions but also non-basic emotions).
2. **Expertise-free, i.e., Requirement #2.** IMGA only requires the observer's perception (i.e., judgment on intuition) rather than the observer's expertise (e.g., no expert knowledge in affective computing, psychology, or certified FACS coders [37]).
3. **Flexibility, i.e., Requirement #3.** Everyone can use IMGA to extract his own mental prototypes of a given emotion thus meeting his needs. That is to say, facial expressions can be personalized.

— The solved drawbacks:

1. **Efficiency: by the online feedback loop, i.e., Requirement #4.** Unlike the traditional RevCor that generates massive trials randomly, in this approach,

based on the observer's feedback, automatically updated trials can contain more valuable information (closer to the mental prototypes of observers).

2. **Efficiency: by acceleration, i.e., Requirement #4.** The way of generating trials for mental prototype computations is intelligent. Moreover, we adopt the microbial genetic algorithm (MGA) [55] as the GA module within IGA to further accelerate the iteration since an elitist GA can converge faster than the non-elitist GA [67].

3. **Diversity, i.e., Requirement #5.** For one emotion, IMGA can provide multiple mental prototypes to each observer. This is closer to reality. In brief, one pipeline brings multiple solutions.

Moreover, differing from the traditional genetic algorithm, we added 2 blocks.

— A population evaluation module to objectively evaluate the quality of the entire population with limited trials.

— A three-state constraint automaton to limit the manipulation of facial expressions and determine the termination of the system.

The organization of this chapter is as follows. We introduce the methodology of IMGA in Section 4.1, then detail the experiments in Section 4.2, and next evaluate the results and compare them with the state of the arts in Section 4.3, and draw conclusions in Section 4.4. All the information about this pipeline can be found at https://yansen0508.github.io/Interactive-Microbial-Genetic-Algorithm/. The code is available at https://github.com/yansen0508/IMGA.

4.1 Methodology

Fig. 4.1 describes our IMGA pipeline. Similar to the traditional Genetic Algorithm (GA), it is an iterative process that repeats four steps: Population (initialization in Section 4.1.1 and update in Section 4.1.5), Selection (in Section 4.1.2), Crossover (in Section 4.1.3), and Mutation (in Section 4.5). For the glossary of our IMGA, the individuals are facial expressions that evolved by iteration. The population in each iteration is called a generation. The entire system, especially the interaction between the human (observer) and the machine (GA system), is detailed by a video demonstration [1].

1. https://yansen0508.github.io/Interactive-Microbial-Genetic-Algorithm/

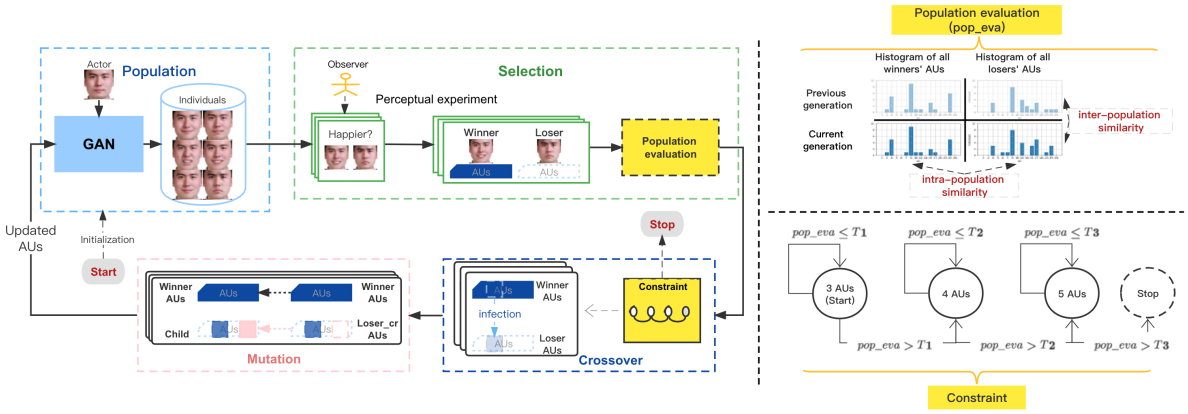


Figure 4.1 – Framework of our interactive microbial genetic algorithm (IMGA). An efficient interdisciplinary approach integrates the psychophysical reverse correlation process (RevCor) into an interactive genetic algorithm (IGA). For the genetic algorithm module within IGA, we adopt the microbial genetic algorithm (MGA) that can obtain various mental prototypes and accelerate the system's convergence. To monitor the convergence of the system and evaluate the quality of the entire population with limited trials, we add a population evaluation module. We also add a three-state constraint automaton to limit the manipulation of facial expressions and to determine the termination of the system. For the tool to generate different facial expressions, we employ GANimation [94] (denoted by "GAN") controlled by facial action units [37]. Please zoom in for better observation.

4.1.1 Population: initialization

We can employ any tool to generate facial expressions defined by low-level attributes. Here, we choose GANimation [94] controlled by facial action units (AUs) [37], i.e., the low-level attributes. This module can manipulate facial expressions with a relatively fine control. Thus, more facial expressions can be produced by different combinations of AU, regardless of whether these expressions belong to a certain type of emotion or not. In this procedure, GANimation (thereafter, *Gan*) takes as input a colored image of the actor's face s (e.g., captured with an emotionally neutral expression) and a n-dimensional binary vector v of AUs to create a deformed face (i.e., individual): $I = Gan(s, v)$.

GANimation is capable of manipulating $n = 16$ AUs from the list of AUs [2]. We define as v=$[\lambda_1, ..., \lambda_n]$, the binary AU vector where each component λ_i represents the activation ($\lambda_i = 1$) or deactivation ($\lambda_i = 0$) of $AU_{\mu[i]}$ (the i_{th} element in the AU list μ). For instance, $\lambda_3 = 1$ represents that AU4 (brow lowerer) is activated, $\lambda_9 = 0$ represents that AU12 (lip corner puller) is deactivated. See the literature [37] for a complete list of AUs. Appendix

2. $\mu = \{1, 2, 4, 5, 6, 7, 9, 10, 12, 14, 15, 17, 20, 23, 25, 26\}$

lists part of the AUs and their descriptions.

While GANimation can, in principle, simultaneously activate AUs, activating too many AUs typically create visual artifacts. Moreover, the state-of-the-art (SOTA) facial expression prototypes of Ekman et al. [37] indicate that most facial expressions have between 3 and 5 AUs activated. Therefore, we initialize the individuals by activating 3 AUs: $\forall v, \sum_{i=1}^{16} \lambda_i = c$, where $c = 3$. There can be $C_{16}^{3} = 560$ possible AU vectors $\mathcal{V} = \{v_1, v_2, ..., v_{560}\}$. Based on the actor's face s, we randomly choose N out of the 560 AU vectors to initialize a population of N individuals. Fig.4.1 ("Population" block) displays some examples of individuals from the initial population.

4.1.2 Selection

Selection has two parts: the perceptual experiment of RevCor and the population evaluation module.

Perceptual experiment.

In the perception experiment, we group the population of N individuals into $N/2$ pairs. Note that each pair of individuals is displayed only once for each iteration. In each trial of the perceptual experiment, a pair of individuals is displayed. Observers are asked to choose which individual best corresponds to the target expression (e.g., "which of these two faces looks happier?"). Note that each observer conducts $N/2$ trials for each generation. According to the answer from the observer, each pair of individuals are annotated by "winner" and "loser". Next, we use the set of $N/2$ winners (W) and the set of $N/2$ losers (L) to evaluate the quality of the current generation.

Population evaluation.

In the traditional GA, the computer can easily assign each individual a fitness value and rank all individuals from the best fit to the worst fit based on a mathematical fitness function. With the winner-loser strategy, $N(N-1)/2$ trials are required to rank the individuals. In our case, since the fitness function is the subjective judgments of observers, we cannot afford so many trials. That is why we only get N/2 trials, and we add a population evaluation module to evaluate the entire population.

For the population evaluation module (*pop_eva* of Fig. 4.1), we compute the similarity between the losers in the previous generation and the current generation, so-called inter-

population similarity (*inter_pop*), and the similarity between winners and losers in the current generation, so-called intra-population similarity (*intra_pop*).

$$pop\_eva = \alpha.inter\_pop + \beta.intra\_pop \tag{4.1}$$

$$inter\_pop = corr(H[L_g], H[L_{g-1}]) \tag{4.2}$$

$$intra\_pop = corr(H[W_g], H[L_g]) \tag{4.3}$$

α and β are positive constants. The similarity is computed by the Pearson correlation $corr(\cdot, \cdot)$. $H[\cdot]$ represents the histogram that counts how many times each AU (from the list μ) occurs in the corresponding set. W_g and L_g denote the AUs of winners W and losers L in the g_{th} generation, respectively. When the inter-population similarity increased, we can infer that there were fewer changes between the successive generations. When the intra-population similarity increased, we can infer that the losers became closer to the winners. Overall, we maximize *pop_eva* to ensure the convergence of the system.

4.1.3 Crossover

Three-state constraint automaton.

Here, we add the three-state constraint automaton to gradually increase the number of activated AUs from 3 to 5 during the crossover. Details are shown in Constraint of Fig. 4.1. The entire population goes through three restricted states: initially 3-AU activation state, then 4-AU activation state, and finally 5-AU activation state (denoted by "3 AUs (Start)", "4 AUs", and "5 AUs"), and three thresholds are given as $[T1, T2, T3]$ accordingly. The population is initialized by activating only 3 AUs for each individual, i.e., $c=3$ defined in Section 4.1.1. During the first state, i.e., 3-AU activation state, each individual can not have more than c AUs activated after the crossover. Once *pop_eva* exceeds the first threshold, i.e., $T1$, the system goes to the next state, i.e., 4-AU activation state. Accordingly, the threshold is updated to $T2$ and $c=4$ for the 4-AU activation state. As shown in Constraint of Fig. 4.1, the procedures in the following states are similar to that in the 3-AU activation state. Finally, the system stops when *pop_eva* exceeds the last threshold, i.e., $T3$.

Infection.

The infection operator is the same as the literature of MGA [55], i.e., uniform infection. The binary AU vector of the loser is infected by that of the winner. Thus, each element in the binary AU vector of the loser can be replaced by the corresponding element of the winner (illustrated in Fig. 4.1). The crossover rate is defined by cr. If the loser after the infection has more than c AUs activated, we randomly deactivate the excess. Note that the AU vectors of winners are unchanged, and the AU vectors of losers after infection are named $Loser\_cr$ in Fig. 4.1.

The infection operator is implemented as follows.

$$v_{loser\_cr} = v_{winner} * MASK_{cr} + v_{loser} * (1 - MASK_{cr}) \tag{4.4}$$

where $MASK_{cr} = [co_1, \cdots, co_n]$ with

$$\forall i \in [1, n], co_i = \begin{cases} 1 & rand(0,1) < cr \\ 0 & otherwise \end{cases}$$

where v_* represents the corresponding AU vector, the mask $MASK_{cr}$ is a vector of the same size as AU vector indicating where the loser will be infected, rand(0,1) generates a uniformly distributed random number between 0 and 1. The top of Fig. 4.2 illustrates the crossover operator.

4.1.4 Mutation

The mutation operator is the same as the literature on MGA [55], i.e., bit mutation. As shown in Fig. 4.1, each element of $Loser\_cr$ has the same mutation rate mr to change its binary value. Note that due to the infection and the mutation, individuals can have, or less than 3 AU activated, even though they are allowed to have more than 3 AU activated.

The mutation operator is implemented as follows.

$$v_{child} = -v_{loser\_cr} * MASK_{mr} + v_{loser\_cr} * (1 - MASK_{mr}) \tag{4.5}$$

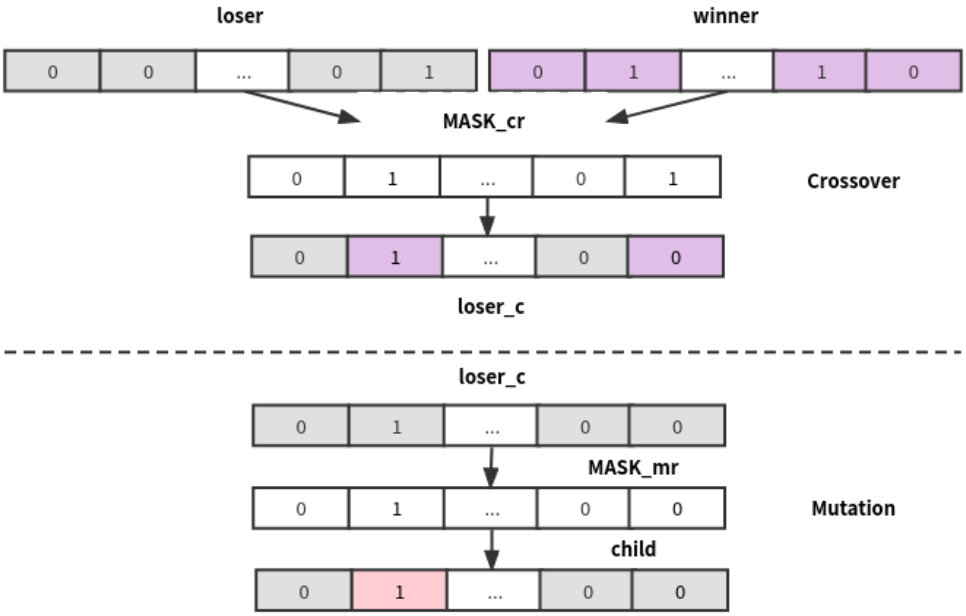


Figure 4.2 – Illustration of crossover operator (top) and mutation operator (bottom).

where $MASK_{mr} = [m_1, \cdots, m_n]$ with

$$\forall i \in [1, n], m_i = \begin{cases} 1 & rand(0,1) < mr \\ 0 & otherwise \end{cases}$$

where the mask $MASK_{mr}$ is a vector of the same size as AU vector indicating where *loser_cr* will be mutated, rand(0,1) represents a uniformly distributed random number between 0 and 1, mr is the probability of mutation. The bottom of Fig. 4.2 illustrates the mutation operator.

4.1.5 Population: update

In order to keep the size of the population unchanged, we only replace AU vectors of the losers with their offspring and keep the winners unchanged. In our approach, we employ the same tool, i.e., GANimation, to create the next generation by the updated low-level attributes, i.e., AU vectors.

The entire algorithm is demonstrated in Algorithm 1.

Algorithm 1 Interactive Microbial Genetic Algorithm

Require: actor's face s, action units vector v, crossover rate cr, mutation rate mr, population size N, maximum generation G

Ensure:

pop = init_pop(s, v, N)
stop = FALSE, g = 1
while (g < G) and stop is FALSE **do**
  for i = 1 to N/2 **do**
    $winner_i$, $loser_i$ = per_exp(pop,i)
  end for
  $W_g = [winner_1, \cdots, winner_{N/2}]$
  $L_g = [loser_1, \cdots, loser_{N/2}]$
  if $g > 1$ **then**
    stop, constraint = pop_eva($W_g, L_g, W_{g-1}, L_{g-1}$)
  end if
  if stop is FALSE **then**
    for i = 1 to N/2 **do**
      $loser\_c_i$ = crossover($winner_i, loser_i, cr, constraint$)
      $child$ = mutation($loser\_cr_i, mr$)
    end for
    $children = [child_1, \cdots, child_{N/2}]$
    pop = update_pop($pop, children$)
    g++
  end if
end while
return pop;

4.2 Experiments

The goal of our experiments is to validate our approach that can efficiently generate multiple prototypes corresponding to a given emotion, even a complex emotion. We first list the implementation details of our process in Section 4.2.1, then detail the experimental protocol in Section 4.2.2, and finally analyze the evolution of the population in Section 4.2.3.

4.2.1 Implementation details

GANimation model. We use the code of GANimation [94] released by its authors. All settings are unchanged.

GA parameter settings. We took the suggestion from the original literature on

MGA [55]: $cr = 0.5$ and $mr = 0.03$. The other GA parameters, i.e., the population size N, the constants α, β (defined in Eq. 4.1), and the thresholds for the three-state constraint automaton, were calibrated empirically. For this purpose, we simulated perceptual experiments by replacing the real observers shown in Fig. 4.1 with an automatic facial expression recognition system. All details and results of simulations are in the Appendix.

According to the simulations, we set the population size as $N = 20$, the constants of population evaluation $\alpha = 0.5, \beta = 0.5$, and thresholds of the three-state constraint automaton as $[T1{=}0.9, T2{=}0.9, T3{=}0.95]$.

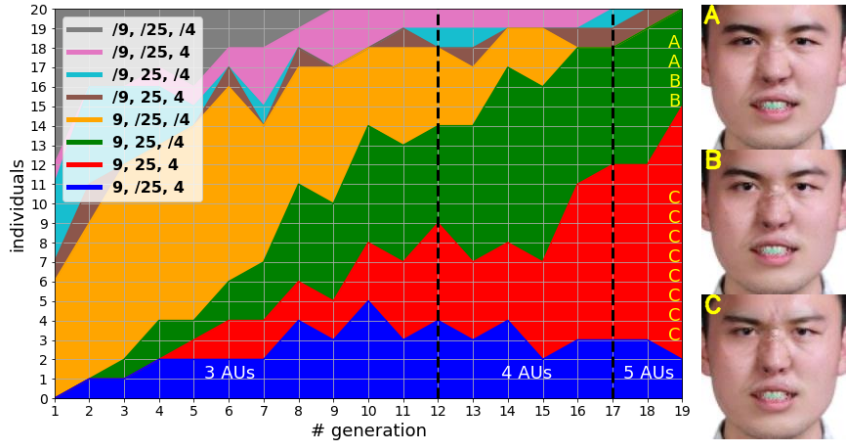
4.2.2 Experimental protocol

Observer demography. To validate our approach, we decided to recruit 12 observers since related works proposing new tools for perceptual experiments recruited a limited number of participants, from 8 to 12 [126, 12]. The 12 observers we recruited are adults (mean age: 34.7 yo) from five cultural groups: Algeria (1), China (1), Brazil (1), France (8), and Russia (1). Only two of the 12 observers have experience in affective computing, whereas nobody is a certified coder in Facial Action Coding System [37]. Each observer signed informed consent, and the experimental data were anonymous.

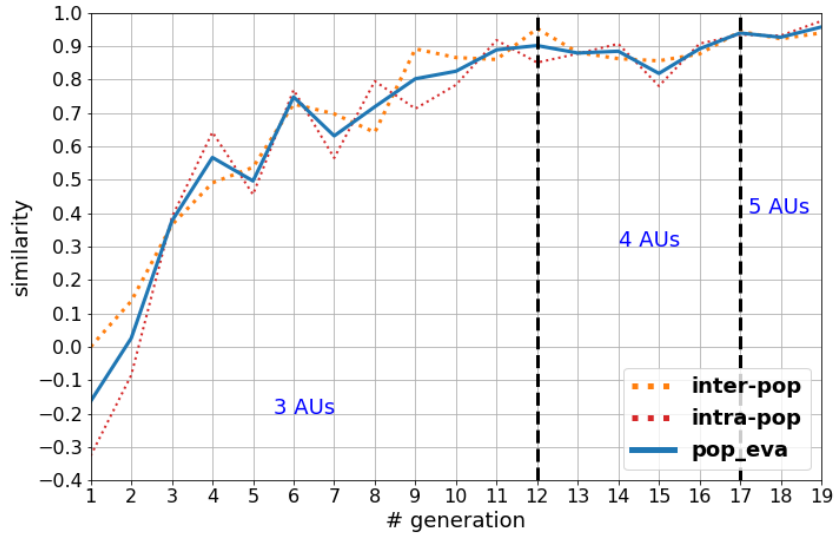
Perceptual experiment. To illustrate the efficiency of our approach, we chose three basic emotions (happiness, sadness, and anger) that existed in deep-learning databases and one complex emotion (confidence) that is not available in existing deep-learning databases. Each of the 12 observers participated in four different experimental tasks to find his/her mental prototypes of happiness, sadness, anger, and confidence. In each task, the question is fixed and unique. For example, "Which of these two faces looks happier?" The order of the four experimental tasks was counterbalanced among observers, and all experimental tasks used the same actor's photograph. Based on the three-state constraint automaton, the experimental task was automatically terminated. However, if the population has evolved for 50 generations, we forcibly stopped the current experimental task. Between the experimental tasks, observers had a 5-minute rest. All experiments were conducted in a quiet room in the laboratory, using a custom computer graphic interface from PsychoPy.

4.2.3 Results: evolution of the population

For all experimental results, we can observe the gradual evolution of the population. Fig.4.3 illustrates this evolution through the experimental results of anger from observer#4.



(a) Evolution of the population



(b) Population evaluation

Figure 4.3 – Experimental results of anger from observer #4. 4.3(a) left: evolution of the population over generations. The x-axis represents the generation number of the population. The y-axis represents the individuals of the current population. The legend lists 8 classes of individuals based on the activation or deactivation (marked by "/") of the related AUs (AU9, AU25, and AU4). For instance, "9, /25, 4" (blue) denotes all the individuals who had AU9 and AU4 activated and had AU25 deactivated. 4.3(a) right: three representative prototypes, i.e., the individuals with the same AU vectors in the last generation, where A: AU7, AU9, AU25; B: AU9, AU25; C: AU4, AU9, AU25. 4.3(b): Population evaluation. We draw the curves of the similarities computed by the population evaluation module. The vertical dotted lines in 4.3(a) and 4.3(b) indicate changing the constraint in the 12_{th} and the 17_{th} generation.

In Fig.4.3(a), we monitored the individuals according to the first, second, and third frequently occurring AUs and presented the representative prototypes. Fig.4.3(b) illustrates the corresponding values computed by the population evaluation module.

The AUs related to the observer's mental prototypes survived, while unrelated AUs gradually disappeared. Based on the subjective judgment from observer #4 on "which of these two faces looks angrier?" AU9 was the most frequently occurring AU in the last generation. AU25 and AU4 were the second and the third. They are more relevant than the other AUs to anger for this observer. If we look at the evolution from the first to the last generation, we observe that the related AUs (AU9, AU25, and AU4) spread throughout the population, and the individuals without the related AUs activated were gradually eliminated (the gray area of Fig. 4.3(a)). Therefore, more individuals with the related AUs activated appeared in subsequent trials.

The related AUs combined with each other. As the experiment proceeded, we noticed the growths of the green, red and dark blue areas and the disappearance of the other areas. This indicates that observer #4 prefers to combine AU9 with AU4 and/or AU25 than the other AUs. During the 3-AU activation state, the system gradually converged. Indeed, 18 of 20 individuals had AU9 activated in the 12_{th} generation. At the end of the experiment, 100% of the population had AU9 activated (green, red, and dark blue areas), 90% of the population had the combination of AU9 and AU25 (green and red areas), 75% of the population had the combination of AU9 and AU4 (red and dark blue areas), and 65% of the population had activated both AU4, AU9, and AU25 (red area).

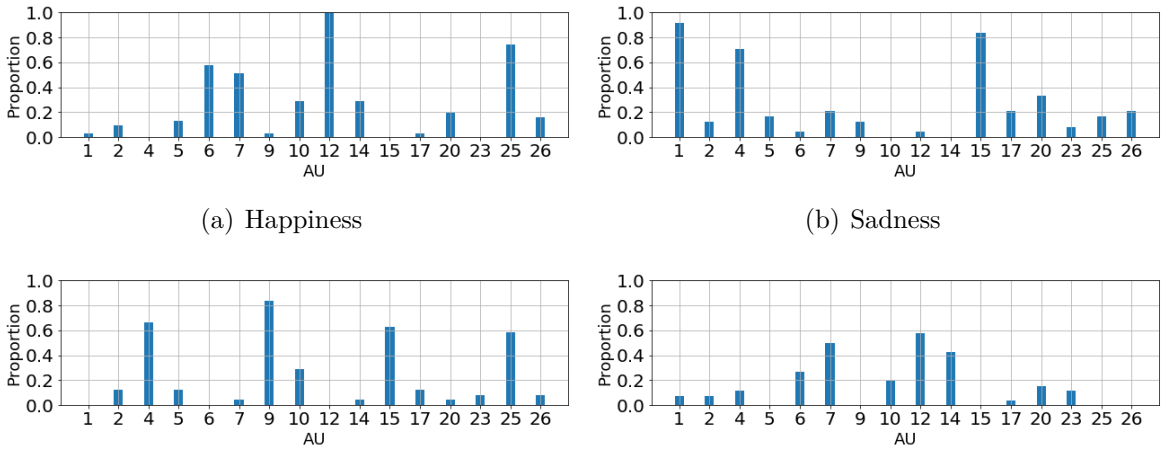
Fig. 4.3(b) reflects the convergence of the system by illustrating the similarities computed by the population evaluation module. Our approach considers not only the inter-population similarity but also the intra-population similarity. During the 3-AU activation state, the system gradually converged. Once changing the constraint, the system searched for results in a broader space and then converged again. That is why the curves dropped and re-converged during the 4-AU activation state and the 5-AU activation state. See the video demonstration for extra information: 1) the AU histograms of winners' and losers' AUs in the previous and the current generations, 2) winners and losers of the current generation, and 3) the computer graphic interface for the experiments.

We define the **representative prototypes** as the individuals with the same AU vectors in the last generation. In Fig. 4.3(a), there are three different representative prototypes of anger from observer #4 denoted by "A" (2 individuals), "B" (2 individuals), and "C" (8 individuals). Next, we quantitatively and subjectively evaluate the representative

prototypes.

4.3 Evaluations and Comparison

For each emotion category, we collected all representative prototypes of observers. All representative prototypes are listed in the Appendix and in Fig. 4.7. In this section, first, we analyze representative prototypes from two perspectives: action units (in Section 4.3.1) and prototypes (in Section 4.3.2). Second, we present our subjective evaluation process (in Section 4.3.3 and Section 4.3.4) for two purposes: 1) to validate that our representative prototypes can reflect observers' mental prototypes and 2) to subjectively compare with the SOTA prototypes. Third, we discuss the efficiency by comparing our approach with the related works using RevCor for affective computing (in Section 4.3.5).



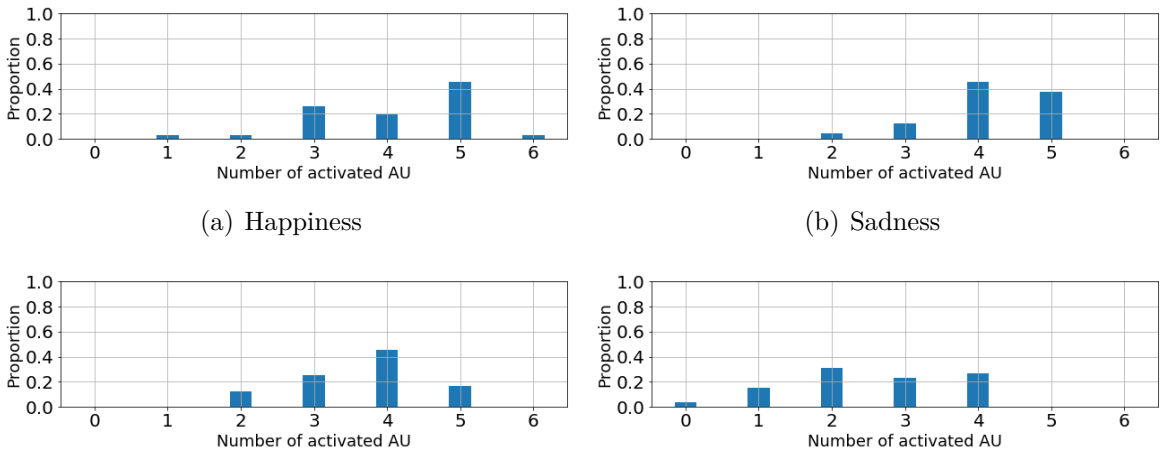
(a) Happiness (b) Sadness

(c) Anger (d) Confidence

Figure 4.4 – Proportion of each AU in the representative prototypes. For the basic emotions, some AUs reveal universality.

4.3.1 Quantitative evaluation: Action units

Our IMGA-generated prototypes are compatible with state-of-the-art prototypes. Our findings indicate all representative prototypes generally convey similar emotional expressions. Fig. 4.4 presents the proportion of each AU that appears in our representative prototypes. We find all AUs from SOTA prototypes [37, 126] in our repre-



(a) Happiness (b) Sadness

(c) Anger (d) Confidence

Figure 4.5 – The proportion of prototypes that have different numbers of AUs activated. There is a discrepancy between basic emotions and confidence.

sentative prototypes, except AU16 (lower lip depressor) for anger [3]. Note that these AUs are not included in the list of editable AUs μ by GANimation aforementioned in Section 4.1.1. To consult the AUs of SOTA prototypes, please see the literature [37, 126].

Within the scope of basic emotions, some AUs reveal universality. Some common AUs can be found in the representative prototypes of basic emotions. In Fig. 4.4, for happiness, 100% of representative prototypes have AU12 (lip corner raiser) activated. For sadness, more than 80% of representative prototypes have AU1 (inner brow raiser) and AU15 (lip corner depressor) activated. For anger, 83% of representative prototypes have AU9 (nose wrinkler) activated. For confidence, the proportions of AUs are not as prominent as those in the basic emotions: 57.7% of representation prototypes have AU12 (lip corner raiser) activated, and 50% of representative prototypes have AU7 (lid tightener) activated.

There is a discrepancy between basic emotions and confidence in terms of the number of activated AUs. Fig. 4.5 summarizes the proportion of prototypes that have different numbers of AUs activated. Typically, most representative prototypes of basic emotions have at least 3 AUs activated. For confidence, although our system initialized with the constraint of 3-AU activation, approximately 50% of the representative prototypes have less than 3 AUs activated.

3. In this thesis, we replaced AU16 with AU25 (lips part) in order to make the reconstructed prototype of anger as close as possible to that of Yu et al. [126].

4.3.2 Quantitative evaluation: Prototypes

The diversity of mental prototypes not only exists within observers, but also between observers. Table 4.1 indicates a great variety of prototypes. From 12 observers, we obtained 31, 24, 24, and 26 different representative prototypes of happiness, sadness, anger, and confidence. See all the representative prototypes in the Appendix or in Fig. 4.7. On average, multiple representative prototypes are acquired per observer. This indicates the diversity of mental prototypes within observers. Furthermore, only a small proportion of these different representative prototypes coexist in at least two observers. Most representative prototypes are different between observers. It can also indicate the diversity of the mental prototypes between observers. Given that neither the SOTA prototypes of Ekman et al. [37] (except sadness) nor Yu et al. [126] could be found in our representative prototypes, this implies that the SOTA prototypes need to be refined. The number of prototypes should be enlarged.

Table 4.1 – Number of different representative prototypes in all observers. Among these different representative prototypes, we also list the proportion of coexisting prototypes between different observers.

	Happiness	Sadness	Anger	Confidence
Number of proto	31	24	24	26
Proportion	22.6%	25%	16.7%	15.4%

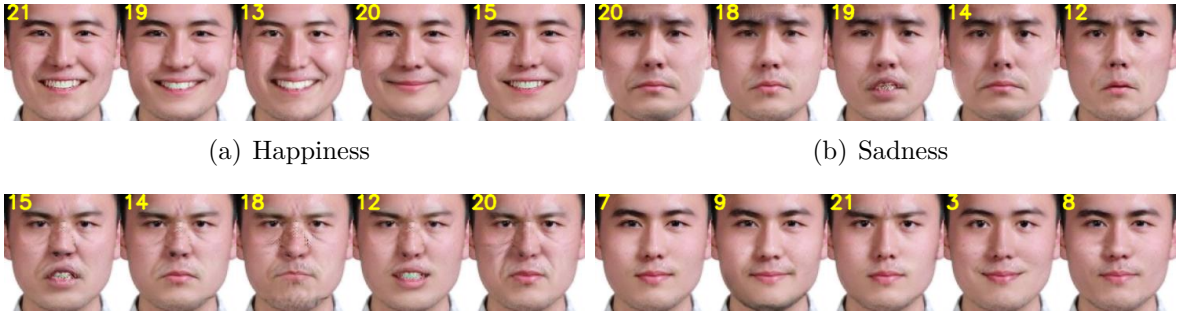
4.3.3 Subjective evaluation: Protocol and measurements

Here, we present our subjective evaluation. We asked the same 12 observers to participate in the subjective evaluation. The subjective evaluation process was divided into four tasks corresponding to the four facial expressions: happiness, sadness, anger, and confidence. In each task, we created an evaluation set including all representative prototypes of observers and the SOTA prototypes [37, 126] (except confidence, for which no SOTA prototype is available). All prototypes in each evaluation set were presented in shuffled order. In the tasks of happiness, sadness, anger, and confidence, observers were asked to select five faces that were the happiest / saddest / angriest / most confident, respectively.

We applied two measurements for the subjective evaluation. First, we count the proportion of observers who still choose at least one representative prototype of theirs. Second, in more detail, we ranked the representative prototypes from the most selected prototype to the least selected prototype by the 12 observers.

Table 4.2 – Proportion of observers who still choose at least one of their representative prototypes. We compute the corresponding baseline by random selections.

	Happiness	Sadness	Anger	Confidence
Ours	83.3%	75%	66.7%	66.7%
Baseline	41.6%	44.3%	39.7%	42.2%



(a) Happiness (b) Sadness

(c) Anger (d) Confidence

Figure 4.6 – We display the top-5 most selected prototypes by the 12 observers. The names of the prototypes are marked in yellow. There is no state-of-the-art prototype appearing in the top-5 prototypes. See the complete ranking in the Appendix.

Table 4.3 – Experiment time (in minutes) of our approach.

	Happiness	Sadness	Anger	Confidence
mean	11.1	12.4	8.6	11.3
std	3.8	3.6	2.9	3.3

4.3.4 Subjective evaluation: Results

Most observers still selected at least one of their mental prototypes. Table 4.2 lists the proportion of observers who still choose at least one representation prototype of theirs. The baseline is derived from random selections of each evaluation task. See the calculation of the baseline in the Appendix. Compared with the baselines, a larger proportion of observers still selected at least one of their representative prototypes. This can indicate that our representative prototypes can reflect the observers' mental prototypes.

Observers less preferred state-of-the-art prototypes. We first sorted all the prototypes according to the proportion selected by observers. Then, we displayed the top-5 prototypes with the highest proportions in Fig. 4.6. We noticed that there is no SOTA prototype appearing in the top-5 prototypes of the three basic emotions. See the complete ranking in the Appendix.

4.3.5 Comparison with related works using RevCor: Efficiency

First, we present the experiment time of our approach. Then, in order to discuss the efficiency, we compare the experiment time and the number of mental prototypes for each observer between our IMGA and the related works using RevCor for affective computing.

Converging speed of our IMGA. In Table 4.3, we present the duration for observers to perform the perceptual experiments. On average, observers performed the perceptual experiments on anger faster than the experiments on the other facial expressions. By calculating the average time for all perceptual experiments, it takes about 10.8 minutes (with 330 trials) for an observer to obtain mental prototypes using our IMGA. The Appendix provides more details about the experiment time and the number of trials.

Table 4.4 – Comparison between our IMGA and the related works using reverse correlation process for affective computing [126, 63, 14, 93, 12]. We list in the first column: the stimuli category, the reverse correlation paradigm, the number of affective states, the number of trials performed by one observer for all affective states, the number of trials performed by one observer for one affective state and the number of mental prototypes for one observer.

	stimuli	paradigm	states	trials/obs	trials/obs/state	proto/obs
IMGA	face	2-AFC	4		**330**	**multiple**
Yu[126]	face	7-AFC	6	2400	(400)	single
Jack[63]	face	7-AFC	6	4800	(800)	single
Chen[14]	face	3-AFC	2	3600	(1800)	single
Ponsot[93]	speech	2-AFC	2		∼700	single
Burred[12]	speech	2-AFC	1		700	single

Comparison with related works. Since all the related works did not provide the experiment time, we compared the number of trials for the perceptual experiment. Table 4.4 illustrates the results of the related work using RevCor, three for facial expression and two for speech. Due to the different paradigms, the numbers of trials are presented in two different ways, i.e., "trials/obs" and "trials/obs/state". By comparing our IMGA ("330") with the works using the two-alternative forced choice (2-AFC) paradigm, ("∼700", and "700" for [93], and [12]), our work reduced the number of required trials (per observer, for one affective state) by approximately a factor of two.

We cannot directly compare with the works [126, 63, 14] that employed different paradigms, since "trials/obs/state" is unknown in the original literature. By calculating "trials/obs"/"number of affective states" to compare these works ("(400)", "(800)", and "(1800)" for [126, 63], and [14]) with ours ("330"), our approach still needs fewer trials than

these works.

In summary, compared with related works using RevCor, our approach has two strengths. First, our approach shrinks the experiment time. Second, only our approach can obtain multiple mental prototypes for each observer.

4.4 Conclusions

In this chapter, we proposed an efficient interdisciplinary approach: Interactive Microbial Genetic Algorithm (IMGA) to **personalize** facial expressions. Such an interdisciplinary approach that integrated the psychological reverse correlation process (RevCor) into an interactive genetic algorithm (IGA) **efficiently** explored **diverse** mental prototypes in a broader range of facial expressions (basic emotion and non-basic emotion, i.e., **exhaustiveness**) for each observer. Our IMGA considered **real-time feedback** from observers to update subsequent trials and further **accelerated** the process by using an elitist GA (i.e., MGA). Similar to the first pipeline, the entire process is **expertise-free**. All you need is subjective judgments. Differing from the traditional genetic algorithm, we added a population evaluation module to evaluate the quality of the entire population with limited trials and a three-state constraint automaton to limit the manipulation of facial expressions and determine the termination of the system.

Compared with the SOTA prototypes [37, 126], we observe that the diversity of mental prototypes exists not only within observers but also between observers. Thus, the prototypes of a given emotion should be enlarged. Furthermore, our approach can extract the emotions that are not available in existing deep-learning databases. Compared with the related works using RevCor [126, 63, 14, 93, 12], our approach is more efficient at two-fold: faster and obtaining multiple mental prototypes. However, the limitation of the related works and our IMGA is that all the stimuli are unimodal. Multimodal stimuli (e.g., video and audio) should be employed to enrich affective computing studies in the future. Another limitation comes from the facial-expression manipulation tool we chose. Indeed, GANimation [94] provides only 16 editable AUs. In the future, GANimation can be replaced by the other FEM (facial expression manipulation) tool that can edit more AUs.

Figure 4.7 – Milestone of this thesis. **Left**: the state-of-the-art prototypes of Ekman [37] and Yu [126]. **Middle**: the prototypes extracted by the first pipeline (MDR). **Right**: the representative prototypes extracted by the second pipeline (IMGA) (see the Appendix for better observation).

Chapter 5

Conclusion and perspective

5.1 General conclusion

Facial expressions (FE) can have a significant impact on social interactions. Reading facial expressions is a way to perceive or interpret the internal (mental) state of humans and can also influence how people interact with each other [111]. Since the publication of Darwin's book *The expression of the emotions in man and animals*, facial expressions are always a hot topic in psychology or, more generally, in cognitive science. In computer science, as we collected from Google scholar in Chapter 2, recently from 2018 to 2022, FE-based research on facial expression recognition, i.e., FER, and facial expression manipulation, i.e., FEM, has attracted increasing attention. In this thesis, according to the application context of Randstad, we focus on one of the downstream tasks: FEM.

As we introduced in Chapter 1, FEM-based techniques and applications have always faced three major challenges. In brief, **Diversity**: for one emotion, there should be multiple facial expression prototypes. **Flexibility**: FEM applications should be personalized to specific users (in this thesis, we called observers). That is to say, the generated facial expressions can meet the need of the users. **Exhaustiveness**: most works only focus on the basic emotions of Ekman, i.e., happiness, sadness, anger, surprise, disgust, fear, and (added later) contempt. Non-basic emotions are not available. As we detailed in Chapter 2, the drawbacks of the database induced these challenges. To address these challenges and fulfill the application requirements (see Section 1.1.3) in brief: **Exhaustiveness, Expertise-free, Flexibility, Efficiency, Diversity**, we propose another way of thinking inspired by the mechanism of how psychologists analyze facial expressions.

We propose our interdisciplinary approach that combines the technique from psychology, i.e., reverse correlation process (RevCor), with the technique from computer science, i.e., FEM and interactive genetic algorithm (IGA). We present our approach in two pipelines: the mental deep reverse-engineering system (MDR) and the interactive microbial genetic algorithm (IMGA). The contributions of the first pipeline (i.e., MDR) are highlighted as

follows.

- **Exhaustiveness.** MDR allows subjective judgments on any emotion, including those not available in existing deep-learning databases. *This meets Requirement #1.*
- **Expertise-free.** Expert knowledge is not required. The only requirement is the observer's perception (i.e., judgment on intuition). *This meets Requirement #2.*
- **Flexibility.** MDR can flexibly personalize facial expressions to fit the expectations of any observers. *This meets Requirement #3.*
- We manipulate facial expressions **on real faces** rather than avatars. In order to be **consistent**, we employ the same FEM module twice: 1) to generate stimuli for RevCor and 2) to generate personalized facial expressions controlled by the mental prototypes obtained from RevCor.
- To enhance the definition of facial expression prototypes, we introduce the concept of **dominant and complementary action units** to precisely describe facial expression prototypes.

The purpose of the second pipeline (i.e., IMGA) is to refine the first pipeline (MDR), mainly in two aspects: efficiency and diversity. Therefore, the contributions of the second pipeline are summarized as follows.

- **Exhaustiveness (an inherited strength from MDR).** The categories of emotions are not limited to those provided in existing deep-learning databases. IMGA can extract the mental prototypes of a broader range of facial expressions (not only basic emotions but also non-basic emotions). *This meets Requirement #1.*
- **Expertise-free (an inherited strength from MDR).** IMGA only requires the observer's perception (i.e., judgment on intuition) rather than the observer's expertise (e.g., no expert knowledge in affective computing, psychology, or certified FACS coders [37]). *This meets Requirement #2.*
- **Flexibility (an inherited strength from MDR).** Everyone can use IMGA to extract his own mental prototypes of a given emotion thus meeting his needs. That is to say, facial expressions can be personalized. *This meets Requirement #3.*
- **Efficiency (a solved drawback of MDR).**
  - **By the online feedback loop.** Unlike the traditional RevCor that generates massive trials randomly, in this approach, based on the observer's feedback, automatically updated trials can contain more valuable information (closer to the mental prototypes of observers).

— **By acceleration.** The way of generating trials for mental prototype computations is intelligent. Moreover, we adopt the microbial genetic algorithm (MGA) [55] as the GA module within IGA to further accelerate the iteration since an elitist GA can converge faster than the non-elitist GA [67].

These two contributions meet Requirement #4.

— **Diversity (a solved drawback of MDR).** For one emotion, IMGA can provide multiple mental prototypes to each observer. This is closer to reality. In brief, one pipeline brings multiple solutions. *This meets Requirement #5.*

Generally, our approach is plug-and-play not only for Randstad but also for other contexts, e.g., a tool for psychological studies, and a tool for psychotherapy application. The components of our approach are replaceable. The tool to generate facial expressions can be replaced or updated by other tools (if they can manipulate the objective low-level attributes), and even the tool can be replaced by the tool for speech manipulation (the reverse-correlation features should be correspondingly replaced such as [93, 51]). The optimization algorithm can be replaced. For instance, our IMGA can be replaced by a simple statistical approach proposed in Section 3.3.3 (performing x trials to determine the dominant AU and then performing another y trials only from $\Omega_{\{d\}}$ to determine the complementary AUs).

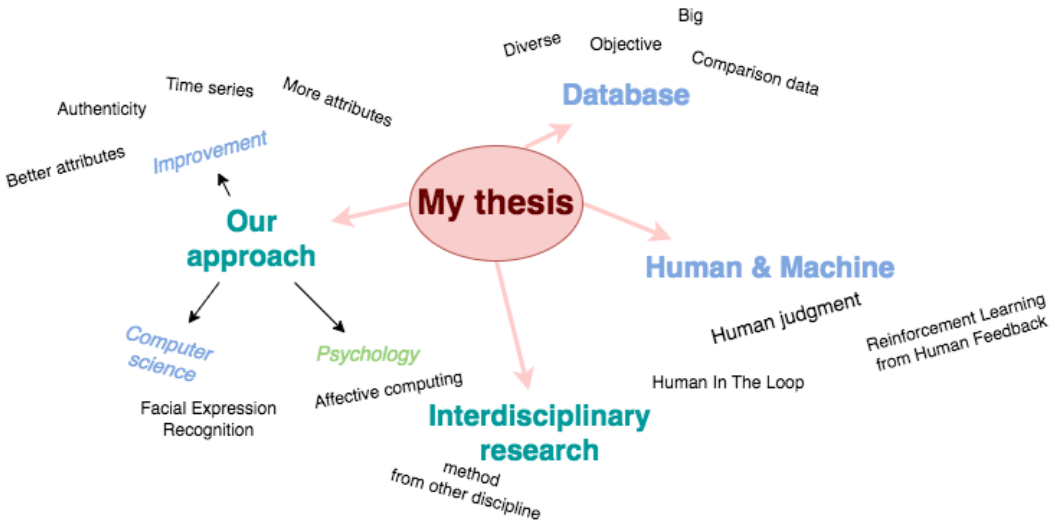


Figure 5.1 – Perspectives of this thesis.

5.2 Perspectives

The organization of the perspectives is as follows (shown in Fig. 5.1). We first present the perspectives of our approach in 3 aspects: the improvement of our approach, the application in psychology, and the application in computer science. Then we present the perspectives of this thesis in more general but different aspects: in terms of building a database, in terms of human and machine, and finally rethinking the interdisciplinary approach.

5.2.1 In terms of our approach

Here, we focus on the perspectives only in terms of our approach. We list the perspectives in terms of the improvement of our approach, in terms of the potential study in psychology, and in terms of the possible solution for the FER task.

Improvement of our approach

In this thesis, we focus on facial expressions. **Based on facial expressions**, there are four points that can be improved.

1. **Action units.** According to the 30 main-coded AUs shown in the Appendix, our FEM tool, i.e., GANimation [94], is incapable to edit nearly half of them, such as AU16 (see Section 4.3.1).
2. **Authenticity.** The authenticity of some AU combinations could be improved. For instance, texture distortions in the teeth region appear when we activate the AUs associated with the mouth (see Appendix Fig. 3).
3. **Other FE low-level attributes.** In addition to AUs, other low-level attributes, such as gaze direction [1], could also be considered.
4. **Time Series.** Currently, our approach only focused on static photographs. For future work, the time series will be considered. That is to say, for the FEM task, we should generate video stimuli rather than images.

In terms of the mental prototypes of emotion, our approach is based on facial expressions, i.e., a visual modality, to describe emotion. However, the concept behind our approach can be employed to extract the mental prototypes of emotion in **other modalities**, e.g., audio, and neural signals.

Affective computing study in psychology

By using our approach, more psychology studies about affective computing can be envisaged. For instance, it would be interesting to use our approach to study the time course of AU activation over a few seconds, and especially if the AU combinations differ in their latency and dynamics [62, 5]. It would also be interesting to use our approach to investigate whether expressions generated by observers of one culture are rated as more prototypical by participants of the same culture [61]. More generally, we envision that systems like ours can be used as a tool to provide experimental control over observers' emotional expression in dyadic interactions to study, e.g., whether one participant's dominant or confident attitude influences the outcome of group behavior [21], or whether emotional convergence improves the quality of the interaction [105].

Application of facial expression recognition

Although the FER task is not the major research objective of this thesis, mental prototypes can be employed for the FER task. For instance, we can create a mental-prototype database if we have collected the mental prototypes from as many and as diverse (on demographics) observers as possible. By comparing the similarity between the source mental prototype (e.g., a list of AUs that are activated) and all possible prototypes of this emotion in our database. Indeed, our mental-prototype-based database is not limited to basic emotions.

5.2.2 In terms of building databases

Considering the state of the art of databases in Chapter 2, here are our suggestions in terms of creating databases, not limited to building FE-based databases but more generally emotional databases.

The more diverse the metadata is, the better the database will be. Many databases usually contain detailed filming information such as camera position, and illumination [50]. Since emotional prototypes are not unique (or so-called universal) but diverse [97, 63, 5, 61, 62], to build an emotional database, detailed demographics such as age, gender, and ethnicity should also be included. Moreover, it is also recommended to add low-level attributes compared with just annotating emotional labels.

Objective low-level attributes are recommended. As we discussed in Chapter 2, the perception of emotions is relatively subjective, thus even experts may have different

judgments. Adding low-level attributes can provide data with an additional description. Therefore, the more objective the low-level attribute is, the more convincing the database will be.

Comparison data are increasingly used. The perceptual experiment in our approach is closer to creating a dataset of comparisons in terms of emotion. Recently, comparison data are increasingly used. For instance, recently, this annotating manner (annotation by human feedback from comparison data) applied in natural language processing (NLP) has achieved great success, such as ChatGPT [1] (see InstructGPT [87]), and the work [3] from Anthropic [2]. The objective of using comparison data is that this annotating manner is more reliable compared to the single annotation. Especially, the data for annotation are relatively subjective (i.e., requiring human judgments without fixed judging criteria/mechanism) such as annotating non-basic emotion (like our work) in computer vision (CV), and annotating harmful/helpful dialogue [3] in NLP.

Furthermore, the comparison data does not need to be unimodal. The **multimodal data pairs** can also be considered. An influential example is CLIP [95, 44] which learned transferable visual models from the supervision of NLP signals. This work is pre-trained by 400 million image-text paired data (i.e., multimodal pairs: images and their captions). The SOTA image representations can be obtained by "predicting which caption goes with which image" and achieves zero-shot transfer. Although comparison data can be a good idea to create a database, the database still has to be very large in order of magnitude. As aforementioned examples [95, 87, 3], at least hundreds of millions of data are required.

5.2.3 In terms of human and machine

We should take full advantage of human judgments. Indeed, perceiving emotions are relatively subjective, and the perception can be different in different conditions (e.g., across cultures). It is a very large area of study in psychology [62, 5, 61, 21, 105]. More generally, human judgments or intuitions are easy to elicit for humans but complex for machines to formalize and automate. Indeed in some of the CV downstream tasks such as FEM and FER and in some of the NLP downstream tasks such as QA query and sentiment analysis, due to the knowledge learned by machine learning cannot win human knowledge, humans have better performance than machines. In order to minimize the gap between machines and humans, incorporating human involvement into the system can be

1. chat.openai.com
2. AI research and safety company. https://www.anthropic.com/

beneficial. **Indeed, our second pipeline employed this mindset (incorporating human involvement to control a pre-trained GAN).** That is also why there are increasing works using Human-in-the-loop (HITL) [119] or Reinforcement Learning from Human Feedback (RLHF) [18].

5.2.4 In terms of interdisciplinary research

Just like the invention of the flying machine is inspired by the wings of birds from zoology, the solution does not have to be entirely limited to computer science. Indeed, in terms of research on facial expressions, emotions, or affective states, **incorporating the technique from other disciplines**, such as RevCor from psychology, **can be helpful**. Especially, in other domains such as medical and healthcare research or for many companies such as small and medium-sized enterprises rather than the leading companies like Microsoft, Google, Amazon, OpenAI, and Meta, data is hard to collect and the amount of the data is very limited. It would be inspiring to use an interdisciplinary approach to address the challenge, thus avoiding collecting billions of data.

Overall, we hope this thesis can pave the way for further scientific studies not only in psychology but also in computer science. We also expect our approach can be personalized to users in different application domains, e.g., as mentioned in Chapter 1, in the context of human resources, a digital coach for the online interview; and in the context of psychotherapy, a digital mirror treating psychiatric disorders of emotion.

GLOSSARY

General terms

Actor: the subject whose face will be manipulated.

FE: Facial Expressions.

FACS: Facial Action Coding System.

AU: Action unit, encoded facial muscle movements by Facial Action Coding System.

RevCor: psychophysical reverse correlation process.

FER: Facial Expression Recognition.

FEM: Facial Expression Manipulation.

Observer: user or participant involved in the perceptual experiment.

Personalization: the results (i.e., in this thesis, facial expressions) that can meet the need of users (observers).

Stimuli: a psychological term, in this thesis stimuli represent the generated facial expressions that are used in the first step of our proposed pipelines (i.e., MDR and IMGA).

Mental representation/prototype: the representation extracted from the reverse correlation process that can reflect the mental image of a given emotion [3].

MDR: Mental Deep Reverse-Engineering System, abbreviation of our firs pipeline.

IMGA: Interactive Microbial Genetic Algorithm, abbreviation of our second pipeline.

Metadata: the recording information of the database.

Demographics: including the range of ages, gender ratio, and ethnicity.

Emotional FACS rules: interpreting and categorizing the facial expression as an emotion based on these rules, also called prototypes of Ekman.

Abbreviation of databases

JAFFE: Japanese Female Facial Expression

CK+: Extended Cohn-Kanade database

BU-3DFE: Binghamton University 3D Facial Expression

3. in this thesis, we focus on emotion.

Multi-PIE: the CMU Multi-PIE Face Database

RaFD: Radboud Faces Database

Oulu-CASIA: Oulu-CASIA NIR&VIS facial expression database

DISFA: Denver Intensity of Spontaneous Facial Action database

FER2013: database for ICML 2013 Facial Expression Recognition Challenge

AFEW: Acted Facial Expressions in the Wild database

EmotioNet: database for EmotioNet challenge

RAF-DB: Real-world Affective Face Database

Terms in the first pipeline (MDR)

GANimation: a Generetic Adversarial Network used to generate facial expressions by controlling a list of action units.

S: the actor's face, the input of the first pipeline.

G: GANimation.

I: the deformed (generated) face created by GANimation, i.e., stimulus.

v: a 16-dimension binary vector of available AUs.

μ: a list containing the corresponding AU numbers of v.

λ: the binary value representing the activation or deactivation of the AU.

V: vectors of all the possible AU combinations by activating only 3 AUs.

Φ: the set of all the possible stimuli generated by GANimation.

Φ_i: the set of all the stimuli where AU_i is activated.

$\Phi_{\bar{i}}$ the set of all stimuli in which AU_i is deactivated.

s_I: representing selected ($s_I = 1$) or non-selected ($s_I = 0$) of a given stimulus I.

m: the number of trials in a perceptual experiment.

Ω: all the trials of the perceptual experiment.

$\Omega_{\{i^*\}}$: the subset of trials in which one of the paired stimuli has AU_i activated and another one has AU_i deactivated.

$\Omega_{\{i\}}$: the subset of trials in which both stimuli have AU_i activated.

$\Omega_{\{\bar{i}\}}$: the subset of trials in which both stimuli have AU_i deactivated.

Z_Ω: the set of all selected stimuli in the perceptual experiment.

$P(i|\Omega_{\{i^*\}})$: the proportion of the selected stimuli that have AU_i activated in the subset $\Omega_{\{i^*\}}$

$P(j|\Omega_{\{d,j^*\}})$ the proportion of selected stimuli in subset $\Omega_{\{d,j^*\}}$ that have AU_j activated

d: the subscript number of dominant AU.

$\mathcal{C}$: all the subscript numbers of complementary AUs.

T_q: the threshold to separate complementary AUs and non-complementary AUs.

v_m: mental representation, an AU vector that has dominant and complementary AUs activated.

Terms in the second pipeline (IMGA)

GA: Genetic Algorithm

IGA: Interactive Genetic Algorithm

MGA: Microbial Genetic Algorithm

Generation: the full set of the results of a GA iteration.

Population: the complete set of the generated hypotheses after a given iteration. Note that in IMGA, the generation is the population since all the individuals in the current population participate the following biological operator: selection, crossover, and mutation.

Individual: the smallest unit in a population, Note that in IMGA, an individual refers to a facial expression.

Chromosome: a single hypothesis of which many make up a population. In this thesis, a chromosome is a list of action units.

Gene: a single bit within a chromosome. In this thesis, the gene refers to one single action unit.

Fitness: a metric to measure the best fit of an individual (i.e., a hypothesis). Fitness function is used to evaluate each chromosome, and usually best fits can be identified and more heavily relied upon in order to create new generational chromosomes. In IMGA, the fitness function is human subjective judgment rather than an objective (mathematical) function.

Selection: Just like in biological terms, a group of chromosomes (usually with high fitness) are chosen to breed the next generation. In IMGA, all the individuals are selected to breed the next generation.

Crossover: Just like in biological terms, the selected chromosomes exchange their genes. In IMGA, we employed infection, i.e., a single-direction crossover from the source chromosome to the target chromosome. That is to say, a source chromosome infects its genes to the target chromosome whereas the genes of the source chromosome are fixed.

Mutation: Just like in biological terms, some genes of the chromosome may change their values (e.g., from 1 to 0).

Gan: GANimation.

s: the actor's face, the input of the first pipeline.

I: the deformed (generated) face created by GANimation, i.e., stimulus.

v: a 16-dimension binary AU vector.

μ: a list containing the corresponding AU numbers of v.

λ: the binary value representing the activation or deactivation of the AU.

V: vectors of all the possible AU combinations by activating only 3 AUs.

N: number of individuals in the population.

pop_eva: population evaluation module.

inter_pop: inter-population similarity.

intra_pop: intra-population similarity.

corr: Pearson correlation function.

α: positive constant for population evaluation.

β: positive constant for population evaluation.

H: histogram that counts how many times each AU occurs in the corresponding set.

L_g: the set of all the loser AUs in g_th generation.

W_g: the set of all the winner AUs in g_th generation.

$T1, T2, T3$: thresholds for the three-state constraint automaton, indicating when the system can jump from the current state to the next state.

c: the amount of AUs that can be activated, i.e., the constraint.

PUBLICATION

Conference:

Yan, S., Soladie, C., & Seguier, R. (2023, January). Exploring Mental Prototypes by an Efficient Interdisciplinary Approach: Interactive Microbial Genetic Algorithm. In 2023 IEEE 17th International Conference on Automatic Face and Gesture Recognition (FG) (pp. 1-8). IEEE.
URL: hal-04050608[121]

Journal:

Yan, S., Soladié, C., Aucouturier, J. J., & Seguier, R. (2023). Combining GAN with reverse correlation to construct personalized facial expressions. Plos one, 18(8), e0290612.
URL [122]

Appendix: Facial Action Coding System (FACS)

Here are the main coded action units of the Facial Action Coding System (FACS) [37].

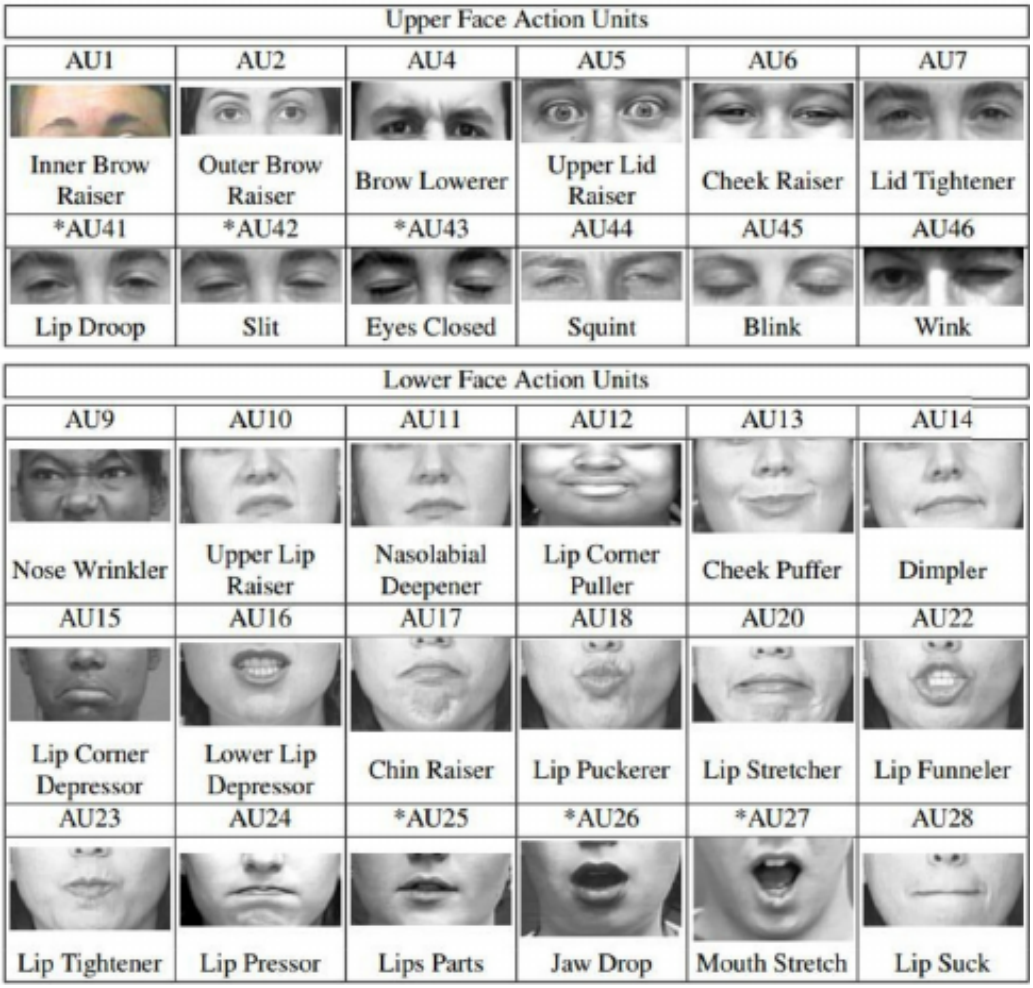


Figure 2 – AUs (main coded) from Facial Action Coding System (FACS) [37]. Figures come from [25].

Appendix: Mental Deep Reverse-Engineering System (MDR)

Please zoom in for better observation, these figures are also available in https://yansen0508.github.io/emotional-prototype/.

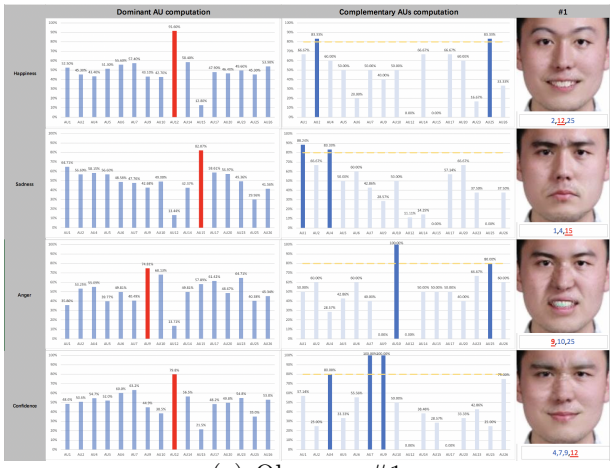
Fig. 3: Bad cases.

Figure 3 – Distortion around the teeth, when AU25 (lips part) is activated.

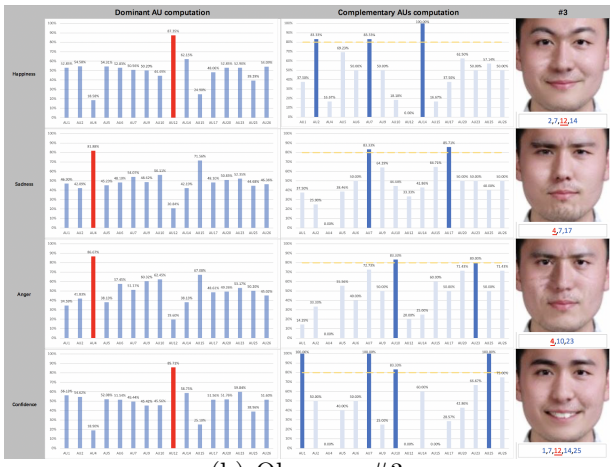
Fig. 4: Mental representation computation from other observers.

Fig. 5: The directed graphs illustrate the voting results for our subjective evaluation by non-observers (Schulze method [98]).

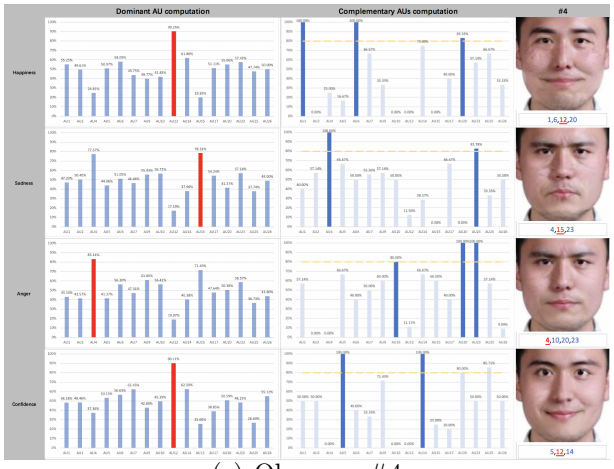
Fig. 6: Converging curves for happiness, sadness, and anger.



(a) Observer #1

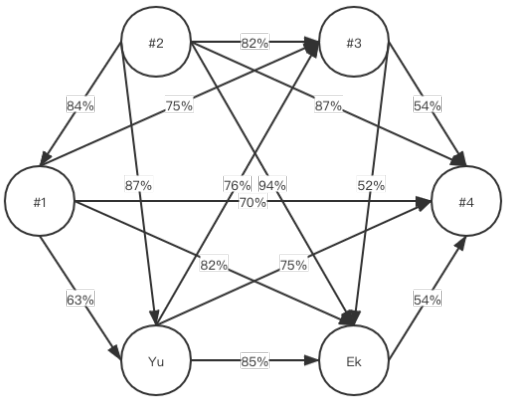


(b) Observer #3

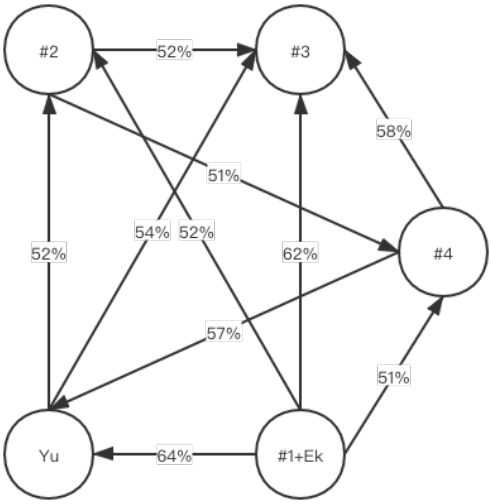


(c) Observer #4

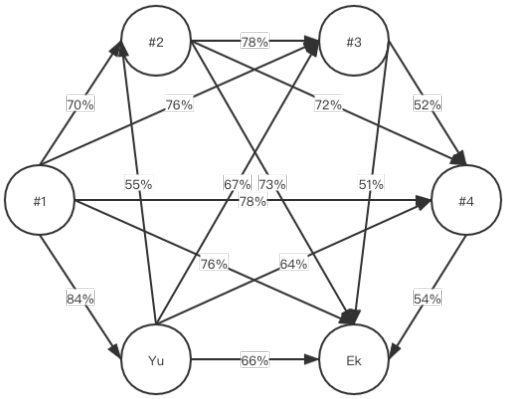
Figure 4 – Mental representation computation from observer #1, #3, and #4.



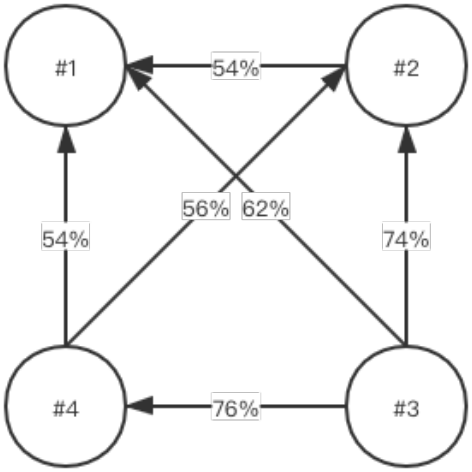
(a) happy



(b) sad

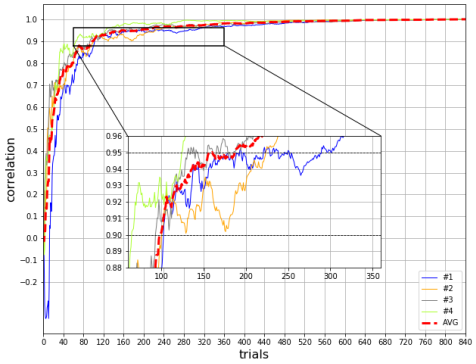


(c) angry

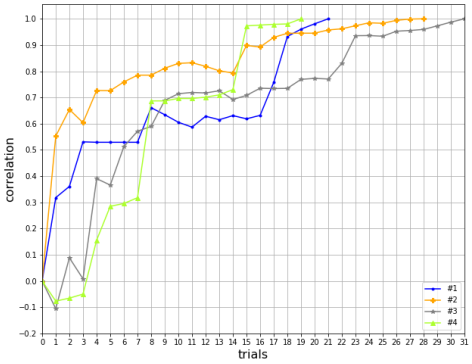


(d) confidence

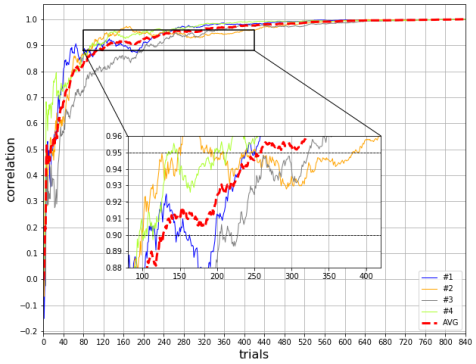
Figure 5 – Directed graphs. We present the voting results for our subjective evaluation by non-observers (Schulze method [98]).



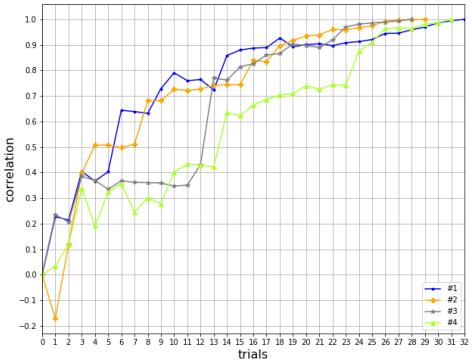
(a) happy-dominant



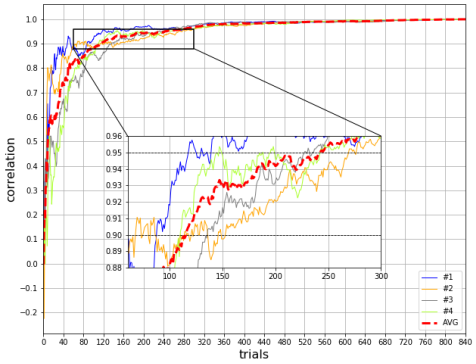
(b) happy-complementary



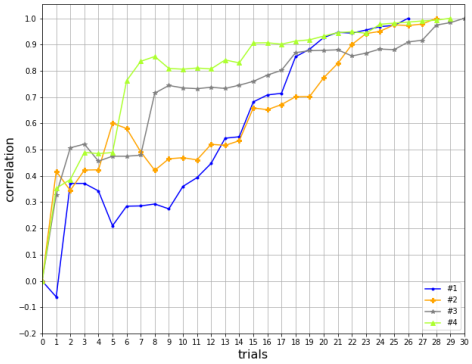
(c) sad-dominant



(d) sad-complementary



(e) angry-dominant



(f) angry-complementary

Figure 6 – Converging curves for happiness, sadness, and anger.

Appendix: Interactive Microbial Genetic Algorithm (IMGA)

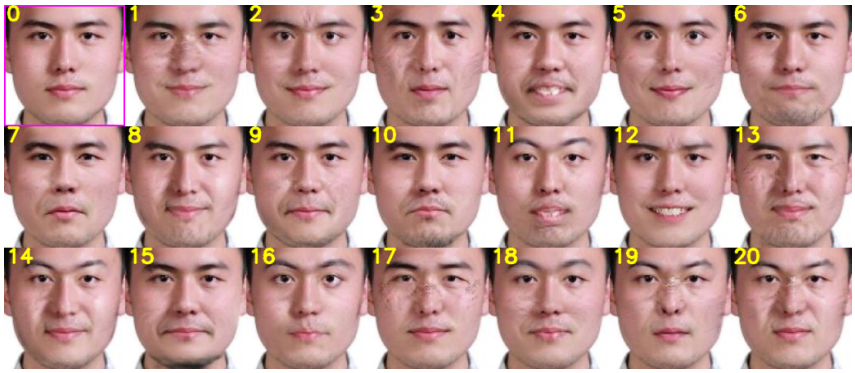


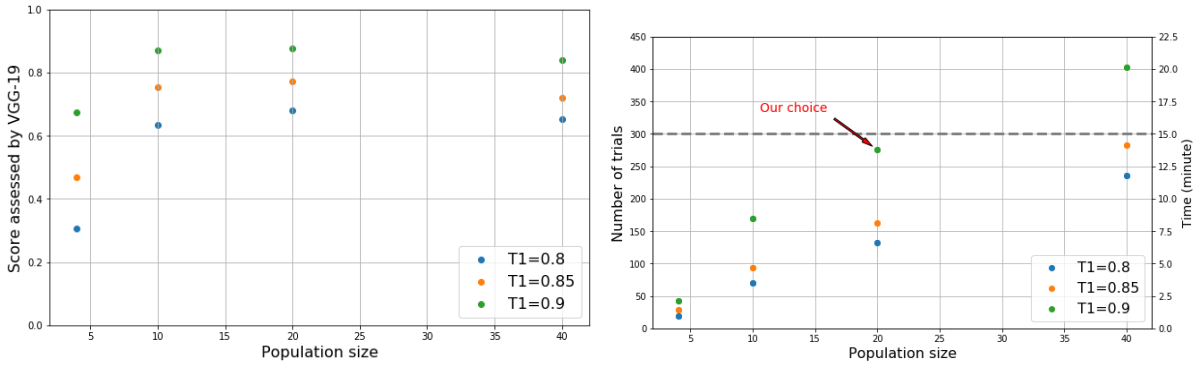
Figure 7 – Example of population initialization. We use GANimation [94] to generate different facial expressions (from #1 to #20) from a neutral face (#0 in red) and different AU (action unit) vectors.

According to the structural order of our manuscript, we present supplementary figures and demonstrations for better understanding. **In Population initialization**: we present an example of population initialization. **In GA parameter settings**: we show the details for setting GA parameters. **In All representative prototypes**: we list all representative prototypes from the 12 observers and the state-of-the-art prototypes [37, 126]. **In Number of trials required for our IMGA**: we detail the number of trials required for our IMGA in each experimental task. **In Subjective evaluation**: For the subjective evaluation of our manuscript, we demonstrate the computation for the baseline (first measurement) and illustrate the entire ranking of representative prototypes (second measurement). In addition, we provide a video demonstration about our approach and the results.

Population initialization

Differing from the majority of facial manipulation techniques that are designed to modify high-level attributes such as hair color, gender, age [17, 66], or emotional expressions

[17, 123], GANimation is trained by the low-level attributes, which are action units (AUs) [37]. Fig. 7 lists an example of initialized generation by GANimation[94]. As the AU vectors were randomly initialized, we can notice that some facial expressions do not correspond to any emotional state, such as #2 with the activation of AU4 (brow lowerer), AU5 (upper lid raiser), and AU12 (lip corner puller).



(a) Score assessed by VGG-19

(b) Number of trials and corresponding times

Figure 8 – We test different combinations of the population size N and the first threshold $T1$ to determine a set of appropriate parameters. 8(a): monitoring the score R_f assessed by VGG-19. 8(b): monitoring the average number of trials required for our IMGA (left Y-axis) and the estimated times (right Y-axis).

GA parameter settings

It is necessary to find a set of relatively appropriate GA parameters before observers perform the perceptual experiments. These parameters are the crossover rate cr, the mutation rate mr, the population size N, the thresholds of the constraint automaton $[T1, T2, T3]$ and constants for the population evaluation α, β. Some parameters were set according to the suggestion from the reference paper [55]: $cr = 0.5$, $mr = 0.03$. Some parameters were taken empirically. Considering user fatigue, the system needs to finally converge in a limited time (we set it to 15 minutes). We need to calibrate the time-sensitive parameters: population size N and thresholds of the constraint automaton $[T1, T2, T3]$, since the first one is related to the number of trials for each iteration and the second one is related to the degree of the system convergence. Empirically, we set $\alpha = \beta = 0.5$ (inter-population similarity and intra-population similarity are both important.) and $T1 = T2 = T3 - 0.05$ (a slightly higher threshold for the stop condition).

We simulated the perceptual experiments by replacing the real observers with an automatic facial expression recognition system. We took the VGG-19 [99] pretrained model as the simulator which outperformed the facial expression recognition tasks in FER2013 dataset [47] (acc=73.112%) and in CK+ dataset [77] (acc=94.64%). The simulators performed the experiments with different GA parameter settings.

For each experimental task of basic emotions (happiness, sadness, anger), the simulator gave each individual in the last generation a score r_i, which is the output value of the corresponding emotional class in the softmax layer of VGG-19. We used the average of all individuals to represent the entire population, $R_k = \frac{1}{N}\sum_{i=1}^{N} r_i$, where N is the population size and k is the simulation number. For each experimental task, the simulator was performed 30 times. Thus, there were, in total, 90 simulations (30 simulations/task $\times$ 3 experimental tasks) for a given set of $N, T1$. We evaluated the simulations by using the average score to represent all populations for the three experimental tasks, $R_f = \frac{1}{90}\sum_{i=1}^{90} R_k$.

With different combinations of the population size N and the first threshold $T1$, we present the final scores (in Fig. 8(a)) and the number of trials required for our IMGA (in Fig. 8(b)). In Fig. 8(a), there are two scores that are rather high: $R_f = 0.87$ (with $N = 10$ and $T1 = 0.9$) and $R_f = 0.88$ (with $N = 20$ and $T1 = 0.9$). Through the attempts of the authors to perform several perception experiments, the response time for each trial was, on average, less than 3 seconds. We set, on average, 3 seconds per trial as the estimated time. Note that the time limit is 15 minutes (dotted line in Fig. 8(b)). We estimated that it took about 8.5 minutes (with $N = 10$, $T1 = 0.9$) and 13.8 minutes (with $N = 20$, $T1 = 0.9$) for one observer to perform one perceptual experiment. Like most genetic algorithms, there is always a trade-off between time and solution diversity. Considering the diversity of the mental prototypes, we finally set $N = 20$ and $T1 = 0.9$. Indeed, in an initialized population of $N = 20$ individuals, each AU has already been activated an average of 3.75 times.

All representative prototypes

In Fig. 9, we list all representative prototypes of observers as well as the state-of-the-art prototypes [37, 126] (in pink) for comparison. Since GANimation [94] does not provide the option to edit AU16 (lower lip depressor), we replace AU16 with AU25 (lips part) to reconstruct the anger prototype of Yu et al.

Figure 9 – Representative prototypes of observers and the state-of-the-art prototypes [37, 126] (noted as Ek. and Yu.). Note that there are no state-of-the-art prototypes of confidence. Since our representative prototype #11-sadness is identical to Ekman-sadness, we merged them and marked them as "Ek/11".

Number of trials required for our IMGA

In Fig. 10, box plots (described by the maximum, the minimum, the median, the average, and the first and third quartiles) show the number of trials performed by observers. The corresponding time for perceptual experiments is shown in Fig. 11. Since the response time for each observer in each trial is different, there are subtle differences between Fig. 10 and

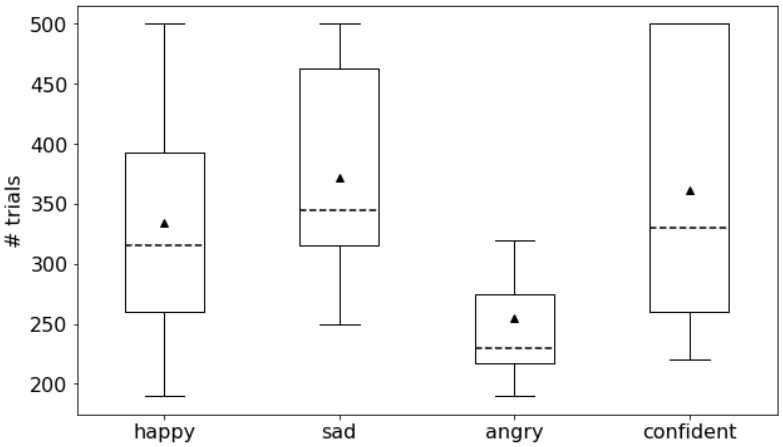


Figure 10 – Box plots: number of trials required for observers to obtain their mental prototypes. The averages for each facial expression are marked by triangles.

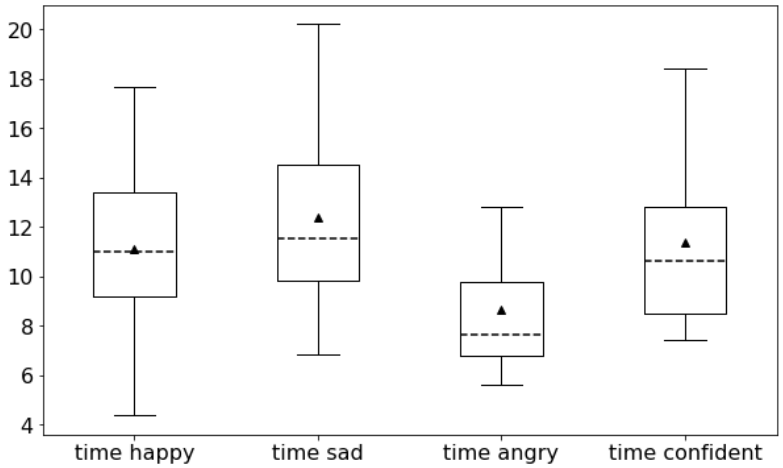
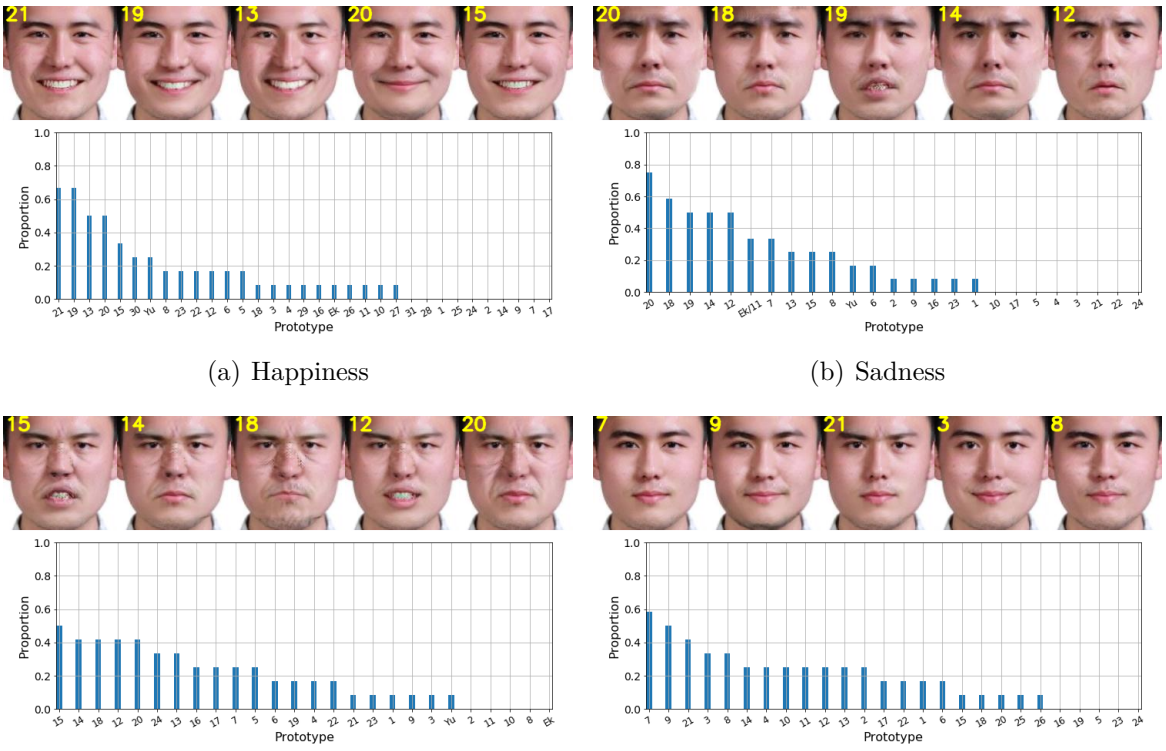


Figure 11 – Box plots: Time (in minutes) for perceptual experiments. The averages for each facial expression are marked by triangles.

Fig. 11. On average, it took about 10.8 minutes (330 trials) for one observer to perform the perceptual experiment.

Subjective evaluation

Here, we present the computation for the baseline of the first measurement: the proportion of observers who still chose at least one representative prototype of theirs. The baseline can be regarded as the probability for one observer who still chose at least one



(a) Happiness (b) Sadness

(c) Anger (d) Confidence

Figure 12 – We displayed by images the top-5 prototypes that observers chose the most. The prototype number is marked on the top left of the image. There is no state-of-the-art prototype appearing in the top-5 prototypes. Since our representative prototype #11-sadness is identical to Ekman-sadness, we merged them and marked them by "Ek/11".

of his/her representative prototypes. For each evaluation task, we define the set of all representative prototypes as P and the set of the representative prototypes of one observer as p_i. We compute the probability for one observer who did not select his/her representative prototypes: $\frac{C_5^{|P|-|p_i|}}{C_5^{|P|}}$, where C is the mathematical combination function and the operator $|.|$ is the cardinality of a set. Hence, we can obtain the probability for one observer who still chose at least one of his/her representative prototypes: $1 - \frac{C_5^{|P|-|p_i|}}{C_5^{|P|}}$. The baseline of one evaluation task is the average probability for the 12 observers: $\frac{1}{12}\sum_{i=1}^{12}\left(1 - \frac{C_5^{|P|-|p_i|}}{C_5^{|P|}}\right)$.

For the second measurement: we sorted all the prototypes according to the proportion selected by observers and displayed the top 5 prototypes with the highest proportions in Fig. 12. The state-of-the-art prototypes are less preferred by observers.

List of Figures

List of Tables

BIBLIOGRAPHY

[1] Reginald B Adams Jr and Robert E Kleck, « Perceived gaze direction and the processing of facial displays of emotion », *in*: *Psychological science* 14.*6* (2003), pp. 644–647.

[2] Al Ahumada Jr and John Lovell, « Stimulus features in signal detection », *in*: *The Journal of the Acoustical Society of America* 49.*6B* (1971), pp. 1751–1756.

[3] Yuntao Bai et al., « Training a helpful and harmless assistant with reinforcement learning from human feedback », *in*: *arXiv preprint arXiv:2204.05862* (2022).

[4] Tadas Baltrušaitis, Peter Robinson, and Louis-Philippe Morency, « OpenFace: An open source facial behavior analysis toolkit », *in*: *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2016, pp. 1–10, DOI: 10.1109/WACV.2016.7477553.

[5] Lisa Feldman Barrett et al., « Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements », *in*: *Psychological science in the public interest* 20.*1* (2019), pp. 1–68.

[6] Eloıse Berson et al., « A robust interactive facial animation editing system », *in*: *Motion, Interaction and Games*, 2019, pp. 1–10.

[7] Rita Berto et al., « An exploratory study of the effect of high and low fascination environments on attentional fatigue », *in*: *Journal of environmental psychology* 30.*4* (2010), pp. 494–500.

[8] Nicola Binetti et al., « Genetic algorithms reveal profound individual differences in emotion recognition », *in*: *Proceedings of the National Academy of Sciences* 119.*45* (2022), e2201380119.

[9] L Brinkman, Alexander Todorov, and R Dotsch, « Visualising mental representations: A primer on noise-based reverse correlation in social psychology », *in*: *European Review of Social Psychology* 28.*1* (2017), pp. 333–361.

[10] Loek Brinkman et al., « Visualizing mental representations in schizophrenia patients: A reverse correlation approach », *in*: *Schizophrenia Research: Cognition* 17 (2019), p. 100138.

[11] Tom Brown et al., « Language models are few-shot learners », *in*: *Advances in neural information processing systems* 33 (2020), pp. 1877–1901.

[12] Juan José Burred et al., « CLEESE: An open-source audio-transformation toolbox for data-driven experiments in speech and music cognition », *in*: *PloS one* 14.*4* (2019), e0205943.

[13] Chen Cao, Qiming Hou, and Kun Zhou, « Displaced dynamic expression regression for real-time facial tracking and animation », *in*: *ACM Transactions on graphics (TOG)* 33.*4* (2014), pp. 1–10.

[14] Chaona Chen et al., « Distinct facial expressions represent pain and pleasure across cultures », *in*: *Proceedings of the National Academy of Sciences* 115.*43* (2018), E10013–E10021.

[15] Xinlei Chen, Saining Xie, and Kaiming He, « An empirical study of training self-supervised vision transformers », *in*: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 9640–9649.

[16] Sung-Bae Cho and Joo-Young Lee, « A human-oriented image retrieval system using interactive genetic algorithm », *in*: *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans* 32.*3* (2002), pp. 452–458.

[17] Yunjey Choi et al., « Stargan: Unified generative adversarial networks for multi-domain image-to-image translation », *in*: *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 8789–8797.

[18] Paul F Christiano et al., « Deep reinforcement learning from human preferences », *in*: *Advances in neural information processing systems* 30 (2017).

[19] Jeffrey F Cohn, « Foundations of human computing: facial expression and emotion », *in*: *Proceedings of the 8th international conference on Multimodal interfaces*, 2006, pp. 233–238.

[20] Timothy F. Cootes, Gareth J. Edwards, and Christopher J. Taylor, « Active appearance models », *in*: *IEEE Transactions on pattern analysis and machine intelligence* 23.*6* (2001), pp. 681–685.

[21] Jean Costa et al., « Regulating feelings during interpersonal conflicts by changing voice self-perception », *in*: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 2018, pp. 1–13.

[22] Sidney K D'Mello, « On the influence of an iterative affect annotation approach on inter-observer and self-observer reliability », *in*: *IEEE Transactions on Affective Computing* 7.*2* (2015), pp. 136–149.

[23] Charles Darwin, *The expression of the emotions in man and animals*, University of Chicago press, 2015.

[24] Charles Darwin and Phillip Prodger, *The expression of the emotions in man and animals*, Oxford University Press, USA, 1998.

[25] Fernando De la Torre et al., « Intraface », *in*: *2015 11th IEEE international conference and Workshops on Automatic Face and Gesture Recognition (FG)*, vol. 1, IEEE, 2015, pp. 1–8.

[26] Abhinav Dhall et al., « Acted facial expressions in the wild database », *in*: *Australian National University, Canberra, Australia, Technical Report TR-CS-11* 2 (2011), p. 1.

[27] Abhinav Dhall et al., « Collecting large, richly annotated facial-expression databases from movies », *in*: *IEEE multimedia* 19.*3* (2012), p. 34.

[28] Abhinav Dhall et al., « Static facial expression analysis in tough conditions: Data, evaluation protocol and benchmark », *in*: *2011 IEEE international conference on computer vision workshops (ICCV workshops)*, IEEE, 2011, pp. 2106–2112.

[29] Djellel Difallah, Elena Filatova, and Panos Ipeirotis, « Demographics and dynamics of mechanical turk workers », *in*: *Proceedings of the eleventh ACM international conference on web search and data mining*, 2018, pp. 135–143.

[30] Uta-Susan Donges, Anette Kersting, and Thomas Suslow, « Women's greater ability to perceive happy facial emotion automatically: gender differences in affective priming », *in*: *PloS one* 7.*7* (2012), e41745.

[31] Shichuan Du, Yong Tao, and Aleix M Martinez, « Compound facial expressions of emotion », *in*: *Proceedings of the National Academy of Sciences* 111.*15* (2014), E1454–E1462.

[32] Shichuan Du, Yong Tao, and Aleix M Martinez, « Compound facial expressions of emotion », *in*: *Proceedings of the national academy of sciences* 111.*15* (2014), E1454–E1462.

[33] Paul Ekman, « An argument for basic emotions », *in*: *Cognition & emotion* 6.*3-4* (1992), pp. 169–200.

[34] Paul Ekman, « Lie catching and microexpressions », *in*: *The philosophy of deception* 1.*2* (2009), p. 5.

[35] Paul Ekman, « Strong evidence for universals in facial expressions: a reply to Russell's mistaken critique. », *in*: (1994).

[36] Paul Ekman and Wallace V Friesen, « Constants across cultures in the face and emotion. », *in*: *Journal of personality and social psychology* 17.*2* (1971), p. 124.

[37] Paul Ekman and Wallace V Friesen, « Facial action coding system », *in*: *Environmental Psychology & Nonverbal Behavior* (1978).

[38] Paul Ekman and Wallace V Friesen, « Nonverbal leakage and clues to deception », *in*: *Psychiatry* 32.*1* (1969), pp. 88–106.

[39] Paul Ekman et al., « Universals and cultural differences in the judgments of facial expressions of emotion. », *in*: *Journal of personality and social psychology* 53.*4* (1987), p. 712.

[40] Jennifer Endres and Anita Laidlaw, « Micro-expression recognition training in medical students: a pilot study », *in*: *BMC medical education* 9.*1* (2009), pp. 1–6.

[41] C Fabian Benitez-Quiroz, Ramprakash Srinivasan, and Aleix M Martinez, « Emotionet: An accurate, real-time algorithm for the automatic annotation of a million facial expressions in the wild », *in*: *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 5562–5570.

[42] William V Gehrlein, « Condorcet's paradox and the likelihood of its occurrence: different perspectives on balanced preferences », *in*: *Theory and decision* 52.*2* (2002), pp. 171–199.

[43] Zhenglin Geng, Chen Cao, and Sergey Tulyakov, « 3d guided fine-grained face manipulation », *in*: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 9821–9830.

[44] Gabriel Goh et al., « Multimodal neurons in artificial neural networks », *in*: *Distill* 6.*3* (2021), e30.

[45] Ian Goodfellow et al., « Generative adversarial nets », *in*: *Advances in neural information processing systems*, 2014, pp. 2672–2680.

[46] Ian Goodfellow et al., « Generative adversarial networks », *in*: *Communications of the ACM* 63.*11* (2020), pp. 139–144.

[47] Ian J Goodfellow et al., « Challenges in representation learning: A report on three machine learning contests », *in*: *Neural Information Processing: 20th International Conference, ICONIP 2013, Daegu, Korea, November 3-7, 2013. Proceedings, Part III 20*, Springer, 2013, pp. 117–124.

[48] Frédéric Gosselin and Philippe G Schyns, « Bubbles: a technique to reveal the use of information in recognition tasks », *in*: *Vision research* 41.*17* (2001), pp. 2261–2271.

[49] Louise Goupil et al., « Listeners' perceptions of the certainty and honesty of a speaker are associated with a common prosodic signature », *in*: *Nature communications* 12.*1* (2021), pp. 1–17.

[50] Ralph Gross et al., « Multi-pie », *in*: *Image and vision computing* 28.*5* (2010), pp. 807–813.

[51] Nadia Guerouaou, Guillaume Vaiva, and Jean-Julien Aucouturier, « The shallow of your smile: the ethics of expressive vocal deep-fakes », *in*: *Philosophical Transactions of the Royal Society B* 377.*1841* (2022), p. 20210083.

[52] Hatice Gunes and Björn Schuller, « Categorical and dimensional affect analysis in continuous input: Current trends and future directions », *in*: *Image and Vision Computing* 31.*2* (2013), pp. 120–136.

[53] Meareg Hailemariam, Ben Goertzel, and Tesfa Yohannes, « Evolving 3D facial expressions using interactive genetic algorithms », *in*: *International Conference on Advances of Science and Technology*, Springer, 2018, pp. 492–502.

[54] Kevin A Hallgren, « Computing inter-rater reliability for observational data: an overview and tutorial », *in*: *Tutorials in quantitative methods for psychology* 8.*1* (2012), p. 23.

[55] Inman Harvey, « The microbial genetic algorithm », *in*: *European conference on artificial life*, Springer, 2009, pp. 126–133.

[56] Kaiming He et al., « Momentum contrast for unsupervised visual representation learning », *in*: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 9729–9738.

[57] Zhenliang He et al., « Attgan: Facial attribute editing by only changing what you want », *in*: *IEEE Transactions on Image Processing* 28.*11* (2019), pp. 5464–5478.

[58] Fang-Cheng Hsu and Peter Huang, « Providing an appropriate search space to solve the fatigue problem in interactive evolutionary computation », *in*: *New Generation Computing* 23.*2* (2005), pp. 115–127.

[59] Gee-Sern Hsu, Chun-Hung Tsai, and Hung-Yi Wu, « Dual-Generator Face Reenactment », *in*: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 642–650.

[60] Phillip Isola et al., « Image-to-image translation with conditional adversarial networks », *in*: *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1125–1134.

[61] Rachael E Jack, Roberto Caldara, and Philippe G Schyns, « Internal representations reveal cultural diversity in expectations of facial expressions of emotion. », *in*: *Journal of Experimental Psychology: General* 141.*1* (2012), p. 19.

[62] Rachael E Jack, Oliver GB Garrod, and Philippe G Schyns, « Dynamic facial expressions of emotion transmit an evolving hierarchy of signals over time », *in*: *Current biology* 24.*2* (2014), pp. 187–192.

[63] Rachael E Jack et al., « Facial expressions of emotion are not culturally universal », *in*: *Proceedings of the National Academy of Sciences* 109.*19* (2012), pp. 7241–7244.

[64] Kerri L Johnson, Masumi Iida, and Louis G Tassinary, « Person (mis) perception: Functionally biased sex categorization of bodies », *in*: *Proceedings of the Royal Society B: Biological Sciences* 279.*1749* (2012), pp. 4982–4989.

[65] Takeo Kanade, Jeffrey F Cohn, and Yingli Tian, « Comprehensive database for facial expression analysis », *in*: *Proceedings fourth IEEE international conference on automatic face and gesture recognition (cat. No. PR00580)*, IEEE, 2000, pp. 46–53.

[66] Tero Karras, Samuli Laine, and Timo Aila, « A style-based generator architecture for generative adversarial networks », *in*: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 4401–4410.

[67] Sourabh Katoch, Sumit Singh Chauhan, and Vijay Kumar, « A review on genetic algorithm: past, present, and future », *in*: *Multimedia Tools and Applications* 80.*5* (2021), pp. 8091–8126.

[68] Hee-Su Kim and Sung-Bae Cho, « Application of interactive genetic algorithm to fashion design », *in*: *Engineering applications of artificial intelligence* 13.*6* (2000), pp. 635–644.

[69] Chih-Chin Lai and Ying-Chuan Chen, « A user-oriented image retrieval system based on interactive genetic algorithm », *in*: *IEEE transactions on instrumentation and measurement* 60.*10* (2011), pp. 3318–3325.

[70] Oliver Langner et al., « Presentation and validation of the Radboud Faces Database », *in*: *Cognition and emotion* 24.*8* (2010), pp. 1377–1388.

[71] Jingting Li, « Facial Micro-Expression Analysis », PhD thesis, CentraleSupélec, 2019.

[72] Shan Li and Weihong Deng, « Deep facial expression recognition: A survey », *in*: *IEEE transactions on affective computing* (2020).

[73] Shan Li, Weihong Deng, and JunPing Du, « Reliable crowdsourcing and deep locality-preserving learning for expression recognition in the wild », *in*: *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2852–2861.

[74] David James Lick et al., « Reverse-correlating mental representations of sex-typed bodies: the effect of number of trials on image quality », *in*: *Frontiers in psychology* 4 (2013), p. 476.

[75] Drew Linsley et al., « What are the visual features underlying human versus machine vision? », *in*: *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2017, pp. 2706–2714.

[76] Ziwei Liu et al., « Deep learning face attributes in the wild », *in*: *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 3730–3738.

[77] Patrick Lucey et al., « The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression », *in*: *2010 ieee computer society conference on computer vision and pattern recognition-workshops*, IEEE, 2010, pp. 94–101.

[78] Michael Lyons et al., « Coding facial expressions with gabor wavelets », *in*: *Proceedings Third IEEE international conference on automatic face and gesture recognition*, IEEE, 1998, pp. 200–205.

[79] Mohammad Mavadati, Peyten Sanger, and Mohammad H Mahoor, « Extended disfa dataset: Investigating posed and spontaneous facial expressions », *in*: *proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2016, pp. 1–8.

[80] S Mohammad Mavadati et al., « Disfa: A spontaneous facial action intensity database », *in*: *IEEE Transactions on Affective Computing* 4.*2* (2013), pp. 151–160.

[81] Albert Mehrabian, « Communication without words », *in*: *Communication theory*, Routledge, 2017, pp. 193–200.

[82] Zbigniew Michalewicz and Marc Schoenauer, « Evolutionary algorithms for constrained parameter optimization problems », *in*: *Evolutionary computation* 4.*1* (1996), pp. 1–32.

[83] Ali Mollahosseini, Behzad Hasani, and Mohammad H Mahoor, « Affectnet: A database for facial expression, valence, and arousal computing in the wild », *in*: *IEEE Transactions on Affective Computing* 10.*1* (2017), pp. 18–31.

[84] Kibum Moon et al., « The Mirror of Mind: Visualizing Mental Representations of Self Through Reverse Correlation », *in*: *Frontiers in Psychology* 11 (2020), p. 1149.

[85] Richard F Murray, « Classification images: A review », *in*: *Journal of vision* 11.*5* (2011), pp. 2–2.

[86] Behnaz Nojavanasghari et al., « Emoreact: a multimodal approach and dataset for recognizing emotional responses in children », *in*: *Proceedings of the 18th acm international conference on multimodal interaction*, 2016, pp. 137–144.

[87] Long Ouyang et al., « Training language models to follow instructions with human feedback », *in*: *arXiv preprint arXiv:2203.02155* (2022).

[88] Maja Pantic et al., « Web-based database for facial expression analysis », *in*: *2005 IEEE international conference on multimedia and Expo*, IEEE, 2005, 5–pp.

[89] Or Patashnik et al., « Styleclip: Text-driven manipulation of stylegan imagery », *in*: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 2085–2094.

[90] Gili Peleg et al., « Hereditary family signature of facial expression », *in*: *Proceedings of the National Academy of Sciences* 103.*43* (2006), pp. 15921–15926.

[91] Adriana Debora Piemonti et al., « Exploration and Visualization of Patterns Underlying Multistakeholder Preferences in Watershed Conservation Decisions Generated by an Interactive Genetic Algorithm », *in*: *Water Resources Research* 57.*5* (2021), e2020WR028013.

[92] Emmanuel Ponsot, Pablo Arias, and Jean-Julien Aucouturier, « Uncovering mental representations of smiled speech using reverse correlation », *in*: *The Journal of the Acoustical Society of America* 143.*1* (2018), EL19–EL24.

[93] Emmanuel Ponsot et al., « Cracking the social code of speech prosody using reverse correlation », *in*: *Proceedings of the National Academy of Sciences* 115.*15* (2018), pp. 3972–3977.

[94] Albert Pumarola et al., « Ganimation: Anatomically-aware facial animation from a single image », *in*: *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 818–833.

[95] Alec Radford et al., « Learning transferable visual models from natural language supervision », *in*: *International conference on machine learning*, PMLR, 2021, pp. 8748–8763.

[96] James A Russell, « A circumplex model of affect. », *in*: *Journal of personality and social psychology* 39.*6* (1980), p. 1161.

[97] James A Russell and Lisa Feldman Barrett, « Core affect, prototypical emotional episodes, and other things called emotion: dissecting the elephant. », *in*: *Journal of personality and social psychology* 76.*5* (1999), p. 805.

[98] Markus Schulze, « A new monotonic, clone-independent, reversal symmetric, and condorcet-consistent single-winner election method », *in*: *Social Choice and Welfare* 36.*2* (2011), pp. 267–303.

[99] Karen Simonyan and Andrew Zisserman, « Very deep convolutional networks for large-scale image recognition », *in*: *arXiv preprint arXiv:1409.1556* (2014).

[100] Lingxiao Song et al., « Geometry guided adversarial facial expression synthesis », *in*: *Proceedings of the 26th ACM international conference on Multimedia*, 2018, pp. 627–635.

[101] Tom Souaille et al., « Extracting Design Recommendations from Interactive Genetic Algorithm Experiments: Application to the Design of Sounds for Electric Vehicles », *in*: *Proceedings of the Design Society* 1 (2021), pp. 1567–1576.

[102] Stefan Steidl et al., « " Of all things the measure is man" automatic classification of emotions and inter-labeler consistency [speech-based emotion recognition] », *in*: *Proceedings.(ICASSP'05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.* Vol. 1, IEEE, 2005, pp. I–317.

[103] Robert C Streijl, Stefan Winkler, and David S Hands, « Mean opinion score (MOS) revisited: methods and applications, limitations and alternatives », *in*: *Multimedia Systems* 22.*2* (2016), pp. 213–227.

[104] Valeriya Strizhkova et al., « Emotion Editing in Head Reenactment Videos using Latent Space Manipulation », *in*: *2021 16th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2021)*, IEEE, 2021, pp. 1–8.

[105] Keita Suzuki et al., « Faceshare: Mirroring with pseudo-smile enriches video chat communications », *in*: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 2017, pp. 5313–5317.

[106] Kevin Sweeney and Cynthia Whissell, « A dictionary of affect in language: I. Establishment and preliminary validation », *in*: *Perceptual and motor skills* 59.*3* (1984), pp. 695–698.

[107] Hideyuki Takagi, « Interactive evolutionary computation: Fusion of the capabilities of EC optimization and human evaluation », *in*: *Proceedings of the IEEE* 89.*9* (2001), pp. 1275–1296.

[108] Hideyuki Takagi, Tomohiro Takahashi, and Ken Aoki, « Applicability of interactive evolutionary computation to mental health measurement », *in*: *2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No. 04CH37583)*, vol. 6, IEEE, 2004, pp. 5714–5718.

[109] Justus Thies et al., « Face2face: Real-time face capture and reenactment of rgb videos », *in*: *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2387–2395.

[110] Christopher A Thorstenson et al., « The role of facial coloration in emotion disambiguation. », *in*: *Emotion* (2021).

[111] Y-I Tian, Takeo Kanade, and Jeffrey F Cohn, « Recognizing action units for facial expression analysis », *in*: *IEEE Transactions on pattern analysis and machine intelligence* 23.*2* (2001), pp. 97–115.

[112] Michel Valstar, Maja Pantic, et al., « Induced disgust, happiness and surprise: an addition to the mmi facial expression database », *in*: *Proc. 3rd Intern. Workshop on EMOTION (satellite of LREC): Corpora for Research on Emotion and Affect*, Paris, France., 2010, p. 65.

[113] Kaisiyuan Wang et al., « Mead: A large-scale audio-visual dataset for emotional talking-face generation », *in*: *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXI*, Springer, 2020, pp. 700–717.

[114] Lina Wang et al., « MGAAttack: Toward More Query-efficient Black-box Attack by Microbial Genetic Algorithm », *in*: *Proceedings of the 28th ACM International Conference on Multimedia*, 2020, pp. 2229–2236.

[115] Nannan Wang et al., « Facial feature point detection: A comprehensive survey », *in*: *Neurocomputing* 275 (2018), pp. 50–65.

[116] Tianxiong Wang and Meiyu Zhou, « A method for product form design of integrating interactive genetic algorithm with the interval hesitation time and user satisfaction », *in*: *International Journal of Industrial Ergonomics* 76 (2020), p. 102901.

[117] Mika Westerlund, « The emergence of deepfake technology: A review », *in*: *Technology Innovation Management Review* 9.*11* (2019).

[118] Wayne Wu et al., « Reenactgan: Learning to reenact faces via boundary transfer », *in*: *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 603–619.

[119] Xingjiao Wu et al., « A survey of human-in-the-loop for machine learning », *in*: *Future Generation Computer Systems* (2022).

[120] Tian Xu et al., « Using psychophysical methods to understand mechanisms of face identification in a deep neural network », *in*: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2018, pp. 1976–1984.

[121] Sen Yan, Catherine Soladié, and Renaud Seguier, « Exploring Mental Prototypes by an Efficient Interdisciplinary Approach: Interactive Microbial Genetic Algorithm », *in*: *2023 IEEE 17th International Conference on Automatic Face and Gesture Recognition (FG)*, IEEE, 2023, pp. 1–8.

[122] Sen Yan et al., « Combining GAN with reverse correlation to construct personalized facial expressions », *in*: *Plos one* 18.*8* (2023), e0290612.

[123] Huiyuan Yang, Zheng Zhang, and Lijun Yin, « Identity-adaptive facial expression recognition through expression regeneration using conditional generative adversarial networks », *in*: *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, IEEE, 2018, pp. 294–301.

[124] Lijun Yin et al., « A 3D facial expression database for facial behavior research », *in*: *7th international conference on automatic face and gesture recognition (FGR06)*, IEEE, 2006, pp. 211–216.

[125] H Peyton Young, « Condorcet's theory of voting », *in*: *The American Political Science Review* (1988), pp. 1231–1244.

[126] Hui Yu, Oliver GB Garrod, and Philippe G Schyns, « Perception-driven facial expression synthesis », *in*: *Computers & Graphics* 36.*3* (2012), pp. 152–162.

[127] AmirAli Bagher Zadeh et al., « Multimodal language analysis in the wild: Cmu-mosei dataset and interpretable dynamic fusion graph », *in*: *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2018, pp. 2236–2246.

[128] Zhihong Zeng et al., « A survey of affect recognition methods: audio, visual and spontaneous expressions », *in*: *Proceedings of the 9th international conference on Multimodal interfaces*, 2007, pp. 126–133.

[129] Jiangning Zhang et al., « Freenet: Multi-identity face reenactment », *in*: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 5326–5335.

[130] Guoying Zhao et al., « Facial expression recognition from near-infrared videos », *in*: *Image and vision computing* 29.*9* (2011), pp. 607–619.

[131] Jun-Yan Zhu et al., « Unpaired image-to-image translation using cycle-consistent adversarial networks », *in*: *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2223–2232.

DOCTORAT BRETAGNE LOIRE / MATHSTIC

CentraleSupélec

Titre : Personnaliser les expressions faciales en explorant les prototypes mentaux émotionnels

Mot clés : Manipulation d'expressions faciales, Réseaux adversaires génératifs, Corrélation renversée, Calcul interactif, Algorithme génétique

Résumé : Les expressions faciales sont une forme essentielle de communication non verbale. Aujourd'hui, les techniques de manipulation des expressions faciales (FEM) ont envahi notre quotidien. Cependant, dans le contexte de l'application, plusieurs exigences doivent être satisfaites. Diversité : les prototypes d'expression faciale doivent être multiples et différents selon les utilisateurs. Flexibilité : les expressions faciales doivent être personnalisées, c'est-à-dire que le système peut trouver le prototype d'expression faciale qui répond aux besoins des utilisateurs. Exhaustivité : la plupart des technologies FEM ne peuvent traiter que les six émotions de base, alors qu'il existe plus de 4000 émotions dans le monde réel. Absence d'expertise : le système FEM doit pouvoir être contrôlé par n'importe qui sans nécessiter de connaissances spécialisées (par exemple, des psychologues). Efficacité : le système avec interaction doit tenir compte de la fatigue de l'utilisateur.

Dans cette thèse, pour répondre à toutes les exigences, nous avons proposé une approche interdisciplinaire en combinant les réseaux adversaires génératifs avec le processus de corrélation renversée psychophysique. De plus, nous avons créé un algorithme génétique microbien interactif pour optimiser l'ensemble du système.

Title: Personalizing facial expressions by exploring emotional mental prototypes

Keywords: Facial expression manipulation, Generative adversarial networks, Reverse correlation, Interactive computation, Genetic algorithm

Abstract: Facial expressions are an essential form of nonverbal communication. Now facial expression manipulation (FEM) techniques have flooded our daily lives. However, in the application context, there are several requirements that need to be addressed. Diversity: facial expression prototypes should be multiple and different between different users. Flexibility: facial expressions should be personalized, i.e., the system can find the facial expression prototype that can meet the need of the users. Exhaustiveness: most FEM technologies can only deal with the six basic emotions, whereas there are more than 4000 emotion labels. Expertise-free: the FEM system should be controllable by anyone without the need for expert knowledge (e.g., psychologists). Efficiency: the system with interaction should consider user fatigue.

In this thesis, to fulfill all the requirements, we proposed an interdisciplinary approach by combining generative adversarial networks with the psychophysical reverse correlation process. Moreover, we created an interactive microbial genetic algorithm to optimize the entire system.