



ÉCOLE POLYTECHNIQUE  
FÉDÉRALE DE LAUSANNE

Projet de semestre - Hiver 2006-2007

Professeur : Michel BIERLAIRE

Assistant : Matteo SORCI

# Modèles de choix discrets pour la reconnaissance des expressions faciales statiques

Antonin DANALET

<antonin.danalet@epfl.ch>

Lausanne, le 8 mars 2007

*Le visage est l'image de l'âme.*

Cicéron

*Si le visage est le miroir de l'âme, alors  
il y a des gens qui ont l'âme bien laide.*

Gustave Flaubert

*Les visages souvent sont de doux  
imposteurs. Que de défauts d'esprit se  
couvrent de leur grâce ! Et que de beaux  
semblants cachent des âmes basses !*

Pierre Corneille

## Résumé

Dans ce projet de semestre, je présente l'utilisation des modèles de choix discret pour construire un modèle de perception des expressions faciales statiques potentiellement utilisable pour la reconnaissance et la classification de ces expressions. Pour décrire ces expressions, je m'inspire du système *Facial Action Coding System* (FACS) de Paul Ekman, basé sur une analyse anatomique de l'action faciale. L'ensemble de choix contient 6 expressions faciales universelles plus l'expression neutre. Chaque alternative est décrite par une fonction d'utilité, définie par une combinaison linéaire de termes exprimant les *Action Units* (AU) du système FACS et par un terme aléatoire capturant l'incertitude. Les utilités permettent une mesure de la vraisemblance pour une combinaison de ces termes de représenter une certaine expression faciale. Elles sont une formalisation de la connaissance *a priori* du processus. Les paramètres de la fonction d'utilité sont définis par une estimation du maximum de vraisemblance. La classification pourrait être obtenue en assignant l'expression obtenant l'utilité maximum.

# Table des matières

<b>Introduction</b>	<b>2</b>
<b>1 La reconnaissance des expressions faciales</b>	<b>5</b>
1.1 L'état de l'art . . . . .	6
1.2 Les 6 expressions faciales universelles . . . . .	7
1.3 Le système FACS . . . . .	8
<b>2 Les modèles de choix discrets</b>	<b>10</b>
2.1 Application à la reconnaissance faciale . . . . .	12
2.2 Le sondage d'évaluation des expressions faciales . . . . .	13
<b>3 Analyse du modèle</b>	<b>16</b>
3.1 Traduction du modèle FACS . . . . .	17
3.2 Définition des fonctions d'utilité . . . . .	20
3.3 Analyse du modèle avec Biogeme . . . . .	23
<b>Conclusion</b>	<b>27</b>
<b>Bibliographie</b>	<b>32</b>

# Introduction

La pensée occidentale se caractérise par une séparation du corps et de l'esprit. Elle existait déjà aux débuts de notre tradition de pensée, d'abord en Grèce puis dans l'Occident chrétien, mais c'est seulement à l'époque classique, dans le sillage du cartésianisme, que s'établit un véritable dualisme entre le corps et l'âme. Dès lors, ce n'est que tardivement que l'Occident et ses chercheurs se sont préoccupés du langage du corps. Il a longtemps constitué un domaine de prédilection pour les artistes - peintres, sculpteurs, poètes - et les émotions ont été considérées comme un obstacle à la science et à la rationalité.

Le XIX<sup>e</sup> siècle voit apparaître les précurseurs. Charles Darwin se préoccupe du code des gestes, dans son ouvrage *Expression des émotions chez l'homme et les animaux*, paru en 1872 et qui reste encore une référence pour bon nombre de scientifiques. Il y défend l'idée que les émotions sont innées, universelles et identiques chez tous les primates. A la même époque, le neurologue Guillaume-Benjamin Duchenne de Boulogne réalise une série d'expérience sur l'expression faciale de l'émotion et rédige en 1862 *Mécanisme de la physionomie humaine, ou Analyse électro-physiologique de l'expression des passions applicable à la pratique des arts plastiques*. Il utilise la stimulation électrique pour mettre en évidence les mouvements associés à l'expression des émotions. Cela lui permet par exemple de conclure qu'un sourire de bonheur est formé non seulement par les muscles buccaux mais aussi par les muscles oculaires. De tels sourires, authentiques, sont appelés "sourire de Duchenne" en son honneur.

Désormais, l'intérêt pour l'étude des expressions faciales et plus généralement pour la "communication non-verbale" est important.

Albert Mehrabian, psychologue américain, a posé que dans les situations de communication, 7% du contenu est donné par le sens des mots, 38% par la façon dont les mots sont prononcés et 55% par l'expression du visage [25].

Selon l'école de Palo Alto, les moyens verbaux ne porteraient que sur 30 à 35% de la signification du message, tandis que les moyens non verbaux véhiculeraient 65 à 70% de l'information [38].

Par conséquent, la communication non verbale est bien une source d'information dans la communication entre deux personnes.



FIG. 1 – Duchenne déclenchant une expression de frayeur par la stimulation électrique

Les expressions faciales sont généralement reconnues comme étant l'un des éléments les plus importants dans ce processus de communication [6, 11, 16, 21, 29, 31] et sont souvent perçues comme étant une source importante d'information concernant l'état émotionnel d'un autre individu durant une interaction sociale [8, 12, 19, 26]. Les expressions faciales peuvent aussi permettre, tout comme d'autres indices non-verbaux, de distinguer entre certaines émotions spécifiques et procurer de l'information à propos de l'intensité des émotions ressenties [4, 14, 18, 36].

L'importance des expressions faciales pour la communication est maintenant reconnue par plusieurs auteurs [11, 21, 29].

Dans ce contexte, il est légitime de tenter de modéliser le processus de perception humaine des expressions faciales, et ainsi permettre la reconnaissance de ces dernières.

Si l'on prend pour exemple l'interface homme-machine, elle passe actuellement principalement par le clavier et la souris. On a vu Steve Jobs nous présenter l'interface du nouvel iPhone uniquement avec les doigts ou encore Spielberg nous présenter ce qu'il considérerait comme l'interface graphique du futur dans A.I., fondée sur les mouvements des bras de l'utilisateur, mais à terme cette interface va tendre vers une approche de plus en plus intuitive, c'est-à-dire de plus en plus proche d'une interaction humaine. Elle va donc

évoluer vers ce qui a été décrit plus haut : une communication où non seulement les gestes et la parole auront une importance, mais aussi les expressions faciales.

Et l'interface homme-machine n'est pas la seule application de la reconnaissance des expressions faciales. On peut aussi imaginer une application dans des domaines tels que l'enseignement (transmission au professeur de l'état des étudiants sous forme d'information de haut niveau), la détection de mensonges, la création d'avatars, le soutien au conducteur en cas de perte de vigilance, la reconnaissance de visages invariante à l'expression ou encore comme indicateur de l'activité émotionnelle moins intrusive que les méthodes actuelles.

Il ne faut cependant pas confondre ici les émotions et les expressions faciales. Les expressions faciales ne sont qu'un moyen parmi d'autres d'exprimer une émotion. Celle-ci peut aussi s'exprimer dans la voix, les gestes, la posture, la direction du regard, ...

Une émotion peut causer une expression faciale ou non, et inversement, une expression faciale peut être causée par une émotion ou non.

Les expressions faciales peuvent dépendre de l'activité physiologique (besoins biologiques, douleur, fatigue), de l'état d'esprit (émotion, réflexion) et fait pleinement partie de la communication non-verbale (émotion simulée, clin d'oeil, froncement des sourcils).

Nous nous intéressons ici uniquement à la reconnaissance des expressions faciales.

Le but de ce projet est donc d'étudier la question de la reconnaissance des expressions faciales à partir d'images statiques dans le cadre de réflexion des modèles de choix discrets. Je ne m'intéresserai cependant qu'à construire un modèle de perception des expressions faciales. Ce modèle pourra être utilisé par la suite pour la reconnaissance et la classification à proprement parler, mais je me limite dans ce travail à la première partie.

Dans une première partie, je vais présenter de manière générale la question de la reconnaissance faciale, puis les modèles de choix discrets pour finalement présenter et analyser le modèle utiliser tout au long de ce projet de semestre.

# Chapitre 1

## La reconnaissance des expressions faciales

Les images faciales sont difficiles à interpréter car elles sont très variables. Les sources de variation proviennent de l'apparence individuelle variable sur la population humaine, de la position dans l'espace, du contexte externe et de la luminosité. Un système automatique de reconnaissance des expressions faciales nécessite la résolution de trois problèmes : la détection et la localisation des visages dans une image, l'extraction de données à partir du visage et la classification des expressions faciales.

Une fois qu'un visage est détecté, la zone est extraite du reste de l'image, puis normalisé pour avoir la même taille et les mêmes niveaux de gris. Je ne m'intéresserais que très peu au problème de la détection dans ce rapport.

L'extraction de données à partir du visage tente de définir une représentation la plus appropriée pour la reconnaissance faciale des expressions. Il y a principalement deux approches : les systèmes holistiques qui tentent de faire correspondre le visage à certains gabarits et les systèmes géométriques basés sur les traits du visage [5]. Les systèmes holistiques traitent l'image comme un tout, et en définissent certaines caractéristiques. Ils tirent de l'image une représentation sous-dimensionnée. Dans les systèmes géométriques, les caractéristiques du visage sont détectées directement dans le visage. Les distances entre certains points et les tailles relatives définissent les données. Ces points définissent une représentation graphique du visage. Dans ce travail, nous avons opté pour une approche géométrique. Le choix de l'ensemble des données caractéristiques est crucial pour la suite de l'analyse.

Reste ensuite à définir une règle de classification des expressions faciales à partir de ces données caractéristiques. Cette règle est apprise à partir d'un ensemble d'images d'entraînement dont l'expression est connue. Ensuite, pour les nouvelles expressions, en phase d'exécution, la représentation est associée avec l'expression correspondante.



## 1.1 L'état de l'art

En comparaison de la reconnaissance faciale, il y a relativement peu de travaux sur la reconnaissance des expressions faciales. Ils se divisent en deux catégories.

Tout d'abord, ceux utilisant des séquences d'images. Suwa et al. [34] font une analyse préliminaire des expressions faciales en suivant vingt taches identifiées. Mase [24] utilise des moyennes et des variances de données de mouvement ("optical flow") sur des petits blocs également divisés. Yaccob et Davis [39] utilisent les mouvements des bords du nez, de la bouche, des yeux et des sourcils. Essa et Pentland [15] construisent un modèle paramétrique dynamique en suivant les mouvements du visage et l'utilisent pour analyser les expressions faciales.

La deuxième catégorie tente de classifier les expressions faciales à partir d'images statiques. C'est ce qui nous intéresse ici. Turk and Pentland [37] représentent les images faciales par un ensemble de composants faciaux standardisés (les "eigenfaces") et effectuent une analyse linéaire des composants principaux. Padgett et Cottrell [27] utilisent une approche similaire, utilisant les "eigenfaces" sur certaines régions du visage (les deux yeux et la bouche). De leur côté, Lanitis et al. [22] tiennent compte à la fois des formes et des niveaux de gris en utilisant des descripteurs déformables paramétrisés et en effectuant une analyse statistique sur un ensemble d'entraînement d'images faciales.

Certains [23] utilisent des modèles de Markov cachés, ou automate à état caché de Markov, un procédé utilisé dans la reconnaissance de formes ou le traitement automatique du langage par exemple. D'autres [40] des filtres se basant sur des ondelettes de Gabor, permettant de construire des modèles pour les expressions faciales. D'autres encore [2, 28] l'analyse en composantes principales et indépendantes ou les réseaux de neurones.

Ces dernières années ont vu l'utilisation de plus en plus fréquente des méthodes basée sur un modèle (*model based method* en anglais) [7, 22]. Le modèle actif d'apparence (Active Appearance Model [33]) est une de ces techniques. Il s'agit d'un algorithme de vision par ordinateur qui permet de faire correspondre à une image une forme selon un modèle statistique. Il est construit à partir d'une phase d'entraînement, et combine des modèles de formes et de texture, pour finalement produire un masque de point sur le visage (cf fig. 1.1).

C'est à partir de ces 55 points que je définirai par la suite les caractéristiques du visage.

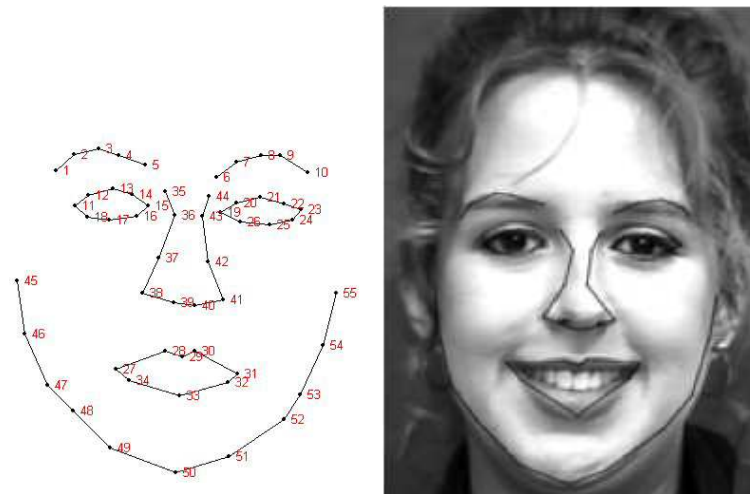


FIG. 1.1 – Masque de point produit par le modèle actif d'apparence (55 points) [1]

## 1.2 Les 6 expressions faciales universelles

La plupart des études actuelles sur les émotions trouvent leur origine dans les travaux de Darwin. Sa thèse, *The Expression of Emotion in Man and Animals* [9], avance que les émotions et leurs expressions ont évolué à travers les espèces, sont innées biologiquement et surtout sont universelles pour l'ensemble des humains et même des primates. Selon Darwin, tous les humains, quelle que soit leur culture et leur origine, expriment les émotions de la même manière.

Les travaux de Darwin ont donné lieu à beaucoup de critiques, la principale étant que ce qu'il avançait n'était pas soutenu par des preuves tangibles. Cependant, jusque vers 1960, peu d'étude ont formalisé et structuré cette recherche. De plus, jusqu'au milieu du XX<sup>e</sup> siècle, il était largement acceptée, et en particulier en anthropologie, que les émotions se différençaient selon les cultures.

Au milieu des années 1960, un psychologue, Sylvan Tomkins, et deux de ces élèves, Paul Ekman et Carrol Izard, effectuèrent une étude, connue sous le nom de *Universality Studies*, sur les émotions humaines à travers différentes cultures et obtinrent des jugements de visages censés exprimer des émotions transculturelles. Alors que la plupart des psychologues pensaient que les expressions étaient culturellement déterminées et dépendaient de conventions sociales, Paul Ekman voyagea au Japon, au Brésil, en Argentine, emmenant avec lui des images de femmes et d'hommes. Partout il eut les mêmes réactions. Et pour vérifier qu'il ne s'agissait pas de règles culturelles occidentales apprises en regardant les mêmes films et les mêmes séries TV, il fit une

deuxième tournée mondiale, dans la jungle de la Papouasie-Nouvelle-Guinée, dans des villages éloignés de tout. Et toujours les mêmes conclusions. Il en tira six expressions universelles : la peur, le dégoût, la colère, le bonheur, la tristesse et la surprise, sur lesquelles tous les individus s'accordaient, quelle que soit leur culture [10]. Dès lors, et suite à d'autres études de Ekman, par exemple sur des aveugles congénitaux qui montraient les mêmes expressions, une base universelle pour les expressions émotionnelle n'est plus débattue dans la communauté des psychologues. Ekman a montré que les expressions faciales étaient en partie un résultat universel de l'évolution, comme Darwin l'avait avancé.

Depuis, Paul Ekman est le principal théoricien des émotions. Il a développé un ensemble de critères qui selon lui définit et délimite ce qu'il considère comme une émotion. Ses *Universality Studies* ont beaucoup influencé la recherche sur les émotions, mais ont néanmoins été sujette à critiques [17, 30].

Actuellement, Paul Ekman entraîne le personnel de sécurité de certains aéroports à lire les mauvaises intentions sur le visage des passagers, et a développé pour cela une nouvelle technique, "Spot", acronyme de "Screening Passengers by Observation Techniques" [35].

### 1.3 Le système FACS

Se basant sur l'universalité, Ekman et Izard ont développé des méthodes de mesure des comportements du visage. En particulier, ils ont créé le système FACS (Facial Action Coding System), largement utilisé et reconnu. Il utilise une quarantaine de caractéristiques anatomiques indépendantes, et définit une taxonomie de toutes les expressions faciales. Le Système FACS est probablement le standard le plus populaire utilisé pour classer systématiquement les expressions physique des émotions du visage, et est utilisé par des psychologues mais aussi par des infographistes.

Ce système définit 46 *Action Units*, qui sont autant de contraction ou de relaxation d'un ou plusieurs muscles, et dont l'association définit une expression faciale. La philosophie de base du système consistait à former des experts pour la reconnaissance et l'interprétation des *Action Units*, mais désormais, le système est aussi utilisé pour automatiser la reconnaissance des *Action Units* et donc des expressions, ainsi que pour la simulation graphique de visages.

Des modèles tels Candide<sup>1</sup> ou Artnatomy<sup>2</sup> permettent non pas d'extraire d'un visage des *Action Units* mais au contraire de donner à un visage l'apparence de telle ou telle *Action Unit*.

---

<sup>1</sup><http://www.bk.isy.liu.se/candide/>

<sup>2</sup><http://www.artnatomia.net/uk/index.html>



















AU1  Inner Brow Raiser	AU2  Outer Brow Raiser	AU4  Brow Lowerer	AU5  Upper Lid Raiser	AU6  Cheek Raiser	AU7  Lid Tightener
AU9  Nose Wrinkler	AU10  Upper Lip Raiser	AU12  Lip Corner Puller	AU15  Lip Corner Depressor	AU16  Lower Lip Depressor	AU17  Chin Raiser
AU20  Lip Stretcher	AU23  Lip Tightener	AU24  Lip Pressor	AU25  Lips part	AU26  Jaw Drop	AU27  Mouth Stretch

FIG. 1.2 – Liste des *Action Units* relatives aux 6 expressions faciales [41]

Les "experts FACS" dissèquent une expression observée et la décomposent en *Action Units* produisant le mouvement. Pour une expression faciale, le résultat consiste en une liste de *Action Units* qui l'ont produite. La durée, l'intensité et l'asymétrie peuvent aussi être enregistrées. Il s'agit donc d'une manière de décrire précisément le visage.

Cette "expertise" est "vendue" par Ekman à travers le Manuel FACS [13]. Il s'agit d'un "guide détaillé et technique expliquant comment catégoriser les comportements faciaux en se basant sur les muscles qui les produisent, autrement dit comment l'action musculaire est liée à l'apparence faciale. (...) Le Manuel FACS permet entre autre une plus grande conscience et sensibilité aux comportements subtils du visage qui peuvent être utiles pour les psychothérapeutes, les interviewers, et tout autre praticien qui doit pénétrer profondément dans la communication interpersonnelle".

Dans ce travail, je fonde mon analyse sur les *Action Units* tels que décrit dans le système FACS, mais il n'y a pas ici d'experts. La connaissance est celle de tous, obtenu par un sondage. C'est à partir des résultats du sondage, à partir de l'avis des sondés, que le modèle s'adaptera et tentera d'appréhender le rôle de chaque *Action Unit* dans la description d'une expression.

Ces *Action Units* vont être transcrites en fonction des 55 points du masque obtenu par le modèle actif d'apparence.

## Chapitre 2

# Les modèles de choix discrets

Les modèles de choix discrets sont utilisés depuis les années 1950 en économétrie. Ils servent à décrire le comportement d'individus face à une situation de choix, lorsque l'ensemble des alternatives disponibles est fini et discret (ensemble de choix). Cette théorie se fonde sur le concept de la maximisation de l'utilité en économie, où l'individu qui choisit est supposé rationnel et capable de faire le choix qui maximisera l'utilité qu'il en retirera. Les alternatives sont supposées mutuellement exclusives et globalement exhaustives, alors que la rationalité des preneurs de décision implique des préférences transitives et cohérentes. On suppose donc une règle de choix s'appuyant sur un opérateur dit de préférence-indifférence, qui a comme propriété d'être réflexif, transitif et que la comparabilité soit garantie<sup>1</sup>. On obtient dès lors un ordre total, et par conséquent cette règle de choix peut s'exprimer numériquement :

$$\exists U_n : C_n \longrightarrow \mathbb{R} : a \longmapsto U_n(a) \quad \text{tel que} \quad a \succsim b \Leftrightarrow U_n(a) \geq U_n(b) \quad \forall a, b \in C_n$$

où  $U_n$  est la fonction d'utilité de l'alternative  $n$ , associant à un choix un valeur numérique, l'utilité de l'alternative. L'alternative  $i$  est choisie si  $U_i \geq U_j \quad \forall j$ .

L'utilité est un concept latent, considérée comme une variable aléatoire et n'est pas directement observée. Cette approche peut-être interprétée comme une tentative de modéliser le processus de décision d'un observateur humain pendant un processus de choix, en supposant que l'analyste a la connaissance de tous les attributs significatifs, une connaissance parfaite de l'opérateur de préférence-indifférence et qu'il ne fait pas d'erreur de mesure.

Dans la réalité, l'analyste manque d'information, que ce soit à cause d'erreurs de mesure ou d'attributs non observés, au niveau des attributs spécifique aux alternatives ou des caractéristiques socio-économiques, des

---

<sup>1</sup>Plus précisément, en notant l'opérateur  $\succsim$ , on a : (i)  $a \succsim a \quad \forall a \in C_n$ , (ii)  $a \succsim b$  et  $b \succsim c \Rightarrow a \succsim c \quad \forall a, b, c \in C_n$ , (iii)  $a \succsim b$  ou  $b \succsim a \quad \forall a, b \in C_n$  où  $C_n$  est l'ensemble de choix

goûts d'une groupe d'individus. C'est pourquoi on utilise des modèles d'utilité aléatoire. Le choix de l'utilisateur est déterminé par un certain nombre de facteurs. Ces facteurs sont pour une part observés par l'analyste, d'autres pas. Appelons les facteurs observés  $x$  et ceux non observés  $\varepsilon$ . Les facteurs sont liés au choix par la fonction  $y = h(x, \varepsilon)$ . Comme  $\varepsilon$  n'est pas observée, le choix n'est pas déterministe et ne peut être prédit exactement. Le terme non observé est considéré comme une variable aléatoire de distribution  $f(x)$ . Par conséquent, la probabilité que le preneur de décision choisisse une alternative  $y$  vaut la probabilité que le facteur non observé soit tel que le processus de choix soit observé :  $P(y|x) = \text{Prob}(\varepsilon \text{ tel que } h(x, \varepsilon) = y)$ . On a donc :

$$\begin{aligned} P(y|x) &= \text{Prob}(I[h(x, \varepsilon) = y] = 1) \\ &= \int I[h(x, \varepsilon) = y] f(\varepsilon) d\varepsilon \end{aligned}$$

Plus précisément, étant donné une population de  $N$  individus et un ensemble de choix  $C_n$  de cardinalité  $J$ , la fonction d'utilité  $U_{in}$  de l'alternative  $i$  perçue par l'individu  $n$  est définie par :

$$U_{in} = V_{in} + \varepsilon_{in}$$

avec  $i = 1, \dots, J$  et  $n = 1, \dots, N$ .  $V_{in}$  représente la partie déterministe de l'utilité, qui est une fonction des attributs de l'alternative et des caractéristiques socio-économiques du preneur de décisions. Le terme  $\varepsilon_{in}$  est une variable aléatoire capturant l'incertitude. Sous l'hypothèse de la maximisation de l'utilité, le résultat du modèle consiste en la probabilité que l'individu  $n$  choisisse l'alternative  $i$  parmi l'ensemble de choix  $C_n$ . Elle est donnée par :

$$\begin{aligned} P_n(i|C_n) &= P(U_{in} = \max_{j \in C_n} U_{jn}) \\ &= P_n(U_{in} \geq U_{jn}, \forall j \in C_n, j \neq i) \\ &= \int_{\varepsilon_n} I(\varepsilon_n < V_{in} - V_{jn}, \forall j \in C_n, j \neq i) f(\varepsilon_n) d\varepsilon_n \end{aligned}$$

où  $\varepsilon_n = \varepsilon_{jn} - \varepsilon_{in}$  et  $I$  est la fonction indicatrice. Selon cette équation, de manière à définir la probabilité de choix, seule la différence entre les utilités intervient. La spécification de la fonction d'utilité représente le moyen du modélisateur d'ajouter sa connaissance *a priori* du processus de choix. Différents modèles de choix discret peuvent être obtenus en faisant différentes hypothèses sur les termes d'erreur. Une famille de modèles largement utilisée est les modèles GEV (*Generalized Extreme Value*). Les modèles GEV proposent une solution proche de l'intégrale de probabilité tout en permettant une certaine flexibilité dans la structure variance / covariance du problème. En supposant que les termes d'erreur soient multivariés de type I distribués selon un modèle aux valeurs extrêmes, l'expression générale de la probabilité

du choix dans le modèle GEV pour un individu donné de choisir l'alternative  $i$ , étant donné l'ensemble de choix  $C$  de  $J$  alternatives, est :

$$P(i|C) = \frac{e^{V_i + \log G_i(y_1, \dots, y_J)}}{\sum_{j=1}^J e^{V_j + \log G_j(y_1, \dots, y_J)}}$$

où  $y_i = e^{V_i}$  et  $G_i = \frac{\partial G}{\partial y_i}$ . La fonction  $G$  est appelée *fonction génératrice* et capture les corrélations entre les alternatives. Plusieurs modèles GEV peuvent être dérivés de cette équation, selon les spécifications différentes de la fonction génératrice. Plus tard, nous allons utiliser un modèle multinomial logit (MNL, *Multinomial Logit Model*), qui est de loin le plus simple et le plus utilisé des modèles de choix discret. Il est obtenu en supposant la fonction  $G$  suivante, qui implique qu'il n'y a aucune corrélation entre les alternatives :

$$G(y_1, \dots, y_J) = \sum_{j \in C} y_j^\mu$$

où  $\mu$  est un paramètre d'échelle positif. Sous ces hypothèses, la probabilité de choix est donnée par :

$$P(i|C) = \frac{e^{\mu V_i}}{\sum_{j \in C} e^{\mu V_j}}$$

## 2.1 Application à la reconnaissance faciale

Le problème de choix qui se pose dans ce cas consiste à choisir l'expression faciale exprimée par un visage. Nous allons tenter de modéliser le processus de choix d'une expression faciale par un individu. Le preneur de décision est tout un chacun, quelle que soit sa culture, son âge, son genre, etc.

L'ensemble de choix considéré ici contient les 6 expressions faciales universelles présentées plus haut. Chaque individu peut choisir parmi l'ensemble universel des 6 alternatives, auquel on a ajouté la possibilité d'exprimer la neutralité de l'expression ("Neutre"), la non-connaissance ("Je ne sais pas") ainsi que de considérer qu'il s'agit d'une expression qui n'est pas dans la liste ("Autre"). Ceci est une hypothèse de départ. Je ne prends en compte dans mon analyse que les réponses qui correspondent aux 6 alternatives ainsi qu'à l'expression neutre. Je ne gère pas ici les réponses "Autre" et "Je ne sais pas".

Je ne tiens par ailleurs pas compte des caractéristiques socio-économiques des preneurs de décision. Les fonction d'utilité sont donc uniquement décrites par les attributs des alternatives et la partie aléatoire, afin de simplifier le modèle.

Les attributs des alternatives se fondent sur le système FACS et ses *Action Units*.

L'utilisation du modèle actif d'apparence nous permet de détecter 55 points sur le visage et de définir à partir de ces points 26 attributs (largeur

des yeux par exemple), pour coller au mieux à la description des *Action Units*.

## 2.2 Le sondage d'évaluation des expressions faciales

De manière générale, un système automatique de reconnaissance d'expressions faciales se fonde sur une représentation des expressions apprise à partir d'un ensemble d'entraînement sur lequel l'expression faciale est connue. Le processus d'apprentissage s'appuie sur les annotations associées à chaque image par un expert ou un groupe d'expert. Les experts doivent associer chaque image de l'ensemble d'entraînement à une expression faciale. Ainsi, plus les experts seront fiables, plus la connaissance sur laquelle le modèle se fondera sera robuste.

Afin de pouvoir utiliser les modèles de choix discret et Biogeme [3], nous avons en effet besoin d'une base de donnée d'images annotées.

C'est dans l'objectif d'avoir une base de donnée la plus robuste possible et de construire une connaissance du sens commun qu'un sondage a été mis sur pied par M. Sorci et al. [32]. Le sondage nous permet d'obtenir une base de données créée par une population hétérogène, ayant des caractéristiques socio-économiques variées, provenant de partout dans le monde, de différentes cultures, de différents âges, etc. Ces données nous permettent non seulement d'avoir une base de données d'images annotées, mais permettraient aussi à terme de définir si les informations socio-économiques des preneurs de décision - les sondés - interviennent dans la reconnaissance des expressions faciales, bien que je ne les ai pas insérées dans mon modèle.

Les images utilisées dans le sondage proviennent de la base de données Cohn-Kanade [20]. La base de données consiste en séquences d'expression de sujets, partant d'une expression neutre et modifiant leur visage jusqu'à une certaine expression. Les sujets sont des étudiants en psychologie de 18 à 30 ans et ont dû effectuer une série de 23 expressions faciales. Il y a 104 sujets, mais seuls 10 ont donné leur accord pour la publication. Le sous-ensemble de la base de donnée Cohn-Kanade utilisé pour le sondage contient 1274 images de ces 10 sujets, 8 femmes et 2 hommes.

Le processus d'annotation consiste à associer une expression à chaque image. Le sondage se fait sur internet<sup>2</sup>. Chaque participant crée un compte, qui lui permet de donner certaines informations pratiques, telle la langue, ainsi que des données socio-économiques, telles que le groupe ethnique par exemple. L'anonymat reste cependant garanti, ainsi que la possibilité de ne pas répondre à ces questions.

---

<sup>2</sup><http://lts5www.epfl.ch/face/>



**Create a new user**

Birth Year :	<input type="text" value="0000"/>
Gender :	<input checked="" type="radio"/> Male <input type="radio"/> Female
Language :	<input type="text" value="English"/>
Studies :	<input type="text" value="High School"/>
Science Knowledge :	<input type="text" value="None"/>
Ethnic group :	<input type="text" value="None"/>
Current location :	<input type="text" value="None"/>
Occupational category :	<input type="text" value="None"/>
Username :	<input type="text"/>
Password :	<input type="text"/>
Password Confirmation :	<input type="text"/>
<input type="button" value="Ok"/>	

FIG. 2.1 – Formulaire socio-économique

Jusqu'à présent, 1488 participants ont pris part au sondage, pour un total de plus de 33 800 annotations d'image.



FIG. 2.2 – Interface d’annotation d’images

## Chapitre 3

# Analyse du modèle

L'idée de ce projet a été de m'inspirer de deux articles parus récemment.

*Discrete choice models for statistic facial expression recognition* (G. Antonini, M. Sorci, M. Bierlaire & J.-P. Thiran, 2006 [1]) présente l'utilisation des modèles de choix discrets dans la reconnaissance des expressions faciales. Il utilise l'*Active Appearance Model* pour l'extraction des caractéristiques faciales, à partir duquel on obtient un masque de points déjà présenté plus haut (fig. 1.1 p. 7, en grand format p. 22). A partir de ce masque, les auteurs définissent une série de caractéristiques faciales se fondant sur des ratios des longueurs et des largeurs des yeux, du nez, des sourcils et de la bouche. Ils utilisent finalement les modèles de choix discrets pour la classification.

Dans *Active and dynamic information fusion for facial expression understanding from image sequences* [41], les auteurs utilisent les *Action Units* du système FACS comme caractéristiques faciales, pour ensuite classifier des séquences d'images en utilisant des réseaux dynamique bayesiens et les hypothèses du système FACS dans la description des expressions faciales.

Je me propose ici de reprendre la base de données d'images statiques et l'extraction des caractéristiques faciales par l'*Active Appearance Model* de l'article de G. Antonini et al. [1], mais d'utiliser les *Action Units* du système FACS comme caractéristiques faciales, en m'appuyant sur l'article de Zhang et al. [41], pour finalement construire un modèle de perception des expressions faciales statiques à l'aide des modèles de choix discret, utilisable pour la reconnaissance de ces expressions.

Dans un premier temps, il a fallu exprimer les *Action Units* du système FACS à partir des 55 points du visage définis par le modèle actif d'apparence.

Ensuite, j'ai rédigé les fonctions d'utilité selon le modèle FACS.

Les données du sondage ont été mises en forme, afin de contenir toute les informations et en particulier les valeurs des *Action Units* désormais définies en fonction des points du visage.

Finalement, ces données ainsi que les fonctions d'utilité ont été testée à l'aide de Biogeme [3].

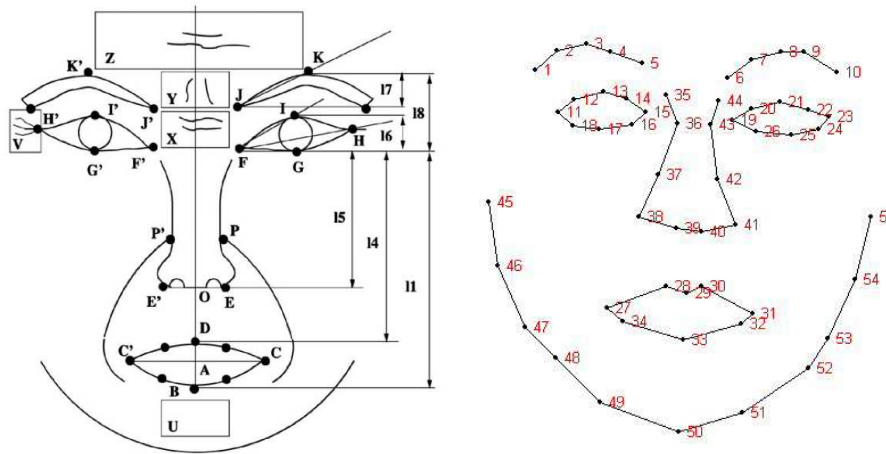


FIG. 3.1 – Les deux masques. A gauche celui de Zhang et al. [41] et à droite celui de Antonini et al. [1]

### 3.1 Traduction du modèle FACS

En s'appuyant sur le système FACS et l'article de Zhang et al. [41], 26 attributs différents ont été définis.

Selon le modèle FACS, les expressions faciales dépendent d'un certains nombres d'*Action Units* définies par l'accroissement ou la diminution de certaines longueur et de *transient features*, des particularités "transitoire" ou "éphémères", tels que rides ou plis de la peau. Je tiens compte ici uniquement des variations de certaines longueur et de certains angles sur le visage, et pas des rides et des plis, n'ayant à ma disposition que le masque du modèle actif d'apparence.

Il s'agit donc ici de traduire les *Action Units* définies à partir du masque de Zhang et al. [41] dans le langage du masque de Antonini et al. [1].

Pour cela, j'ai à disposition les deux masques, ainsi que le descriptif des *Action Units* selon le premier masque, reproduit dans la fig. 3.2, p. 18.

J'ai donc identifié les caractéristiques sur le masque à parti duquel elles avaient été définies, puis j'ai essayé de les exprimer le plus fidèlement dans le masque du modèle actif d'apparence.

Prenons comme exemple l'*Action Unit* 4. Sa définition, selon le modèle de Zhang est :

$l8, \overline{FJ}, \overline{JJ'}, \overline{FP}, \overline{F'P'}$   
 decreased,  $\angle HFI$  increased  
 and wrinkle in  $\square Y$  increased

On obtient ainsi une description de l'*Action Unit* 4 sur le masque de

AUs	Facial Visual Cues	Visual Channel(s)
AU1	$\angle F H J, \overline{J F}$ increased OR $\overline{J F}$ increased, $l 8$ nonincreased	Brow
AU2	$l 8$ increased and $\overline{J F}$ nonincreased frown in $\square Z$ increased	Brow, Wrinkler
AU4	$l 8, \overline{F J}, \overline{J J'}, \overline{F P}, \overline{F' P'}$ decreased, $\angle H F I$ increased and wrinkle in $\square Y$ increased	Brow, Wrinkler
AU5	$l 6, \overline{J F}$ and $\overline{J J'}$ increased	Lid
AU6	nasolabial furrow presence and wrinkle in $\square V$ increased	Nasolabial, Wrinkler
AU7	$\angle H F I$ nonincreased and $\angle H G F$ increased	Lid
AU9	wrinkle increased in $\square X$ nasolabial furrow presence OR $\overline{P F}, \overline{F J}$ decreased	Wrinkler, Nasolabial
AU10	$l 4$ decreased and $ \overline{F C} - \overline{F' C'} $ increased, nasolabial presence OR $\overline{O D}$ decreased, $\overline{D B}, \overline{C' C}$ increased	Lip, Nasolabial
AU12	$\overline{F C}, \overline{F' C'}$ decreased, $\overline{C C'}$ increased, $\overline{G I}$ nonincreased	LipCorner
AU15	$\overline{F C}, \overline{F' C'}, \overline{C C'}$ increased	LipCorner
AU16	$\overline{O D}$ non-change, $\overline{D B}$ decreased	Lip
AU17	$\overline{O B}$ decreased and wrinkle in $\square U$ presence	Chin, Wrinkler
AU20	$\overline{C C'}$ increased and $\overline{F C}, \overline{F' C'}$ nonchange	LipCorner
AU23	$\overline{D B}, \overline{C C'}$ decreased	Lip
AU24	$\overline{D B}$ decreased, $\overline{C C'}$ nonchange	Lip
AU25	$\overline{D B}$ increased, $\overline{D B} < T_1$ , $\overline{C C'}$ nonincreased	Mouth
AU26	$T_1 < \overline{D B} < T_2$ , $\overline{C C'}$ nonincreased	Mouth
AU27	$\overline{D B} > T_2$ , $\overline{C C'}$ nonincreased	Mouth

FIG. 3.2 – Descriptif des *Action Units* selon le masque du système FACS [41]

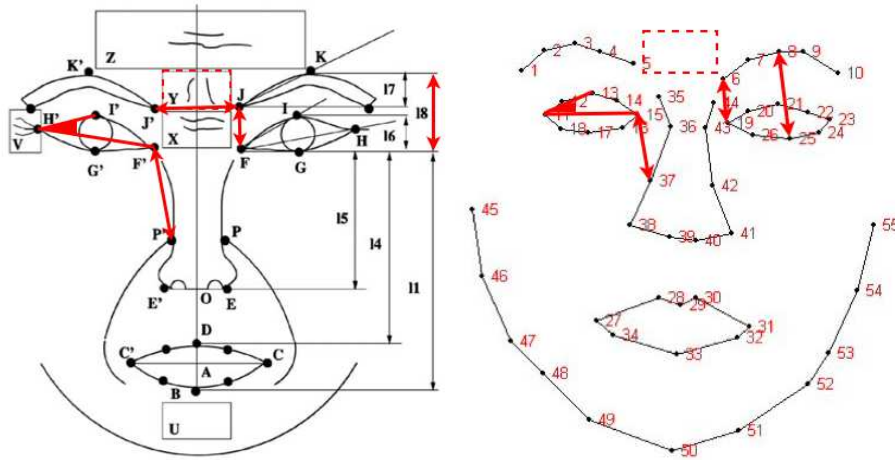


FIG. 3.3 – L'Action Unit 4 sur les deux masques

Zhang et al., que l'on peut approximer sur le masque de Antonini et al. Comme dans toute traduction, la transcription parfaite n'est pas possible. Certaines informations se perdent dans la transcription. Il en est ainsi en l'occurrence pour toutes les *transient features*, ces caractéristiques telles que les rides et les plis de la peau. Dans notre exemple, les rides présentes dans  $\square Y$  ne sont pas exprimables dans le modèle de Antonini et al.

A partir de ces informations, j'ai défini une série de descripteurs faciaux (*facial descriptors*, fd) représentant chaque longueur et chaque angle utilisé dans la description des *Action Units*. Pour notre exemple, dans le cas de l'Action Unit 4, on obtient donc trois descripteurs faciaux pour les longueurs  $l8$ ,  $\overline{FJ}$ ,  $\overline{FP}$ , présent à gauche et à droite du visage, ainsi qu'un angle,  $\angle HFI$ , lui aussi présent des deux côtés du visage. On a donc 8 descripteurs faciaux pour cette Action Unit. Si l'on prend le descripteur facial  $l8$ , il correspond dans le masque du modèle actif d'apparence à la distance entre les points 8 et 25. On peut donc décrire de descripteur facial comme suit dans Matlab :

```
fd.broweye_r(i) = sqrt((imageLand(8,3) -
imageLand(25,3))^2 + (imageLand(8,4) -
imageLand(25,4))^2);
```

où  $fd.broweye\_r$  représente la distance entre les points 8 et 25, et  $imageLand$  est la référence aux coordonnées des points des visages (la deuxième composante, 3 ou 4, représente respectivement l'abscisse ou l'ordonnée du point du visage considéré).

A partir de ces descripteurs faciaux, on peut donc définir l'Action Unit 4 selon le masque du modèle actif d'apparence :

## AU4

```
fd.broweye_l(i), fd.broweye_r(i),
fd.brow_dist(i), fd.eye_angle_l(i),
fd.eye_angle_r(i), fd.eye_nose_dist_l(i),
fd.eye_nose_dist_r(i)
```

On peut par conséquent traduire complètement le tableau représentant la description des expressions faciales selon le modèle FACS (fig. 3.4, ci-dessous).

Le résultat complet de cette transcription est présente en annexe à ce rapport.

Comme précisé plus haut, le résultat ne tiens bien entendu pas compte des *transient features*, des rides et des plis du visage.

Emotional Category	Primary Visual Cues					Auxiliary Visual Cues					
	AU	AU	AU	AU	AU	AU	AU	AU	AU	AU	Transient Feature(s)
Happiness	6	12				25	26	16			wrinkles on outer eye canthi presence of nasolabial furrow
Sadness	1	15	17			4	7	25	26		
Disgust	9	10				17	25	26			presence of nasolabial furrow
Surprise	5	26	27	1+2							furrows on the forehead
Anger	2	4	7	23	24	17	25	26	16		vertical furrows between brows
Fear	20	1+5	5+7			4	5	7	25	26	

FIG. 3.4 – Utilisation des *Action Units* dans la description des expressions faciales [41]

## 3.2 Définition des fonctions d'utilité

Il faut désormais construire des fonctions d'utilité pour chaque expression faciale. Pour cela, on peut s'appuyer sur le tableau de la fig. 3.4 ci-dessus. Il ne s'agit pas dans l'article de Zhang et al. de fonction d'utilité, mais simplement d'intervention de certaines *Action Units* dans la description d'expressions faciales. Les *Action Units* sont classées selon une intervention dite "primaire" ou "secondaire".

Nous allons par conséquent définir des fonctions d'utilité se fondant sur ce modèle, à la différence près que nous ne tiendrons pas compte des *transient features*, que nous ne prendrons pas en compte *a priori* les remarques dynamiques (croissant, décroissant) des descriptions des *Action Unit* et que nous ne gérerons pas explicitement les *Action Units*, mais que nous définirons

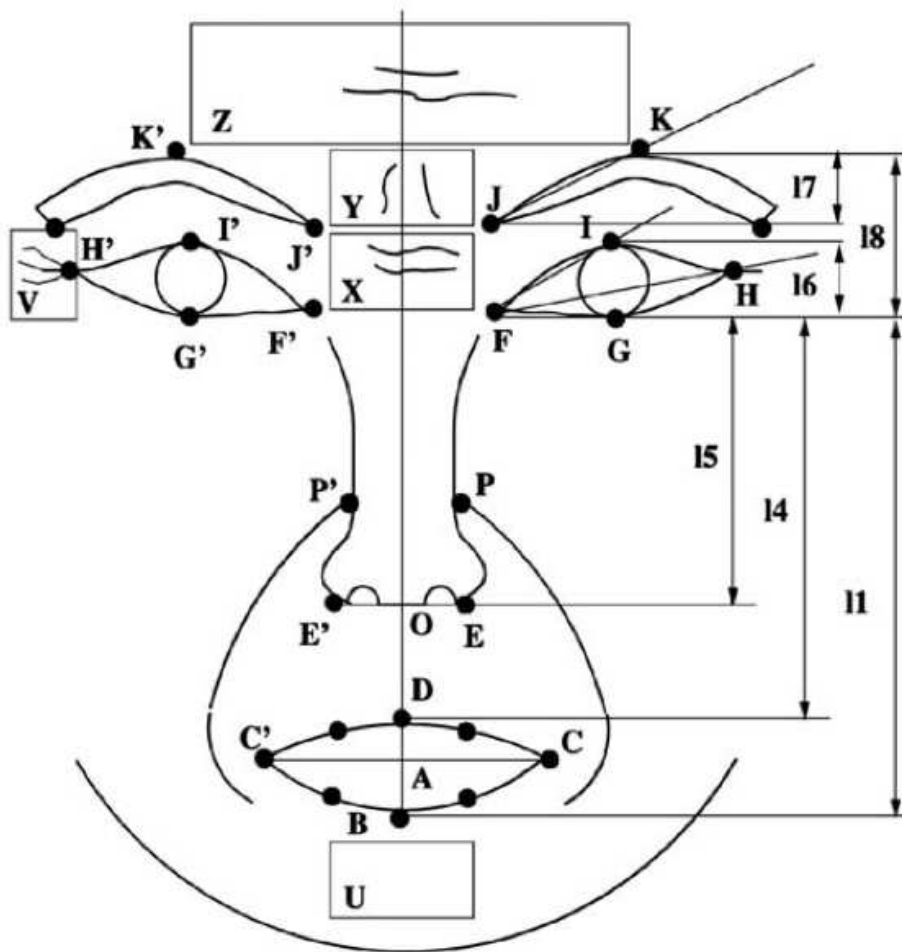


FIG. 3.5 – Masque de point utilisé par Zhang et al. [41]



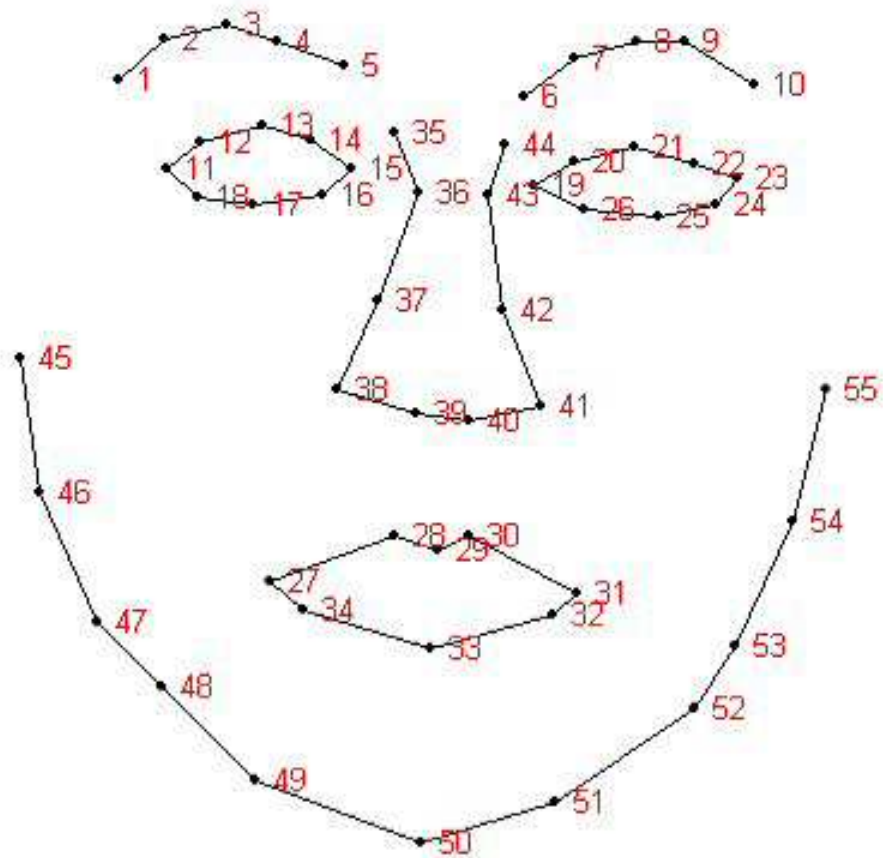


FIG. 3.6 – Masque de point produit par le modèle actif d'apparence (55 points) [1]

les fonctions d'utilités comme des combinaisons linéaires des descripteurs faciaux.

Il est effectivement impossible de gérer *a priori* les remarques dynamiques, car dans l'approche de Zhang et al., ils s'appuient sur des séquences d'images, alors que nous analysons des images statiques. Ils ont par conséquent la possibilité de se référer à une image précédente, alors que nous ne le pouvons pas. Implicitement, dans notre définition des fonctions d'utilités, nous nous référons à l'expression faciale neutre. Nous verrons dans l'analyse, par la suite, que nous pouvons *a posteriori* retrouver ces informations dynamiques.

Les fonctions d'utilités utilisées ont donc la structure suivante :

$$V_i = \alpha_i + \sum_{k=1}^K I_{ki} \beta_{ki} \text{fd}_k$$

où  $i$  est l'indice de l'alternative,  $K$  le nombre de descripteurs faciaux,  $\beta_{ki}$  le poids du  $k$ -ième descripteur facial dans l'alternative  $i$ ,  $\alpha_i$  la constante spécifique à l'alternative,  $I$  la fonction indicatrice exprimant l'utilisation ou non du descripteur facial dans la fonction d'utilité de telle ou telle expression faciale et  $\text{fd}_k$  les descripteurs faciaux.

Si nous voulions exprimer explicitement les *Action Unit* dans les fonctions d'utilités, nous devrions ajouter un paramètre particulier multipliant une combinaison linéaire de descripteurs faciaux. Dans ce cas, la linéarité du modèle serait perdue. Il est parfaitement possible de gérer des modèles non linéaires *via* Biogeme, mais je n'ai pas eu le temps de le faire.

Les fonctions d'utilité permettent d'exprimer la connaissance *a priori* du modèle. Ici, elles permettent d'exprimer la connaissance du système FACS. Dans la prochaine section, nous allons utiliser Biogeme [3] afin de tester ces hypothèses. Nous allons utiliser les données du sondage pour infirmer ou confirmer l'hypothèse générale selon laquelle la reconnaissance des expressions faciales suit le modèle FACS.

### 3.3 Analyse du modèle avec Biogeme

Biogeme (pour Bierlaire's Optimisation Toolbox for GEV Model Estimation) est un logiciel créé par Michel Bierlaire pour l'estimation du maximum de vraisemblance des modèles de choix discrets aux valeurs extrêmes généralisées (*Generalized Extreme Value*, GEV).

Il s'agit donc d'estimer un modèle, afin de déterminer s'il est valide.

J'ai testé deux modèles, un s'appuyant sur les caractéristiques principales du système FACS, et un sur les principales et les auxiliaires.

Dans le cas du modèle simple, en s'appuyant uniquement sur les caractéristiques principales, et absolument sans modifier le modèle FACS, nous arri-

Résumé des statistiques			
Nombre d'observations = 33147		$\rho^2 = 0.343683$	
L(0) = -64501.1		Nombre de paramètres estimés = 73	
L( $\beta$ ) = -42333.1		Norme finale du gradient = 0.000229707	
Likelihood ratio test : 44335.9			
Nom de la variable	Coefficient estimé	Erreur standard robuste	T statistique robuste
b_eye_nose_dist_r_D	-1.26e+02	+1.07e+01	-1.18e+01
b_brow_dist_A	+2.83e+01	+2.48e+00	+1.14e+01
b_broweye_r3_SD	+7.83e+01	+3.44e+00	+2.27e+01

FIG. 3.7 – Résumé des résultats pour le modèle s'appuyant sur les caractéristiques principales

vons déjà à des bons résultats (73 paramètres estimés, une log-vraisemblance finale de 44335.9, une norme finale du gradient à 0.0336163 et 10 paramètres considérés par Biogeme comme non significatifs seulement).

Avec le modèle un peu plus complexe, en tenant compte à la fois des caractéristiques principales et secondaires, on augmente les nombre de paramètres estimés à 101 tout en gardant d'excellents résultats.

Résumé des statistiques			
Nombre d'observations = 38286		$\rho^2 = 0.302309$	
L(0) = -84122.9		Nombre de paramètres estimés = 101	
L( $\beta$ ) = -58691.8		Norme finale du gradient = 9.8044e-006	
Likelihood ratio test : 50862.3			
Nom de la variable	Coefficient estimé	Erreur standard robuste	T statistique robuste
b_mouth_w_H	+6.95e+01	+1.73e+00	+4.01e+01
b_mouth_h_SU	+7.02e+01	+1.89e+00	+3.71e+01
b_eye_nose_dist_r_SA	-2.85e+02	+1.01e+01	-2.81e+01

FIG. 3.8 – Résumé des résultats pour le modèle s'appuyant sur les caractéristiques principales et secondaires

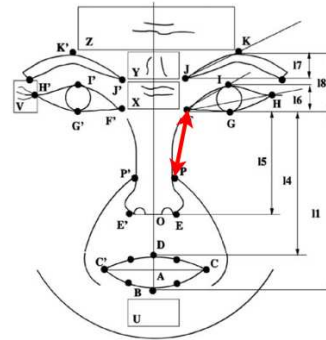
Dans les deux cas, le modèle est bien défini et la norme du gradient est suffisamment faible.

Analysons plus en détail un paramètre particulier.

Nom de la variable	Coefficient estimé	Erreur standard robuste	T statistique robuste
b_eye_nose_dist_r_D	-1.26e+02	+1.07e+01	-1.18e+01

Considérons donc le paramètre lié à la distance entre l'oeil et le nez à droite intervenant dans la fonction d'utilité du dégoût. Ce descripteur facial intervient dans la définition du dégoût par l'Action Unit 9. Il s'agit de la distance  $\overline{FP}$  sur le masque de Zhang et al.. Le coefficient associé, b\_eye\_nose\_dist\_r\_D, est significativement différent de zéro et vaut -126.

AU9 | wrinkle increased in □X  
nasolabial furrow presence OR  
 $\overline{PF}$ ,  $\overline{FJ}$  decreased



Par conséquent, selon notre modèle, plus la distance entre le nez et les yeux ainsi définie est petite, plus l'utilité associée au dégoût sera grande. En effet, si la distance est petite et qu'elle est multiplié par -126, l'utilité sera plus grande que dans la situation d'un visage exprimant l'expression neutre.

Dans l'Action Unit 9, il est de plus précisé que la distance  $\overline{FP}$  décroît. On voit que cette remarque dynamique, que nous n'avions pas pu prendre

en compte dans la définition des fonctions d'utilité, se retrouve dans le signe du coefficient associé. On tient donc compte de ces remarques dynamiques. Il faut cependant noter qu'il y a une différence notable entre les deux modèles dans le sens donné à ces remarques. En effet, dans l'article de Zhang et al., ces remarques font référence à l'image précédente, à l'état du visage précédent dans la séquence d'image considérée. Dans l'article de Antonini et al., il s'agit d'une croissance ou d'une décroissance faisant référence à l'état neutre, puisque nous considérons des images statiques.

L'association d'une centaine de ces descripteurs faciaux nous donne un modèle de perception des expressions faciales fondé sur le système FACS et confirmé par les données du sondage.

# Conclusion

Dans ce travail, j'ai utilisé les modèles de choix discrets afin de modéliser la perception humaine des expressions faciales. Pour cela, j'ai traduit le système FACS, et j'ai analysé ce modèle avec Biogeme. Il résulte de cette analyse une confirmation du modèle, ainsi qu'une expression quantitative du modèle FACS, principalement qualitatif dans sa définition originale.

La force de ce modèle réside dans le fait qu'il s'appuie à la fois sur une connaissance large et robuste à travers l'échantillon d'images d'apprentissage issues du sondage et sur le modèle FACS, à travers les fonctions d'utilités, qui lui donne un sens. On peut ainsi rendre significatives les données brutes obtenues par le sondage.

Parmi les améliorations potentielles du modèle et les avancées futures, on peut imaginer tenir compte des caractéristiques socio-économiques dans les fonctions d'utilités. Les données sont déjà disponibles dans le résultat du sondage, et il ne reste plus qu'à tester et analyser ces données. On peut aussi imaginer prendre en compte les *transient features*. Pour cela, il faudrait utiliser un autre masque et être capable d'extraire ces données à partir des images statiques. Finalement, il serait intéressant de modéliser explicitement les *Action Units* comme de réelles unités. Cela implique d'utiliser des modèles non-linéaires, et pourrait poser problème puisque les *Action Units* n'apparaissent souvent que dans la description d'une seule expression faciale.

En changeant radicalement le cadre de réflexion, on peut aussi penser appliquer cette technique sur des séquences d'images et non plus sur des images statiques.

A partir de ce modèle, on peut facilement imaginer valider la reconnaissance des expressions faciales.

Ce travail a été effectué comme projet de semestre lors de ma troisième année de Bachelor en mathématique. Il a constitué le premier travail de recherche dans mon cursus. Le bilan est excellent. J'ai pu avoir un aperçu de la recherche, collaborer avec Matteo Sorci, doctorant au laboratoire de traitement du signal, et Barbara Cerretani, en projet de master dans le même laboratoire, utiliser les connaissances acquises au cours de Modélisation mathématique du comportement du Prof. Bierlaire que je suivais en parallèle

et faire un projet à la frontière entre mathématiques et ingénierie.

Le sujet abordé est par ailleurs passionnant, autant du point de vue des concepts mathématiques utilisés - les modèles de choix discret - que de l'application - les expressions faciales.

Je regrette cependant de ne pas avoir eu plus de temps à disposition afin de mener à bien quelques analyses plus approfondies.

# Bibliographie

- [1] G. Antonini, M. Sorci, M. Bierlaire, and J. P. Thiran. Discrete choice models for static facial expression recognition. In *Advanced Concepts for Intelligent Vision Systems*, pages 710–721, 2006.
- [2] M. Bartlett. *Face image analysis by unsupervised learning and redundancy reduction*. Springer, 1998.
- [3] M. Bierlaire. Biogeme : A free package for the estimation of discrete choice models. In *Proceedings of the 3rd Swiss Transportation Research Conference*, Ascona, Switzerland, 2003.
- [4] J. D. Boucher and P. Ekman. Facial areas and emotional information. *Journal of Communication*, 25 :21–29, 1975.
- [5] R. Chellappa, C. L. Wilson, and S. Sirohey. Human and machine recognition of faces : A survey. *Proceedings of the IEEE*, 83(5) :705–741, Mai 1995.
- [6] G. Collier. *Emotional Expression*. Hillsdale, NJ : Lawrence Erlbaum Publisher, 1985.
- [7] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. *Lecture Notes in Computer Science*, 1407 :484, 1998.
- [8] S. Coren and J. A. Russell. The relative dominance of different facial expressions of emotion under conditions of perceptual ambiguity. *Cognition and Emotion*, 6 :339–356, 1992.
- [9] C. Darwin. *The Expression of Emotion in Man and Animals*. 1872.
- [10] P. Ekman. Universals and cultural differences in facial expressions of emotions. Paper presented at the Nebraska Symposium of Motivation, Lincoln, Nebraska, 1972.
- [11] P. Ekman. Asymmetry in facial expression. *Science*, 209 :833–834, 1980.
- [12] P. Ekman, W. V. Friesen, and P. Ellsworth. Research foundations. In *P. Ekman (Ed.), Emotion in the human face*, pages 1–43. Cambridge : Cambridge University Press, 1982.



- [13] P. Ekman and W.V. Friesen. *Facial Action Coding System (FACS) : Manual*. Consulting Psychologists Press, Palo Alto, California, 1978.
- [14] P. Ekman and M. O’Sullivan. Facial expression : Methods, means and moues. In R. S. Feldman and B. Rimé, editors, *Fundamentals of non-verbal behavior*. New-York : Cambridge University Press, 1991.
- [15] I. A. Essa and A. P. Pentland. Coding, analysis, interpretation, and recognition of facial expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7) :757–763, Juillet 1997.
- [16] N. L. Etcoff and J. J. Magee. Categorical preception of facial expressions. *Cognition*, 44 :227–240, 1992.
- [17] J. M. Fernández-Dols. Facial expression and emotion : A situationist view. In *The Social Context of Nonverbal Behavior*, volume 1, pages 242–261. Cambridge University Press, 1999.
- [18] A. J. Fridlund, P. Ekman, and H. Oster. Facial expression of emotion : Review of the litterature, 1970-1983. In A. W. Siegman and S. Feldstein, editors, *Nonverbal behavior communication*. Hillsdale, NJ : Lawrence Erlbaum Associates, 1987.
- [19] U. Hess, A. Kappas, and K. Scherer. Multichannel communication of emotion : Synthetic signal production. In K. R. Scherer, editor, *Facets of emotion : Recent research*, pages 161–182. Hillsdale, NJ : Lawrence Erlbaum Publisher, 1988.
- [20] T. Kanade, J. Cohn, and Y. L. Tian. Comprehensive database for facial expression analysis. In *Proceedings of the 4th IEEE International Conference on Automatic Face and Gesture Recognition (FG’00)*, pages 46–53, 2000.
- [21] A. Kappas. The illusion of the neutral observer : On the communication of emotion. *Cahiers de linguistique française*, 12 :153–168, 1991.
- [22] A. Lanitis, C. J. Taylor, and T. F. Cootes. Automatic interpretation and coding of face images using flexible models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7) :743–756, 1997.
- [23] J. J. Lien. *Automatic recognition of facial expressions using hidden Markov models and estimation of expression intensity*. Pittsburgh, Pa. : Carnegie Mellon University, The Robotics Institute, 1998.
- [24] K. Mase. Recognition of facial expression from optical flow. *IEICE transactions*, 74(10) :3473–3483, 1991.
- [25] A. Mehrabian. Communication without words. *Psychology Today*, pages 53–55, September 1968.

- [26] P. Noller. Video primacy - a further look. *Journal of Nonverbal Behavior*, 9 :28–47, 1985.
- [27] C. Padgett and G. Cottrell. Identifying emotion in static images. In *Proceedings of the 2nd Joint Symposium on Neural Computation*, volume 5, pages 91–101, 1997.
- [28] C. Padgett and G. Cottrell. Representing face images for emotion classification. *Advances in Neural Information Processing Systems*, pages 894–900, 1997.
- [29] W. E. Rinn. Neuropsychology of facial expression. In *Fundamentals of Nonverbal Behavior*, pages 3–30. Cambridge University Press, 1991.
- [30] J. A. Russell. Is there universal recognition of emotion from facial expression? a review of cross-cultural studies. *Psychological Bulletin*, 115(1) :102–141, 1994.
- [31] C. A. Smith and H. S. Scott. A componential approach to the meaning of facial expressions. In *The psychology of facial expression*, pages 229–254. Cambridge University Press, 1997.
- [32] M. Sorci, G. Antonini, J.-P. Thiran, and M. Bierlaire. Facial expressions evaluation survey. Technical report, EPFL, Novembre 2006.
- [33] M. B. Stegmann. Active appearance models : Theory and cases. Master’s thesis, Department of Mathematical Modelling, Technical University of Denmark, Lyngby, 2000.
- [34] M. Suwa, N. Sugie, and K. Fujimora. A preliminary note on pattern recognition of human emotional expression. In *Proceedings of the 4th International Joint Conference on Pattern Recognition*, pages 408–410, 1978.
- [35] G. Traufetter and N. Bougeard. Le visage, reflet de l’âme. *L’Hebdo*, 3, Janvier 2007.
- [36] A. Tremblay, J. Deschênes, B. Poulin, J. Roy, G. Kirouac, and A. Kappas. Identification de régions faciales critiques dans le jugement d’authenticité d’un sourire d’expressions de joie véritablement ressentie. Papier présenté au XVIe Congrès annuel de la SQRP, Québec, Canada, Octobre 1993.
- [37] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1) :71–86, 1991.
- [38] P. Vayer and C. Roncin. *Le corps et les communications humaines : l’animation des groupes adultes*. Vigot, 1986.

- [39] Y. Yacob and L. Davis. Recognizing facial expressions by spatio-temporal analysis. In *Proceedings of the International Conference on Pattern Recognition*, volume 1, pages 747–749, 1994.
- [40] J. Ye, Y. Zhan, and S. Song. Facial expression features extraction based on gabor wavelet transformation. *IEEE International Conference on Systems, Man and Cybernetics*, 2004.
- [41] Y. Zhang and Q. Ji. Active and dynamic information fusion for facial expression understanding from image sequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(5) :699–714, May 2005.