Università degli Studi di Siena

Facoltà di Ingegneria

Corso di Laurea Magistrale in Ingegneria delle Telecomunicazioni

Modelling human perception of static expressions by discrete choice models

Relatore

Chiar.mo Prof. Alessandro Mecocci

Correlatore

Ing. Matteo Sorci

Tesi di Laurea di

Barbara Cerretani

Anno Accademico 2006/2007

Riassunto

Nelle comunicazioni interpersonali, il volto dei soggetti che interagiscono svolge un ruolo fondamentale, persino più importante della comunicazione verbale: costituisce infatti il mezzo per identificare chi ci sta davanti, per capirne lo stato emotivo e le intenzioni, per interpretare il significato di quello che ci viene detto, etc. Negli ultimi anni, grazie anche allo sviluppo delle tecnologie nel settore dell'analisi delle immagini, molte ricerche sono state portate avanti per arrivare ad un'automazione del processo di riconoscimento e di classificazione delle espressioni facciali. L'approccio generale che viene seguito nell'Analisi Automatica delle Espressioni Facciali (AFEA) è sostanzialmente unico, e prevede tre passi:

1. Ricerca del volto nell'immagine analizzata;
2. Estrazione di informazioni relative all'espressione facciale;
3. Riconoscimento dell'espressione.

Tuttavia, gli algoritmi di AFEA proposti sono molti, e differiscono principalmente nelle tecniche utilizzate per i vari passi.

Il metodo che qui proponiamo prevede, per i passi 1 e 2, la scelta degli *Active Appearance Models* (AAM), che si sono dimostrati efficaci per la rappresentazione di volti e l'estrazione di features. Tale tecnica, infatti, permette di ottenere modelli statistici sia della forma che della texture del volto, offrendone quindi una descrizione globale. I modelli vengono costruiti facendo apprendimento su un training set, costituito da

immagini in cui sono stati segnati alcuni punti di riferimento (*landmarks*) intorno ai principali tratti del viso umano (naso, occhi, bocca, mento). Questi punti verranno successivamente utilizzati nell'analisi per ottenere misure del volto utili all'identificazione delle espressioni facciali.

Una volta che i volti siano stati correttamente rappresentati, ci proponiamo di dimostrare la validità e l'efficienza dell'utilizzo di *Discrete Choice Models* (DCM) per associare ad una immagine di un volto la particolare espressione che sembra rappresentare (passo 3). I DCM sono modelli econometrici, introdotti solo recentemente nell'ambito dell'analisi di immagini, definiti per descrivere e predire il comportamento di individui (*decisori*) in situazioni di scelta, in cui l'insieme delle alternative disponibili (*choice set*) è finito e discreto. Generalmente, si ipotizza che il decisore scelga l'alternativa per lui più vantaggiosa. Quindi, se supponiamo che ogni alternativa i, $i = 1, ..., J$, porti al decisore n un certo profitto U_{ni} $\forall i$ (*utilità*), l'alternativa per lui più vantaggiosa sarà quella (j) con il livello di utilità più alto: $U_{nj} > U_{ni}$, $\forall i \neq j$. Il grado di appetibilità viene stabilito dal decisore valutando alcune caratteristiche delle alternative disponibili; inoltre, la scelta dipende da fattori personali del decisore stesso, come l'età, il lavoro che svolge, etc. Il ricercatore che vuole analizzare il comportamento delle persone che devono compiere la scelta, può fare delle ipotesi su quali siano le caratteristiche che la influenzano, e quindi quali siano gli attributi da usare per ricostruire l'utilità percepita dal decisore. Tuttavia, il ricercatore riesce ad osservare solo una parte dei fattori valutati dal decisore e, di conseguenza, è in grado di modellare solo una parte, V_{ni}, dell'utilità reale U_{ni}. La parte rimanente, non osservata, viene modellata utilizzando un termine random ε_{ni}, che tiene conto di tutti i fattori che influenzano la scelta ma che non sono stati inclusi in V_{ni}. L'utilità può quindi essere scritta come:

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

Il problema che stiamo studiando si presenta appunto come una situazione di scelta, in cui il choice set, discreto e finito, è costituito da nove espressioni facciali: Neutra, Gioia, Tristezza, Rabbia, Sorpresa, Disgusto, Paura, Altra e Non saprei. In questo contesto, dietro la scelta dei DCM c'è l'esigenza di modellare il modo in cui osservatori umani riconoscono su un volto una espressione. Vogliamo dimostrare che la scelta di una particolare espressione facciale non è influenzata solamente dalle caratteristiche del volto analizzato, ma anche da quelle socio-economiche della persona che lo analizza.

I dati su cui è basato il nostro studio sono stati ottenuti proponendo un questionario sul web a soggetti provenienti da tutto il mondo. Un approccio tipico per il riconoscimento di espressioni facciali è basato su una rappresentazione dell'espressione, appresa da un training set. Nel processo di apprendimento, a un esperto (o a un gruppo di esperti) viene chiesto di associare etichette (*labels*) ai campioni di training. In questo caso, la label che l'esperto deve associare ad ogni immagine dell'insieme di training è una espressione facciale, scelta tra nove possibili. Per noi, un "esperto" è qualcuno che ha una *forte conoscenza* del problema, in modo da garantire la correttezza di quello che stiamo cercando di apprendere. Nell'ambito della valutazione di espressioni facciali, ogni essere umano può essere considerato un esperto. Per questa ragione è stato proposto un sondaggio nel web, al fine di ottenere un insieme di immagini etichettate direttamente da esperti quanto più possibile eterogenei.

Il *Facial Expressions Evaluation Survey* è nato nell'agosto 2006 con lo scopo di acquisire in maniera diretta la conoscenza degli esperti sul problema del riconoscimento di espressioni facciali. Lo scopo finale del questionario è collezionare un insieme di dati creati da una popolazione di osservatori umani reali, provenienti da tutto il mondo, con impiego, cultura, età e genere diversi, appartenenti a gruppi etnici differenti. Questa eterogeneità nel campione di esperti ci permetterà di analizzare quali

fattori influenzano il modo in cui sono percepite le espressioni facciali. Allo stesso tempo, saremo in grado di capire quali tratti del volto umano le diverse categorie di persone considerano essere più importanti per riconoscere le espressioni. Infine, l'analisi dei dati fornirà informazioni utili per applicazioni basate sull'interazione tra uomo e computer. Infatti, ogni modello a priori costruito sui dati può essere utilizzato per migliorare il design di un sistema automatico di riconoscimento delle espressioni facciali.

Figura 1: *Esempi di soggetti del Cohn-Kanade database.*

Le immagini proposte ai partecipanti durante il questionario fanno parte del *Cohn-Kanade AU-Coded Facial Expression Database* [24], in cui sono raccolte sequenze che mostrano l'evoluzione temporale di espressioni facciali, partendo da un'espressione neutra e terminando nel momento di picco dell'espressione che viene mimata. Le sequenze sono registrate con una videocamera posta di fronte ai soggetti, che sono stati istruiti a riprodurre 23 espressioni facciali, azionando uno o più muscoli del volto secondo quanto spiegato da Ekman nel suo *Facial Action Coding System* (FACS). Di queste 23 espressioni, sei sono basate sulla descrizione di "emozioni principali" (gioia, rabbia, paura, disgusto, tristezza e sorpresa). Nelle immagini sono ritratti circa 100 studenti di psicologia, di età compresa tra 18 e 30 anni. Il 65 per cento sono donne, il 15 per cento neri-americani e il tre per cento asiatici o latini. Di questi, solamente dieci hanno dato il consenso per la pubblicazione delle fotografie.

Le immagini utilizzate per il nostro sondaggio sono state quindi scelte tra quelle in cui sono ritratti questi soggetti (otto donne e due uomini). In particolare, abbiamo utilizzato 1274 immagini estratte dalle sequenze relative a espressioni di "emozioni principali".

Il questionario, che si trova alla pagina web

`http://lts5www.epfl.ch/face`

è disponibile in tre lingue (inglese, italiano e francese) e i partecipanti possono scegliere quella che preferiscono. All'inizio del sondaggio, e solamente la prima volta che viene effettuato il login, il partecipante deve creare un nuovo account e inserire alcune informazioni personali (Figura 2). Queste informazioni, che chiameremo *caratteristiche socio-*

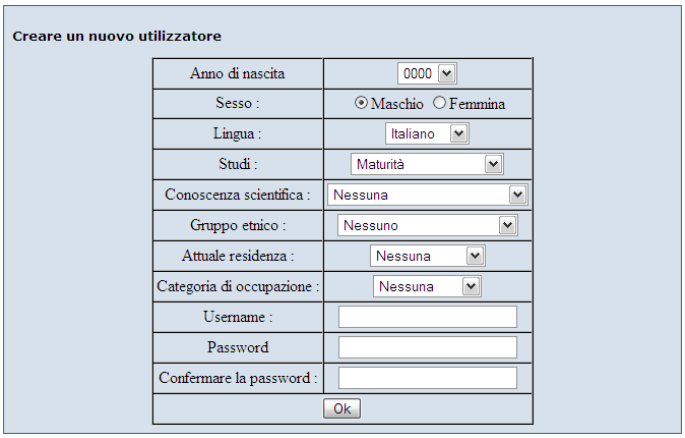


Figura 2: *Interfaccia per l'inserimento delle informazioni personali del partecipante.*

economiche, sono importanti per poter segmentare la popolazione dei decisori rispetto al loro background culturale, all'età, all'occupazione e al titolo di studio. Il gruppo etnico è utile per poter investigare il comportamento di scelta delle persone quando si trovano davanti soggetti della stessa razza o di razze diverse. La lista completa e la descrizione delle caratteristiche socio-economiche analizzate è riportata nella Tabella 5.1. L'utente può garantire la propria privacy scegliendo liberamente

un username e una password personali. I dati sono trattati in maniera confidenziale e solamente a scopo scientifico; tuttavia nella maggior parte dei campi da riempire è disponibile un'opzione "Nessuno" per coloro che non vogliono fornire informazioni personali. Dopo aver effettuato l'accesso nel suo account, il partecipante deve specificare il luogo in cui si trova in quel momento (casa, lavoro o altro) e decidere quante immagini vuole etichettare. A questo punto, il partecipante può iniziare il processo di annotazione, cliccando sul pulsante "Iniziare il questionario".

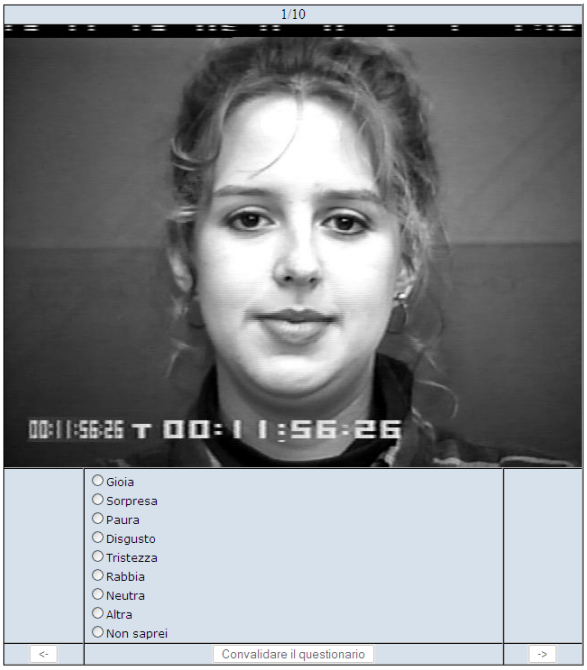


Figura 3: *Interfaccia per l'annotazione delle immagini.*

Il processo di annotazione consiste nell'associare un'etichetta (una espressione) ad ogni immagine che viene mostrata. L'etichetta deve essere scelta in un insieme di nove alternative: Gioia, Sorpresa, Paura, Disgusto, Tristezza, Neutra, Altra, Non saprei. Nella lista delle espressioni disponibili abbiamo incluso, oltre alle sette espressioni di base, le opzioni "Altra" e "Non saprei", per tenere conto di eventuali ambiguità nell'interpretazione delle immagini. Per rendere più semplice il processo di annotazione, è stata disegnata un'interfaccia grafica semplice e in-

tuitiva, mostrata in Figura 3. Dopo aver selezionato una delle opzioni disponibili, il partecipante può cliccare sulla freccia destra per convalidare la scelta corrente e passare all'immagine successiva. Il questionario può essere interrotto in qualsiasi momento, uscendo dal proprio account, e ricominciato alla connessione successiva a partire dalla prima immagine non ancora etichettata. Quando il partecipante ha raggiunto il numero di fotografie che aveva deciso di etichettare, può convalidare l'intero sondaggio cliccando sul pulsante "Convalidare il questionario". E' possibile partecipare al questionario anche più di una volta.

Fino ad ora, 1700 persone hanno risposto al questionario, annotando in totale circa 39000 immagini. Alcune statistiche sui partecipanti, disponibili anche alla pagina web

`http://itswww.epfl.ch/~sorci/SurveyStat.php`

sono mostrate in Figg. 5.4-5.5. Possiamo osservare che la maggioranza dei partecipanti vive in Europa e che il gruppo etnico "Bianco" è il più numeroso. Sono comunque presenti rappresentanti di tutti i continenti popolati e di tutti i gruppi etnici. Se prendiamo in considerazione il background culturale, circa la metà del campione ha una laurea universitaria e tutte le categorie di occupazione sono abbastanza ben rappresentate. Le categorie "Scienze dell'informazione" e "Altre" sono le più numerose per quanto riguarda la "Conoscenza scientifica", ma è presente anche un buon numero di partecipanti con conoscenze nelle scienze sociali, cognitive e del comportamento. Possiamo concludere che il campione di popolazione analizzato è piuttosto eterogeneo, dato che tutte le categorie sono ben rappresentate. L'informazione raccolta è quindi sufficiente a stabilire quali tra i fattori presi in esame intervengano nella percezione delle espressioni facciali.

Il nostro approccio alla comprensione delle espressioni facciali si basa su descrizioni linguistiche delle espressioni stesse (le etichette). Voglia-

mo analizzare la procedura di scelta delle etichette, per identificare quali fattori vengono presi in esame dagli individui quando associano un'espressione ad un volto. Facciamo l'ipotesi che la scelta dipenda da due tipi di fattori:

- Caratteristiche del volto esaminato,
- Caratteristiche personali del decisore.

Per dimostrare la nostra assunzione, proponiamo quattro DCM, seguendo un approccio progressivo, in cui ad ogni passo viene aggiunto un insieme di nuovi attributi al modello costruito al passo precedente. I primi tre modelli sono costruiti combinando tra loro solamente gli attributi relativi alle caratteristiche del volto, mentre nel quarto sono introdotte anche le caratteristiche del decisore. Per ciascuna delle nove espressioni considerate è definita una funzione di utilità, lineare rispetto ai parametri. La scelta di una forma lineare è legata principalmente alla necessità di costruire modelli non troppo complessi, in modo da non dover utilizzare per la stima risorse di calcolo eccessive, anche nel caso in cui i parametri incogniti siano molti. La forma generale delle utilità è data da:

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

dove $i = 1, ..., C$ con $C = 9$ è il numero di espressioni, K è il numero di attributi inclusi nel modello, I_{ki} è una funzione di attivazione, uguale a 1 se il k-esimo attributo è incluso nell'utilità per l'espressione i e 0 altrimenti, e α_i è una costante di specificità dell'alternativa (ASC). I coefficienti α_i rappresentano il valore medio della parte non osservata dell'utilità corrispondente, e uno di loro deve essere normalizzato a 0, affinchè il modello sia consistente con la teoria dei DCM (vedi [5]). In questo caso, normalizziamo rispetto all'espressione neutra. Ogni attributo k in ogni utilità i è pesato da un coefficiente deterministico incognito, β_{ki}, che deve essere stimato. Per semplificare il modello, nella funzione

di utilità per l'alternativa "Non saprei" è stata inclusa solo la costante di specificità.

Nei paragrafi seguenti illustreremo in dettaglio i quattro modelli che proponiamo, dal più semplice al più complesso, spiegando, ad ogni passo, quali attributi sono stati inclusi e perché. Nelle funzioni di utilità scelte per ciascuno dei modelli sono stati inclusi solamente gli attributi corrispondenti a parametri statisticamente significativi (secondo una statistica t-test rispetto a zero). E' importante notare che le espressioni finali proposte per le utilità sono il risultato di un robusto processo iterativo, in cui ipotesi diverse sono state testate e validate. Prima di costruire e stimare i modelli, il dataset viene parzialmente ripulito mediante un'analisi degli outliers e, successivamente, viene diviso in due blocchi: l'80 per cento delle osservazioni sarà utilizzato nella fase di training, mentre il rimanente 20 per cento per la validazione dei modelli.

Le espressioni facciali rappresentano la conseguenza visibile dell'attività muscolare del volto. Tale attività può essere descritta e codificata utilizzando il *facial action coding system* (FACS) proposto da Ekman, in cui le azioni dei muscoli facciali vengono rappresentate mediante le cosiddette *Action Units* (AU). E' opinione generale che almeno le sei "espressioni principali" possano essere rappresentate come combinazioni di AU, perciò anche noi abbiamo deciso di utilizzare modelli basati sulle AU, per descrivere sia queste espressioni che l'alternativa "Altra". Prendendo spunto dal lavoro di Zhang e Ji [44], distinguiamo le AU in primarie e ausiliarie. Con AU primarie, intendiamo quelle AU o combinazioni di AU che possono essere chiaramente classificate come una delle sei espressioni senza ambiguità. Al contrario, le AU ausiliarie possono essere solo combinate additivamente con le AU primarie per fornire ulteriori indizi per il riconoscimento dell'espressione facciale. Di conseguenza, un'espressione può essere descritta sia usando le sole AU primarie che considerando

anche le AU ausiliarie. Inoltre, anche cambiamenti nelle caratteristiche transitorie del volto, come le rughe, possono aiutare a identificare certe espressioni. La Tabella 5.3 riassume quali sono le AU primarie e ausiliarie e le caratteristiche transitorie associate alle sei "espressioni principali".

Il FACS è un sistema pensato per osservatori umani, disegnato per individuare cambiamenti anche piccoli nei tratti del volto. Per poter estrarre in maniera automatica queste caratteristiche, le AU devono essere codificate quantitativamente mediante descrittori facciali che possano essere estratti direttamente dall'immagine del volto. In Fig. 5.6a sono mostrate le relazioni geometriche tra le varie parti del volto e sono evidenziate con rettangoli le regioni in cui si formano le rughe. Una corretta associazione tra ogni AU e i corrispondenti punti di riferimento del volto è fondamentale per un'accurata interpretazione dell'espressione. Per questo, tali associazioni sono state fatte a mano, come mostrato nella Tabella 5.4, in modo che i cambiamenti facciali visivi siano automaticamente misurabili. I descrittori delle AU sono stati quindi ottenuti misurando distanze e angoli tra i punti appropriati dell'*active appearance model* del volto. Tutte le corrispondenze sono mostrate nella Tabella 5.5.

Basandoci sulle relazioni riportate nella Tabella 5.3, includiamo in ciascuna funzione di utilità le AU legate alla corrispondente espressione. In particolare, ogni AU è definita come la combinazione lineare di descrittori facciali, secondo quanto espresso dalla Tabella 5.4. Proponiamo, quindi, tre modelli che si basano solamente sulle caratteristiche dei volti esaminati. Ogni modello è ottenuto aggiungendo al precedente un insieme di attributi nuovi.

Nel primo modello, le espressioni sono descritte solamente dagli indizi visivi primari. I parametri da stimare sono 67. L'espressione generale delle funzioni di utilità per ciascuna alternativa in questo modello è la seguente:

$$V_i = \alpha_i + \sum_{m=1}^{M} I_{mi}\beta_{mi}prim_m$$

Nel secondo modello, invece, in cui anche le AU ausiliarie sono considerate, i parametri incogniti sono 82, e la funzione di utilità assume la forma:

$$V_i = \alpha_i + \sum_{m=1}^{M} I_{mi}\beta_{mi}prim_m + \sum_{n=1}^{N} I_{ni}\beta_{ni}aux_n$$

Un sommario degli attributi inclusi in questi due modelli è riportato nella Tabella 5.6. Ogni colonna corrisponde ad una espressione, mentre le righe sono i descrittori facciali delle AU incluse in ciascuna utilità. Il simbolo '$\star$' corrisponde agli attributi $prim_m$ e il '$\bullet$' ad aux_n.

L'attivazione di muscoli facciali produce anche, in certe regioni del volto, rughe transitorie perpendicolari alla direzione del movimento del muscolo. Anche se le rughe presenti sulla fronte, nella regione nasolabiale e agli angoli degli occhi possono diventare permanenti all'aumentare dell'età, generalmente i movimenti muscolari causano cambiamenti nel loro aspetto, intensificandole o distendendole. Le caratteristiche transitorie possono dunque fornire informazioni ausiliarie per il riconoscimento delle espressioni facciali. Le regioni in cui compaiono le rughe sono quelle racchiuse da rettangoli in Fig. 5.6a. Il cambiamento delle rughe nella regione $\square X$ è incluso esplicitamente nella descrizione dell'AU9 (*Nose Wrinkler*); gli altri, invece, aiutano solamente ad identificare certe AU. Per esempio, le rughe che si formano nelle regioni $\square Z$, $\square Y$, $\square V$, $\square U$ offrono informazioni disgnostiche per l'identificazione, rispettivamente, delle AU2 (*Outer Brow Raiser*), AU4 (*Brow Lowerer*), AU6 (*Cheek Raiser*), e AU17 (*Chin Raiser*).

La presenza di rughe in una fotografia di un volto può essere determinata analizzando gli *edge* nelle regioni in cui ci si aspetta che tali rughe appaiano. A questo scopo, applichiamo, prima, una *edge detection with embedded confidence*, proposta da Meer e Georgescu [32], in cui la procedura comune di edge detection in tre passi: stima del gradiente, soppressione dei non massimi, thresholding a isteresi; è generalizzata per includere l'informazione data da misure di confidenza. Secondariamente, controlliamo la direzione degli edge estratti, per filtrare il rumore. Infatti, come possiamo vedere in Figura 5.6a, le rughe nelle regioni $\square Z$ e $\square X$ dovrebbero essere soprattutto orizzontali, e principalmente verticali quelle nella regione $\square Y$. Questa tecnica produce buoni risultati anche nel caso in cui sia presente qualche piccola ciocca di capelli sulla fronte, dal momento che i capelli sono spesso rappresentati da edge verticali, mentre ci si aspetta che le rughe della fronte siano prevalentemente orizzontali.

Per inserire nel nostro modello queste caratteristiche transitorie del volto, scegliamo di considerare il rapporto tra i pixel degli edge (rughe) e i pixel del background (pelle) per rappresentare le rughe nelle regioni $\square X$ e $\square Y$. Inoltre, usiamo due variabili categoriche per le rughe nasolabiali e per quelle della fronte: queste variabili hanno valore 1 se le rughe corrispondenti sono presenti nel viso, e zero altrimenti. Le funzioni di utilità per il terzo modello sono quindi ottenute aggiungendo queste variabili a quelle che descrivono le AU. Il modello risultante, in cui i parametri incogniti sono 93, è espresso da funzioni nella forma:

$$V_i = \alpha_i + \sum_{m=1}^{M} I_{mi}\beta_{mi}prim_m + \sum_{n=1}^{N} I_{ni}\beta_{ni}aux_n + \sum_{l=1}^{L} I_{li}\beta_{li}tr_l$$

La Tabella 5.7 mostra quali caratteristiche transitorie (tr_l) sono state incluse nelle funzioni di utilità di ciascuna espressione facciale.

L'idea che la scelta fatta dal decisore dipenda anche da alcune sue caratteristiche personali è uno dei nuovi aspetti affrontati in questo lavoro. Per identificare i fattori coinvolti nel processo di decisione, analizziamo le caratteristiche socio-economiche dei decisori. Tutte queste caratteristiche sono rappresentate nelle funzioni di utilità usando variabili categoriche. In altre parole, definiamo una variabile per ciascun sottoinsieme in cui sono divise le categorie di attributi socio-economici analizzati. Una categoria socio-economica è, ad esempio, "Studi" e i suoi sottoinsiemi sono "Maturità", "Laurea Universitaria", "Dottorato" e "Altro" (cfr. Tabella 5.1). La variabile categorica è uguale a 1 se il partecipante appartiene al sottoinsieme corrispondente, e zero altrimenti. Per esempio, $reg5$ avrà valore 1 per i decisori europei e zero per tutti gli altri. Per stimare e interpretare correttamente il modello, abbiamo bisogno di definire un sottoinsieme di riferimento per ogni categoria socio-economica. Laddove sono disponibili, scegliamo a questo scopo l'opzione "Nessuna". Negli altri casi, consideriamo l'inglese come la lingua di riferimento, e l'opzione "Altro" come il sottoinsieme di riferimento per "Studi" e per il luogo in cui si trova il partecipante quando risponde al questionario.

Nella Tabella 5.8 sono riportati gli attributi socio-economici inclusi nelle funzioni di utilità per ciascuna espressione in aggiunta ai descrittori facciali. Le funzioni di utilità per questo modello sono definite come segue:

$$V_i = \alpha_i + \sum_{m=1}^{M} I_{mi}\beta_{mi} prim_m + \sum_{n=1}^{N} I_{ni}\beta_{ni} aux_n + \sum_{l=1}^{L} I_{li}\beta_{li} tr_l + \sum_{h=1}^{H} I_{hi}\beta_{hi} socec_h$$

In questo modello completo, in cui sono stati considerati sia attributi del volto che del decisore, i parametri da stimare sono 177. Il modello qui riportato è quello con i parametri più significativi, rispetto ad una statistica t-test.

Dopo che abbiamo definito ciascun modello, decidendo quali attributi includere, stimiamo il valore dei parametri utilizzando BIOGEME. La stima dei parametri incogniti si basa sulla massima verosimiglianza. L'algoritmo usato per la massimizzazione individua massimi locali della funzione di verosimiglianza. Abbiamo ripetuto la stima più volte, scegliendo punti di partenza diversi (un modello iniziale con tutti i parametri a zero, e il valore stimato di diversi modelli intermedi). Tutti hanno dato la stessa soluzione. BIOGEME genera diversi file, che contengono informazioni sul processo di stima. In particolare, analizzeremo ora il file "Report", che contiene i risultati della stima a massima verosimiglianza, per capire quanto il modello proposto descriva accuratamente il comportamento di scelta studiato. Le tabelle contenenti i risultati della stima di ciascun modello sono riportate nell'Appendice A. Le ultime tre colonne delle tabelle contengono, rispettivamente, il valore stimato dei parametri β, l'errore standard associato e il valore del t-test. Un segno $(*)$ è applicato se il t-test fallisce, rispetto alla soglia specificata. Altrimenti, il parametro è significativamente diverso da zero, almeno al 95 per cento.

Nel nostro primo modello abbiamo fatto l'ipotesi che solo le AU primarie influenzino la scelta del decisore. Ai fini della stima, abbiamo normalizzato a zero l'alternativa "Neutra" e i coefficienti stimati sono quindi interpretati relativamente ad essa. Un segno negativo per le ASC di alcune alternative può essere interpretato come una preferenza del decisore per l'alternativa "Neutra" rispetto a quelle, mantenendo costante tutto il resto. Al contrario, il segno positivo di ASC_H indica una preferenza per l'alternativa "Gioia" rispetto a "Neutra".
Per quanto riguarda i coefficienti dei descrittori facciali, possiamo interpretarli ricordandoci che ogni AU è definita come una combinazione di misure il cui valore aumenta o diminuisce. Poichè valori dei descrittori facciali più elevati rispetto all'espressione neutra comportano valori di utilità più alti per le espressioni corrispondenti, ci aspettiamo coef-

ficienti positivi per le misure che dovrebbero aumentare. Viceversa, un segno negativo è atteso per distanze e angoli che si suppone debbano essere minori in una particolare espressione rispetto alla neutra. Confrontando con la Tabella 5.4 i risultati ottenuti, possiamo osservare che quasi tutti i coefficienti β sono consistenti con l'interpretazione proposta. Per esempio, il coefficiente $\beta_{brow\_dist\_A}$ è negativo e statisticamente diverso da zero, indicando che distanze maggiori tra le sopracciglia per l'espressione "Rabbia" rispetto a "Neutra" implicano un valore minore per l'utilità. Il valore positivo del parametro relativo all'altezza della bocca $\beta_{mouth\_height\_SU}$ è intuitivo: spesso associamo una bocca aperta verticalmente ad una espressione di sorpresa. Oltre a questo, il coefficiente $\beta_{eyeMouth\_dist\_r\_SA}$ ha segno positivo, indicando che gli angoli della bocca si abbassano in "Tristezza" rispetto a "Neutra", mentre il coefficiente $\beta_{eyeMouth\_dist\_r\_H}$ per la stessa misura in "Gioia" è negativo, come ci aspettavamo. E' interessante notare che per alcune misure i coefficienti hanno segno diverso se consideriamo il lato destro o sinistro del volto (i.e. $\beta_{eyeNose\_dist\_l\_O}$ e $\beta_{eyeNose\_dist\_r\_O}$). Questo può dipendere dal fatto che le maschere generate con l'AAM non corrispondono esattamente ai *landmark* reali del volto. Un'altra possibile spiegazione è che il volto stesso sia asimmetrico, anche nel caso neutro o solamente perché i muscoli non si muovono nello stesso modo su entrambi i lati del viso. Diversi studi hanno infatti misurato questa asimmetria nel movimento per quanto riguarda le espressioni simulate, dimostrando che certe azioni si manifestano in maniera più intensa qualche volta nel lato destro del volto e qualche volta nel sinistro.

Nel secondo modello, anche le AU ausiliarie sono state incluse nelle funzioni di utilità. Solamente pochissimi parametri sono risultati essere non statisticamente significativi per il modello. I valori stimati per le ASC mostrano che, mantenendo tutto il resto costante, c'è una preferenza nella scelta di "Gioia" rispetto alle altre alternative. Infatti, solamente

ASC_H ha segno positivo. Si potrebbe ipotizzare che una espressione di felicità sia più facile da riconoscere rispetto alle altre anche quando le sue caratteristiche distintive sono appena accennate sul volto. Tuttavia, potrebbero esistere anche altre spiegazioni, per esempio le immagini rappresentanti "Gioia" potrebbero essere più numerose delle altre nel database analizzato. Il parametro ASC_F ha cambiato segno, ma ora è significativo, mentre nel modello precedente non era significativamente diverso da zero.
I coefficienti dei descrittori facciali relativi alle AU ausiliarie sono significativamente diversi da zero e i loro segni sono consistenti con l'interpretazione proposta anche per le AU primarie. Infatti, le misure che dovrebbero decrescere hanno coefficienti negativi (e.g. $\beta_{brow\_dist\_SA}$), mentre per quelle che aumentano i coefficienti sono positivi (e.g. $\beta_{mouth\_height\_F}$). E' importante notare che i parametri relativi alle AU primarie sono ancora significativi, e i loro segni e ampiezze sono rimasti circa gli stessi del modello precedente.

Per testare se questo modello sia effettivamente migliore del precedente, utilizziamo il *likelihood-ratio test* (vedi Ben-Akiva e Lerman [5]). Il likelihood-ratio test (LRT) è un test statistico in cui un modello relativamente più complesso è confrontato con uno più semplice per vedere se è significativamente più adatto a descrivere un certo dataset. Se questa ipotesi è verificata, i parametri aggiuntivi del modello più complesso sono utilizzati in analisi successive. Il LRT è valido solamente se usato per confrontare modelli gerarchicamente annidati, cioè se il modello più complesso differisce da quello più semplice solamente per l'aggiunta di uno o più parametri. In questo caso, il modello da testare è quello in cui le AU ausiliarie sono state aggiunte alle primarie. Introdurre nuovi parametri comporta sempre un likelihood più alto. Tuttavia, non sempre questo incremento è tanto significativo da giustificare la ulteriore complessità del modello. Il LRT fornisce un criterio oggettivo per scegliere tra diversi modelli possibili. Il likelihood-ratio test inizia con un confronto dei

valori del log-likelihood per i due modelli:

$$LR = -2(lnL_0 - lnL_1)$$

dove L_0 e L_1 sono, rispettivamente, il massimo della funzione di likelihood sotto l'ipotesi nulla (modello semplice) e il massimo con quel vincolo rilassato (modello complesso). Questa statistica asintoticamente segue una distribuzione chi-quadro. Per determinare se la differenza tra i valori del likelihood dei due modelli è statisticamente significativa, dobbiamo poi considerare i gradi di libertà, d. Nel LRT, i gradi di libertà sono pari al numero di nuovi parametri aggiunti nel modello più complesso:

$$d = D_1 - D_0$$

dove D_0 e D_1 sono il numero di parametri nei due diversi modelli. Usando questa informazione possiamo quindi ricercare il valore critico della statistica test in tabelle statistiche standard. Rifiutiamo l'ipotesi nulla che il modello senza AU ausiliarie sia più appropriato per il dataset analizzato, se

$$-2(lnL_0 - lnL_1) > \chi^2_{((1-\alpha),d)}$$

in cui α è il livello di significatività. In questo caso specifico, i gradi di libertà sono $d = 82 - 67 = 15$ e scegliendo $\alpha = 0.01$ otteniamo

$$-2(-47143.9 + 46109.2) = 2069.4 > 30.58$$

Possiamo quindi rifiutare l'ipotesi nulla e concludere che i coefficienti delle AU ausiliarie dovrebbero essere inclusi nel modello.

Nel caso del modello in cui sono state aggiunte le caratteristiche transitorie, quasi tutte le variabili sono statisticamente significative e hanno il segno atteso. I parametri stimati sono ancora consistenti con l'interpretazione proposta nei paragrafi precedenti. In particolare, i coefficienti delle caratteristiche transitorie hanno tutti segno positivo, e questo indica che

la presenza di rughe nel volto ha un impatto positivo nelle utilità corrispondenti. Per esempio, il fatto che $\beta_{forehead\_SU}$ e $\beta_{forehead\_A}$ siano positivi riflette una preferenza per le alternative "Sorpresa" e "Rabbia" rispetto a "Neutra" quando sono presenti rughe sulla fronte. Inoltre, il valore maggiore di $\beta_{forehead\_SU}$ indica un impatto più forte di questo attributo nell'utilità di "Sorpresa" che in quella di "Rabbia". Una considerazione simile vale per $\beta_{nasolabial\_D}$, il coefficiente di una variabile categorica uguale a 1 se la ruga nasolabiale tipica di "Disgusto" appare sul viso. Per quanto riguarda i parametri $\beta_{naswrink\_D}$ e $\beta_{browwrink\_A}$, possiamo osservare che sono significativamente diversi da zero, indicando che più rughe sono presenti rispettivamente nelle regioni $\square X$ e $\square Y$ di Figura 5.6a e più è forte l'impatto di questi parametri nelle utilità corrispondenti. L'interpretazione delle ASC è analoga a quella dei modelli precedenti. Tali parametri sono tutti significativamente diversi da zero, i loro segni non sono cambiati e anche il loro valore è rimasto sostanzialmente lo stesso.

A questo punto, possiamo applicare il likelihood-ratio test tra il modello con le caratteristiche transitorie (M_2) e il modello con le sole AU primarie e ausiliarie (M_1). In questo caso, l'ipotesi nulla è che tutti i coefficienti per i descrittori delle rughe valgano zero. Anche in questo caso, $-2(lnL_1 - lnL_2)$ ha una distribuzione chi-quadro, con $d = 93 - 82 = 11$ gradi di libertà. Se consideriamo di nuovo $\alpha = 0.01$, otteniamo:

$$-2(-46109.2 + 45563.2) = 1092 > 24.73$$

Possiamo quindi rifiutare l'ipotesi nulla e concludere che aggiungere variabili per le caratteristiche transitorie migliora il modello.

Nell'ultimo modello aggiungiamo alcune variabili per descrivere il decisore. L'interpretazione dei parametri per le misure dei tratti del volto e delle ASC data per i modelli precedenti è applicabile anche in questo caso. I coefficienti per le variabili socio-economiche sono stati stimati e la maggior parte di essi è significativamente diversa da zero al 95 per cento.

Possiamo osservare che il parametro β_{gender_SA} nell'alternativa "Tristezza" ha segno negativo. Questo implica che gli uomini hanno probabilità più bassa delle donne di scegliere l'alternativa "Tristezza" rispetto alla "Neutra", dal momento che la variabile "Genere" vale 1 se il decisore è di sesso maschile. Il segno positivo del coefficiente dell'età β_{age1_F} riflette una preferenza degli individui più giovani (18-40 anni) per l'alternativa "Paura" rispetto a "Neutra". Il coefficiente relativo alla lingua del decisore è positivo nell'utilità di "Disgusto" per gli individui di lingua italiana. Questo può indicare che gli italiani riescono a distinguere tra l'espressione neutra e disgustata con più facilità dei partecipanti di lingua inglese.

Per testare se l'aggiunta di variabili socio-economiche migliora la descrittività del modello, usiamo ancora una volta il LRT. Confrontiamo le funzioni di log-likelihood di questo modello e del precedente, in cui venivano considerati solamente i descrittori del volto. In questo caso, i gradi di libertà sono $d = 177 - 93 = 84$. Il valore critico per $\chi^2_{(0.99,90)}$ risulta essere 124.1, quindi la condizione per rifiutare l'ipotesi nulla è

$$-2(-45563.2 + 45285.4) = 555.6 > 124.1$$

ed è soddisfatta. E' dunque possibile concludere che il modello contenente sia le caratteristiche socio-economiche che gli attributi che descrivono il volto dovrebbe essere preferito agli altri modelli presentati nei paragrafi precedenti.

Abbiamo dimostrato che gli attributi che abbiamo inserito nei modelli proposti procurano informazione significativa e influenzano la scelta del decisore. Ora vogliamo testare quanto questi modelli riescano a predire bene l'alternativa scelta, dato un insieme di caratteristiche note. Per farlo, utilizziamo BioSim, uno strumento che può predire, per ogni modello stimato con BIOGEME, le probabilità di ciascuna alternativa per ogni osservazione. Infatti, i coefficienti stimati di un DCM possono esse-

re usati per calcolare la probabilità di scelta di ciascuna alternativa per ciascuna osservazione nel campione analizzato. Se diamo il file dei dati e il file del modello in input a BioSim, viene generato un file di output che contiene, per ciascuna osservazione, l'alternativa scelta e le probabilità per tutte le alternative del choice set. In Figura 6.1 sono riportati i grafici delle probabilità predette dell'alternativa scelta per ciascuna osservazione del dataset, ottenute con i quattro diversi modelli. In blu sono riportate le probabilità calcolate sul dataset utilizzato per stimare il modello (circa 31000 osservazioni) e in rosso le probabilità per l'insieme di validazione (circa 8000 immagini etichettate). Per decidere se una particolare scelta è stata riprodotta correttamente, confrontiamo la probabilità dell'alternativa scelta con una soglia, definita come la probabilità nel caso peggiore. Il caso peggiore si verifica quando tutte le alternative hanno la stessa probabilità predetta, cioè $p = \frac{1}{9} \approx 0.12$ per questo particolare problema. Infatti, questo significa che non abbiamo informazione per predire la scelta. Quindi, se la probabilità dell'alternativa scelta è maggiore della soglia, possiamo dire che il modello ha riprodotto bene il comportamento del decisore. Come risulta dalla Tabella 6.5, i quattro modelli mostrano buone prestazioni. La percentuale di osservazioni correttamente predette aumenta passando dal primo modello (74.55% per il training set e 73.44% per le nuove osservazioni) all'ultimo (76.14% per il grafico blu e 75.36% per quello rosso). Questo significa che aggiungere attributi nelle funzioni di utilità migliora i risultati. Inoltre, è importante notare che l'andamento dei grafici è simile per il training set e per il validation set. Da questo possiamo dedurre che i nostri modelli riescono a generalizzare bene. Possiamo quindi concludere che i *discrete choice models* hanno offerto prestazioni buone e incoraggianti, dimostrandosi un valido approccio per affrontare il problema dell'analisi automatica di espressioni facciali.

Questo lavoro rappresenta il primo tentativo di utilizzare la *Discrete Choice Analysis* per modellare le espressioni facciali, perciò varie strade sono percorribili per sviluppare ulteriormente il metodo di analisi proposto. Dal momento che uno degli aspetti chiave dei DCM è la possibilità di scegliere quale forma usare per le funzioni di utilità, alcune modifiche possono essere apportate in questa direzione. Una prima idea potrebbe essere quella di *segmentare* la popolazione che ha partecipato al questionario, suddividendola in gruppi sulla base delle caratteristiche socio-economiche. Per esempio, si potrebbe studiare il comportamento di uomini e donne analizzando i due insiemi di persone separatamente, invece che tramite una variabile categorica. In questo modo potrebbero infatti venire alla luce informazioni utili a stabilire con maggiore precisione quali caratteristiche siano realmente coinvolte nel processo di decisione. Allo stesso tempo, si potrebbe ridurre la complessità del problema, diminuendo il numero di parametri incogniti. I modelli potrebbero inoltre essere resi più robusti grazie all'aggiunta di nuovi parametri, come descrittori dell'evoluzione dinamica delle espressioni facciali. Infine, potrebbero essere scelti altri tipi di espressioni per le funzioni di utilità, come, ad esempio, combinazioni non-lineari di parametri. Tuttavia, la stima di modelli di questo tipo richiede risorse computazionali elevate, quindi, affinchè il problema sia risolvibile, in questi casi si dovrebbe limitare il numero di parametri del modello.

Acknowledgements

This thesis work was done during six months I spent at the Signal Processing Institute of EPFL (Ecole Polytechnique Fédérale de Lausanne), in Switzerland. I would like to acknowledge all the people supporting this project.

I thank my supervisor, Prof. Alessandro Mecocci, and my supervisor at EPFL, Prof. Jean Philippe Thiran, for giving me the opportunity to work on this interesting project and for their help and guidance throughout it.

I would also like to thank my academic advisor, Ph.d. student Matteo Sorci, for his large and consistent interest in my project and for his constructive feedback.

I gratefully acknowledge the technical guidance of Prof. Michel Bierlaire throughout the discrete choice analysis and his constructive comments on my work.

Thanks should also go to Dr. Gianluca Antonini, for his interest and fruitful discussions on my work during the thesis period.

Finally, I am very grateful to all the people who partecipated in the survey, without whom I couldn't have done this thesis.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER

ONE

Introduction

When people interact to communicate, a very important role is played by the face. In fact, thanks to it we can get information about our interlocutors: who they are, what they feel, what their intentions are, etc. Studies demonstrated that facial expression is even more significant than verbal communication, and we can daily experience this in our life (when we want to tell something important to someone, we often prefer to have this person in front of us). This is why various researchers have been interested in this topic, during past centuries and until nowadays. In particular, in the last years, a new aspect of facial expression analysis has been tackled: how expression recognition could be automatically performed by a computer. A lot of different algorithms have been proposed, most of them using traditional classification techniques for identifying an expression on a face. In the present work, we want to demonstrate the validity of a new approach, based on discrete choice analysis, for associating a face image with the expression it seems to represent. Moreover, we want to prove that the process of recognizing an expression depends not only on the characteristics of the analysed face, but also on the characteristics of the analysing people.

The data we use for this study are obtained from a survey proposed

on the web to people from all over the world. A series of face images is shown to the participants, who are asked to select the most suitable expression for each face. The expression must be chosen in a set of nine: the six Ekman's "basic expressions" (happiness, sadness, anger, disgust, surprise, fear), the neutral face, and two options, "Other" and "I don't know", included to account for possible interpretation ambiguity.

These expressions are thus the *choice set* for the discrete choice models we propose. Each alternative (expression) is represented in the model by an *utility function*. These functions are linear combinations of parameters, describing both features of the face and personal characteristics of the decision maker. A random term is also included in each utility function, to represent the uncertainty given by attributes not included in the model. Facial expressions are described by using the facial action coding system (FACS), proposed by Ekman and Friesen in 1978. In this system, each muscular activity in the face is coded and quantified by measures named "action units" (AUs). Facial descriptors are extracted from face images by applying active appearance modelling (AAM) techniques.
The utility function measures how well a particular combination of parameters represent a certain expression. Each sample in the data set is therefore linked with the utility function it maximizes. Unknown model's parameters are estimated with a maximum likelihood technique, by means of BIOGEME package, developed by Bierlaire at EPFL (Ecole Polytechnique Fédérale de Lausanne).

Anyway, it is worth noting that some boundaries are posed in our study. In the first place, we consider static images, so we don't include in the analysis the aspects related to the dynamic evolution of facial expressions. Secondly, images represent posed expressions, that can look different from spontaneous ones. Finally, we analyse data collected with the survey as of a certain date, but new people have answered ever since, as the survey is still available on the web. So our data set is only a part

of all the observations that could be studied in the future.

Under this conditions, our operating procedure is the following. We define four discrete choice models, where each one is obtained by adding new parameters to the precedent, in order to improve the descriptiveness of each utility function. In the first model, we include only primary action units, namely, facial features linked without ambiguity to each basic expression. In the second model, we introduce in each utility function the auxiliary action units, that are not necessarily present on the face for recognizing a particular expression. The third model uses also attributes related to transient changes of the face, like wrinkles and furrows. Finally, in the fourth model we add the socio-economic characteristic of the participants in the survey. The performances of all the models are then compared, in order to demonstrate that most of the considered attributes are important to recognize the expression on a face, included descriptors of the decision maker.

This work is structured as follows:

- In Chapter 2, we introduce the issue of facial expression analysis, and in particular we focus on the work of Paul Ekman.
- In Chapter 3, we present the principal techniques for automatic facial expression analysis.
- In Chapter 4, we review the AAM and introduce the DCM theory.
- In Chapter 5, a detailed description of the utility functions of each model is given, along with the attributes description and the results of the learning process.
- In Chapter 6, we report the experiments and the comparison between the different models.
- In Chapter 7, conclusions and future works are finally reported.

- The Appendix A is dedicated to the full tables of the estimation results of the four proposed models.

CHAPTER

TWO

Overview on Facial Expression

Facial expressions are an important channel of nonverbal communication: they provide cues about emotional response, regulate interpersonal behaviour, communicate aspects of psychopathology, etc. Expression implies a revelation about the characteristics of a person, a message about something internal to the expresser. Even though the human species has acquired the powerful capabilities of a verbal language, the role of facial expressions in person-to-person interactions remains substantial, since messages of the face provide significant commentary and illustration about verbal communications.

The study of human facial expressions has therefore many aspects, from computer simulation and analysis to understanding its role in art, nonverbal communication and the emotional process, and many different investigators have dealt with it. In particular, three basic questions have arisen: is there any relationship between emotion and facial expression? Are facial expressions culturally bound or universal? And, are any universals in expressions biologically based? In this section we will review the answers that have emerged to these questions.

2.1 What Does a Facial Expression Mean?

The term "expression" implies the existence of something that is expressed. Some psychologists deny that there is really any specific organic state that corresponds to our naive ideas about human emotions. Other psychologists think that the behaviours referenced by the term "expression" are part of an organized emotional response, and thus, the term "expression" captures these behaviours' role less adequately than a reference to it as an aspect of the emotion reaction. Still other psychologists think that facial expressions have primarily a communicative function and convey something about intentions or internal state, and they find the connotation of the term "expression" useful. Regardless of approach, the evidence on the universe of facial expression indicates that it is a large and complex set. The relation of expressions to emotions is precise and refined. Anyway, it does not mean that facial expression *is* emotion, since emotions are not the only source of facial expression (see Fig. 2.1). The next paragraphs briefly discuss how some important theories view the relation between facial expressions and emotions.

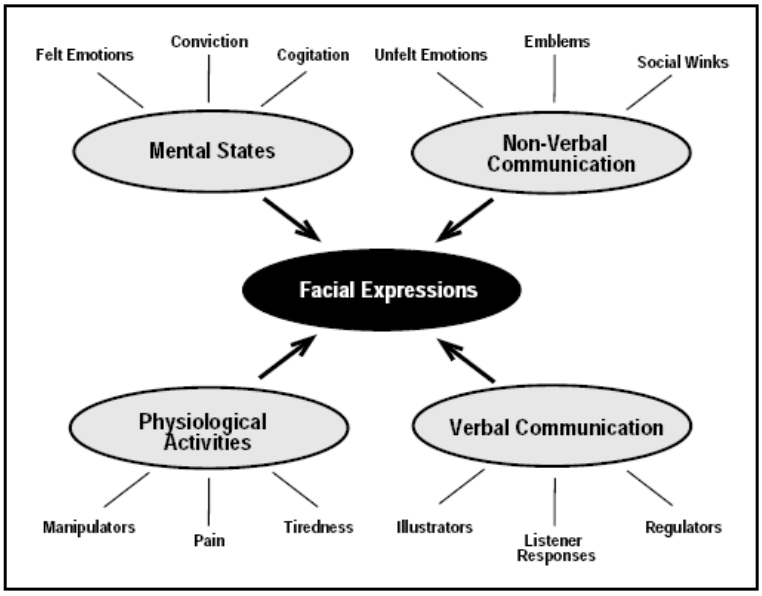


Figure 2.1: *Sources of facial expressions.*

2.1.1 Psychological Interpretations

The study of facial expression is a very old matter and it is not clear where such a story should begin. Observations about emotions appearing on the face can be found in various ancient and medieval writers. However, only recently psychologists have assessed the relation between feelings and expression nearly unanimously: the face is the key to understanding emotion, and emotion is the key to understanding the face.

Linking face to emotions may be common sense, but it has turned out to be the single most important idea in the psychology of emotion. It is central to a research program that claims Silvan Tomkins, a mid-20th century psychologist, as its modern theorist. He proposed that the sensations provided by emotional expressions, vascular changes, and other changes in the face are the source of the qualitatively different feelings of emotion, e.g., happy from sad, fear from anger. Perception of other bodily changes provides less specific feelings of emotion. Tomkins also argued that there are specific categories of emotion that have evolved for certain functional, adaptive reasons, which are likewise reflected in neural organization. These categories of emotion correspond to specific categories of facial expressions and are organized around their facial expressions. For example, emotions related to disgust derive from the prototype of rejecting food that is noxious or dangerous to eat, with a core expression of opening the mouth and lips, and pushing out with the tongue. Silvan Tomkins's new look at facial expression and emotion was largely responsible for encouraging the work of colleagues in the late 20th century that resulted in a heightened place in psychology for these topics.

2.1.2 A Modern Scientific Treatment of Facial Expressions

The scientific study of the facial expression of emotion began with Charles Darwin's *The Expression of Emotions in Man and Animals*, first pub-

lished in 1872, which provided a basis for considering facial expressions as behaviours that evolved as a mechanism of communication. He can be considered as the originator of the theory furthered by Tomkins. Moreover, Darwin was one of the first scientists to use photographs as illustrations (see Fig. 2.2) and to use the judgement method for studying the signal value of an expression – which has become the most frequently used method in the psychology of expression.

Figure 2.2: *Pictures from Darwin's "The Expression of Emotions in Man and Animals" [9].*

Among his many extraordinary contributions, Darwin gathered evidence that some emotions have a universal facial expression, cited examples and published pictures suggesting that emotions are evident in other animals, and proposed principles explaining why particular expressions occur for particular emotions – principles that, he maintained, applied to the expressions of all animals.

In his work, Darwin argued that certain emotional expressions are innate and the same for all people. His evidence for universality was the answers

to 16 questions he sent to Englishmen living or travelling in eight parts of the world: Africa, America, Australia, Borneo, China, India, Malaysia and New Zealand. Even by today's standard, that is a very good, diverse, sample. They wrote that they saw the same expressions of emotion in these foreign lands as they had known in England, leading Darwin to say: "It follows, from the information thus acquired, that the same state of mind is expressed throughout the world with remarkable uniformity[1]". However, there are three problems that make Darwin's evidence on universality unacceptable by today's scientific standards. First, Darwin did not ask a sufficient number of people in each country to answer his questions. Second, Darwin relied upon the answers of these Englishmen, rather than asking the people who were native in each country (or asking his English correspondents to do so). Current research always studies the people who are native to each country, not a foreign observer's interpretation of their behaviour. Third, the way in which Darwin worded his questions often suggested the answer he wanted.

Although Darwin himself put little emphasis on the communicative potential of facial expression of emotion as an object of adaptive selection, the thrust of his general work suggests this connection and encouraged later scientists to elaborate upon this mechanism. One branch of this tradition is the approach to studying animal behaviours known as *ethology*. Early ethologists, such as Konrad Lorenz, studied stimulus-response patterns in animals, where fixed action patterns are elicited by distinct sign-stimuli having evolutionary significance, or releasers. Later, ethologists studied human behaviour in light of these concepts and findings, and began to elaborate the communicative significance of both human and animal facial expressions. Eibl-Eibesfeld, for example, studied facial expressions, such as smiling, and other specific facial behaviours, such as the eyebrow flash, in the context of their adaptive value in a communica-

[1] C. Darwin, *The Expression of Emotions in Man and Animals*, 1872, p. 10

tive framework.

2.2 Ekman and Friesen: a New Important Contribution

Darwin's theory on universality of facial expressions of emotion was largely ignored by scientists in the subsequent century. Instead, the view that facial expressions are not valid indicators of emotion was widely accepted even though the evidence was contradictory. Ekman, Friesen, and Ellsworth (1972, 1982) resolved this issue definitively by pointing out methodological problems that had confused other researchers. They showed that observers could agree on how to label both posed and spontaneous facial expressions in terms of either emotional categories or emotional dimensions. The labels judges assigned to posed expressions tended to agree with the poser's intended message. For spontaneous expressions, judges selected labels consistent with emotions appropriate in the situations that elicited the expressions. They proposed that the universal

Figure 2.3: *Six of the photographs used by Ekman, in 1966, in his cross-cultural research for demonstrating the universality of facial expressions.*

in facial expressions of emotion was the connection between particular facial configurations and specific emotions. That does not mean that expressions will always occur when emotions are experienced, for we are capable of inhibiting our expressions. Nor does it mean that emotions will always occur when a facial expression is shown, for we are capable of fabricating an expression. Contrary to the belief of some anthropologists at the time, Ekman found that facial expressions of emotion are not culturally determined, but universal to human culture and thus biological in origin, as Charles Darwin had once theorized. He also developed the Facial Action Coding System (FACS) to taxonomise every conceivable human facial expression.

2.2.1 Basic Emotions

To match a facial expression with an emotion implies knowledge of the categories of human emotions into which expressions can be assigned. For millennia, scholars have speculated about categories of emotion, and, in 1971, Ekman and Friesen [14] postulated six primary emotions that possess each a distinctive content together with a unique facial expression. These prototypic emotional displays are also referred to as "basic emotions". They seem to be universal across human ethnicities and cultures and include happiness, sadness, anger, disgust, fear and surprise, though many other categories are possible and used by philosophers, scientists, actors, and others concerned with emotion. The recent development of scientific tools for facial analysis, such as the FACS, has facilitated resolving category issues. The most robust categories are discussed in the following paragraphs.

Happiness

Happy expressions are universally and easily recognized, and are interpreted as conveying messages related to enjoyment, pleasure, a positive

disposition, and friendliness. Examples of happy expressions are the easiest of all emotions to find in photographs, and are readily produced by people on demand in the absence of any emotion. In fact, happy expressions may be practised behaviours because they are used so often to hide other emotions and deceive or manipulate other people. Consider this point when viewing invariably smiling political figures and other celebrities on television. Detecting genuine happy expressions may be as valuable as producing good simulations.

Sadness

Sad expressions are often conceived as opposite to happy ones, but this view is too simple, although the action of the mouth corners is opposite. Sad expressions convey messages related to loss, bereavement, discomfort, pain, helplessness, etc. Until recently, American culture contained a strong censure against public displays of sadness by men, which may account for the relative ease of finding pictures of sad expressions on female faces. A common sense view, shared by many psychologists, is that sad emotion faces are lower intensity forms of crying faces, which can be observed early in newborns, but differences noted between these two expressions challenge this view, though both are related to distress. Although weeping and tears are a common concomitant of sad expressions, tears are not indicative of any particular emotion, as in tears of joy.

Anger

Anger expressions are seen increasingly often in modern society, as daily stresses and frustrations underlying anger seem to increase, but the expectation of reprisals decrease with the higher sense of personal security. Anger is a primary concomitant of interpersonal aggression, and its expression conveys messages about hostility, opposition, and potential attack. Anger is a common response to anger expressions, thus creating a positive feedback loop and increasing the likelihood of dangerous con-

flict. Until recent times, a cultural prohibition on expression of anger by women, particularly uncontrolled rage expressions, created a distribution of anger expressions that differed between the sexes. The uncontrolled expression of rage exerts a toxic effect on the angry person, and chronic anger seems associated with certain patterns of behaviour that correspond to unhealthy outcomes, such as Type A behaviour. Although frequently associated with violence and destruction, anger is probably the most socially constructive emotion as it often underlies the efforts of individuals to shape societies into better, more just environments, and to resist the imposition of injustice and tyranny.

Disgust

Disgust expressions are often part of the body's responses to objects that are revolting and nauseating, such as rotting flesh, insects in food, or other offensive materials that are rejected as suitable to eat. Obnoxious smells are effective in eliciting disgust reactions. Disgust expressions are often displayed as a commentary on many other events and people that generate adverse reactions, but have nothing to do with the primal origin of disgust as a rejection of possible foodstuffs.

Fear

Fear expressions are not often seen in societies where good personal security is typical, because the imminent possibility of personal destruction, from interpersonal violence or impersonal dangers, is the primary elicitor of fear. Fear expressions convey information about imminent danger, a nearby threat, a disposition to flee, or likelihood of bodily harm. The specific objects that can elicit fear for any individual are varied. The experience of fear has an extremely negative felt quality, and is reduced, along with the bodily concomitants, when the threat has been avoided or has passed. Organization of behaviour and cognitive functions are adversely affected during fear, as escape becomes the peremptory goal.

Anxiety is related to fear and may involve some of the same bodily responses, but is a longer term mood and the elicitors are not as immediate. Both are associated with unhealthy physical effects if prolonged.

Surprise

Surprise expressions are fleeting, and difficult to detect or record in real time. They almost always occur in response to events that are unanticipated, and they convey messages about something being unexpected, sudden, novel, or amazing. The brief surprise expression is often followed by other expressions that reveal emotion in response to the surprise feeling or to the object of surprise, emotions such as happiness or fear. For example, most of us have been surprised, perhaps intentionally, by people who appear suddenly or do something unexpected ("to scare you"), and elicit surprise, but if the person is a friend, a typical after-emotion is happiness; but if a stranger, fear. A surprise seems to act like a reset switch that shifts our attention. Surprise expressions occur far less often than people are disposed to say "that surprises me", etc., because in most cases, such phrases indicate a simile, not an emotion. Nevertheless, intellectual insights can elicit actual felt surprise and may spur scholarly achievements. Surprise is to be distinguished from startle, and their expressions are quite different.

2.2.2 Facial Action Coding System

The evidence shows that facial expressions are related to emotion both biologically and culturally, but, until recently, all the evidence was based on observers' judgements of the face and few studies have tried to measure every possible facial expression and what the cues are for each emotion. Over the years various procedures for facial measurement have been invented. Nevertheless, most research on facial behaviour has measured the information that observers were able to infer from the face, instead

of measuring the face itself, mainly because of the problems of devising an adequate technique. Another problem which has plagued previous attempts to measure facial movement has been how to describe most precisely each measurement unit. A new versatile method for measuring and describing facial behaviours, the Facial Action Coding System, was developed in 1976 by Paul Ekman and Wallace Friesen. This system is now the most used standard for measuring facial expressions in the behavioural sciences and is widely accepted by researchers in the area of automated facial expression recognition. Likewise, we also adapt the FACS to describe the six primary emotional expressions in our model.

Facial expressions represent a visible consequence of facial muscle and autonomic nervous system actions: is it possible to describe and quantify every action the face can perform? Ekman and Friesen provided an answer to this question with their FACS, by measuring all visible facial movements. Ideally, FACS would differentiate every change in muscular action, but it is limited to what a user can reliably discriminate when movements are inspected repeatedly, in stopped and slowed motion. It does not measure invisible changes (e.g., certain changes in muscle tonus) or vascular and glandular changes produced by the autonomic nervous system. Limiting FACS measurement to visible movements was consistent with an interest in those behaviours which may be social signals, usually detected during social interactions. FACS can be applied to any reasonably detailed visual record of facial behaviour. If the technique were to measure invisible or autonomic nervous system activity, it would be limited to situations where sensors were attached (e.g., EMG electrodes) or special sensing and recording methods were used (e.g., thermography).

The primary goal in developing FACS was to create a *comprehensive* system, which could measure and distinguish all possible, visually discriminable facial actions. Comprehensiveness is important because

many of the fundamental questions about the universe and nature of facial expressions cannot be answered if just a subset of behaviours is measurable. FACS was derived from an analysis of the anatomical basis for facial movement. A comprehensive system was obtained by discovering how each muscle of the face acts to change visible appearances. With this knowledge it is possible to analyse any facial movement into anatomically based, minimal action units.

Another consideration that guided the development of FACS was the need to separate description from inferences about the meanings of behaviours. Scoring is less likely to be biased if the observer does not have to evaluate or attach meanings to behaviours. Almost all the previous descriptive systems have included some inferential scores, such as "aggressive frown" (Grant, 1969), "lower lip pout" (Blurton-Jones, 1971), etc. Each of these actions should be described in noninferential terms. Blurton-Jones (1971) noted that facial activity could be described in three ways: the location of shadows and lines; the muscles responsible; or the main positions of landmarks, such as mouth corners or brow location. He opted for the last basis, since it seemed to him "more convenient if description could be given which did not require that anyone who uses them should learn the facial musculature first, although knowledge of the musculature obviously improves the acuity of one's observations[2]". On the contrary, Ekman and Friesen have adopted almost the opposite position. FACS emphasizes patterns of movement, the changing nature of facial appearance. Distinctive actions are described: the movements of the skin, the temporary changes in size and location of the features, and the gathering, pouching, bulging, and wrinkling of the skin. The user of FACS learns the mechanics or muscular basis of facial movement, not simply the consequences of actions or a description of static landmarks. As time passes, FACS users increasingly focus on behavioural description

[2]N.G. Blurton-Jones, *Criteria for use in describing facial expressions in children*, 1971, p. 369

and are rarely aware of "meanings". So, by emphasizing measurement of the face in terms of muscle actions, inferences about meanings are minimized.

FACS's emphasis on movement and muscular action also helps overcome problems due to physiognomic differences between people. Individuals differ in the size, shape, and location of their features and in permanent wrinkles, bulges, or pouches which become permanent in mid-life. The particular shape of a landmark may vary from one person to another; for example, when the lip corner goes up, the angle, shape, or wrinkle pattern may not be the same for all people. If only the end result of movement is described, scoring may be confused by physiognomic variation. Knowledge of the muscular basis for actions helps deal with these differences.

FACS measurement units are called "action units" (AUs) and represent the muscular activity that produces momentary changes in facial appearance. There are two reasons for using the term "action unit" instead of "muscle unit". Firstly, this is because a few times they have combined more than one muscle in a single AU. Secondly, the appearance changes produced by one muscle (as defined by anatomists) were sometimes separated into two or more AUs, to represent relatively independent actions of different part of that muscle. For example, following Hjorstjo's lead [21], the frontalis muscle which raises the brow was separated into two action units, depending upon whether the inner or outer portion of this muscle lifts the inner or outer portions of the eyebrow.

There are 46 AUs which account for changes in facial expression, and 12 AUs which describe changes in gaze direction and head orientation in coarser terms (as shown in Tables 2.1- 2.2). Table 2.1 lists the names, numbers and anatomical basis of each action unit. Most of the action units involve a single muscle. The numbers are arbitrary and do not have any significance except that 1 through 7 refer to brows, forehead or eyelids. The table indicates where more than one muscle is combined into

AU	FACS Name	Muscular Basis
1	Inner Brow Raiser	Frontalis, Pars Medialis
2	Outer Brow Raiser	Frontalis, Pars Lateralis
4	Brow Lowerer	Depressor Glabellae; Depressor Supercilli; Corrugator
5	Upper Lid Raiser	Levator Palpebrae Superioris
6	Cheek Raiser	Orbicularis Oculi, Pars Orbitalis
7	Lid Tightener	Orbicularis Oculi, Pars Palebralis
8	Lips Toward Each Other	Orbicularis Oris
9	Nose Wrinkler Levator	Labii Superioris, Alaeque Nasi
10	Upper Lip Raiser Levator	Labii Superioris, Caput Infraorbitalis
11	Nasolabial Furrow Deepener	Zygomatic Minor
12	Lip Corner Puller	Zygomatic Major
13	Cheek Puffer	Caninus
14	Dimpler	Buccinnator
15	Lip Corner Depressor	Triangularis
16	Lower Lip Depressor	Depressor Labii
17	Chin Raiser	Mentalis
18	Lip Puckerer	Incisivii Labii Superioris; Incisivii Labii Inferioris
20	Lip Stretcher	Risorius
22	Lip Funneler	Orbicularis Oris
23	Lip Tightner	Orbicularis Oris
24	Lip Pressor	Orbicularis Oris
25	Lips Part Depressor	Labii, or Relaxation of Mentalis or Orbicularis Oris
26	Jaw Drop	Masetter; Temporal and Internal Pterygoid Relaxed
27	Mouth Stretch	Pterygoids; Digastric
28	Lip Suck	Orbicularis Oris
38	Nostril Dilator	Nasalis, Pars Alaris
39	Nostril Compressor	Nasalis, Pars Transversa and Depressor Septi Nasi
41	Lid Droop	Relaxation of Levator Palpebrae Superioris
42	Slit	Orbicularis Oculi
43	Eyes Closed	Relaxation of Levator Palpebrae Superioris
44	Squint	Orbicularis Oculi, Pars Palpebralis
45	Blink	Relaxation of Levator Palpebrae and Contraction of Orbicularis Oculi, Pars Palpebralis
46	Wink	Orbicularis Oculi

Table 2.1: *Action Units: changes in facial expression.*

AU	Description
51	Head turn left
52	Head turn right
53	Head up
54	Head down
55	Head tilt left
56	Head tilt right
57	Head forward
58	Head back
61	Eyes turn left
62	Eyes turn right
63	Eyes up
64	Eyes down

Table 2.2: *Action Units: changes in head direction and in gaze orientation.*

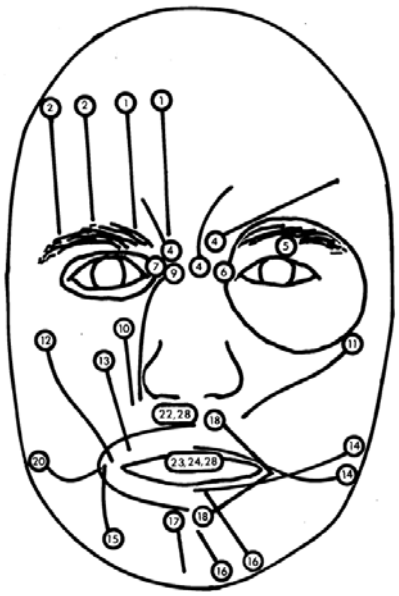


Figure 2.4: *Schematic portrayal of FACS measurement units. The number in the circle indicates the action unit. The circle represent a relatively fixed point towards which the skin is pulled along the radiating line. Picture from [17].*

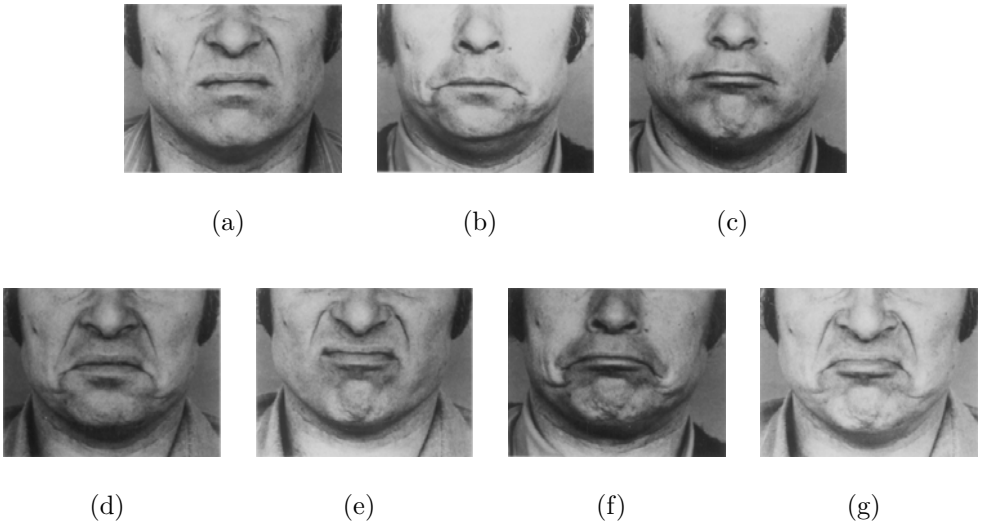


Figure 2.5: *Action Units 10, 15, 17 and their combinations. (a) AU10, (b) AU15, (c) AU17, (d) AU10+AU15, (e) AU10+AU17, (f) AU15+AU17, (g) AU10+AU15+AU17. Pictures from [17].*

a single action unit, or where more than one action unit is separated from a single muscle. The FACS names given in the table are a shorthand, not meant to describe the appearance changes, but a convenience to call them to mind.

After determining the single AUs, Ekman and Friesen performed and examined between 4000 and 5000 AU combinations. Study of these combinations showed that most of the appearance changes were additive (i.e., each AU was clearly recognizable and virtually unchanged). There were a few AU combinations which were not additive, but instead showed new appearances. All of these distinctive combinations are described in FACS in the same detail as the single AUs.

Reliability was a major concern in the development of FACS. Ekman and Friesen (1978) studies have repeatedly shown good reliability even when the learner uses only the self-instructional FACS manual without direct guidance from FACS's authors. The evidence shows that FACS can successfully measure the visibly distinctive facial actions as its authors

intended.

Besides being reliable, FACS has revealed the answer to many basic questions about expressions. From the single AUs and their combinations, Ekman and Friesen have estimated that there are several hundred thousand possible visibly distinguishable facial expressions, most of which are never seen on people's faces in everyday life. FACS has been used to score pictures of faces which observers have judged to express emotion and to score faces of people in emotionally arousing situations. Based on evidence from such scoring, the expressions produced by different combinations of AUs which convey emotional meanings appear to number in the hundreds, if not thousands. If the strength of muscular contraction and the timing or sequence of muscular recruitment were included, this number would be substantially increased. Of course, people do not have a different emotion name for each of these expressions. Instead, many emotional expressions are synonyms or convey different connotations of particular emotions. Observers also perceive differences in the intensity of emotion expressions which may be based on strength of muscular contraction, number of muscles recruited, or area of the face in which contractions occur. The number of expressions conveying emotional meanings is much greater than researchers have typically acknowledged, but it is much smaller than the number of possible expressions.

Every facial muscle can be involved in one or more emotional expressions, so there is no distinction between emotional and nonemotional muscles. Some muscles always signal a particular emotion, such as zygomatic major which produces a smile and is characteristic of happiness. It is never involved in a negative emotional expression without blending its own message. Other muscles, such as the corrugator, are involved in expressions which convey many different emotional messages and nonemotional messages. Some emotions, such as happiness and disgust, can be signalled by the action of only one muscle, but other emotions, such as sadness, need the action of more than one muscle to be signalled unam-

biguously.

The disadvantage of this approach is that it is labour intensive and insensitive to very slow changes in muscle tonus. Another approach for measuring facial expressions in muscular or anatomical terms, the electromyography (EMG), overcomes these problems. In this technique, surface electrodes placed over different regions of the face measure electrical discharge from contracting muscular tissue through the skin. The EMG signal lends itself to immediate recording and is sensitive to slight muscular movements that may not be visible even to trained eyes. However, also this technique has at least two main disadvantages. The first one is that EMG is highly obtrusive: the application of surface electrodes makes subjects aware of the facial measurement. The second drawback is that the recording selectivity of facial EMG is not muscle specific, but rather regionally specific, and it is not yet certain whether EMG allows the differentiation of as many different emotions as can be done with measurement that relies upon observer scoring of visible muscular actions. On the contrary, the FACS is precise, able to specify which muscles were active, and allows measurement of any movement, not just an a-priori set pre-determined by the placement of EMG leads. Moreover, the facial action coding system is unobtrusive, performed from videotape records or pictures, without intruding on the subject.

FACS is a very elaborate system, much more comprehensive than any previous technique. There is no facial action described by other systems which cannot be described by FACS, and there are many behaviours described by FACS not previously distinguished. FACS allows for scoring asymmetries, either in terms of different AUs or different intensities. A means for measuring the intensity and the timing of actions is also detailed.

FACS scores are descriptive only, and provide no implications about the meaning of the behaviour. Analysis of the data can use only these raw FACS scores, or the scores can be translated into more psychologi-

cally meaningful concepts by techniques such as FACSAID, a database interpretation system available to researchers.

CHAPTER

THREE

Automatic Facial Expression Analysis

Since the mid 1970s, different approaches have been proposed for facial expression analysis, but in the nineties this field of study gained much inertia starting with the pioneering work of Mase and Pentland [30]. The reasons for this renewed interest in facial expressions are multiple, but mainly due to advancements accomplished in related research areas such as face detection, face tracking and face recognition as well as the recent availability of relatively cheap computational power. Various applications using automatic facial expression analysis can be envisaged in the near future, fostering further interest in doing research in different areas, including image understanding, psychological studies, facial nerve grading in medicine, face image compression and synthetic face animation, video-indexing, robotics as well as virtual reality.

Facial expression analysis includes both measurement of facial motion and recognition of expression. The general approach to Automatic Facial Expression Analysis (AFEA), shown in Figure 3.1, consists of three steps:

1. Face detection;

2. Facial expression data extraction;

3. Facial expression recognition.

Figure 3.1: *Basic architecture of Automatic Facial Expression Analysis.*

Before a facial expression can be analysed, the face must be detected in a scene. Next step is to devise mechanisms for extracting the facial expression information from the observed facial image. In the case of static images, the process of extracting the facial expression information is referred to as localizing the face and its features in the scene. The face can be represented in various ways, e.g., as a whole unit (*holistic representation*), as a set of features (*analytic representation*) or as a combination of these (*hybrid approach*). The applied face representation and the kind of input images determine the choice of mechanisms for automatic extraction of facial expression information. The final step is to define some set of categories, which we want to use for facial expression classification and/or facial expression interpretation, and to devise the mechanism of categorization. In the next paragraphs we survey the facial expression analysis techniques presented in literature in the past decade.

3.1 Face Detection

For most works in automatic facial expression analysis, the conditions under which a facial image is obtained are controlled. Usually, the image has the face in frontal view. Hence, the presence of a face in the scene is ensured and some global location of the face in the scene is known a priori. However, determining the exact location of the face in a digitized facial image is a complex problem, e.g., the scale and the orientation of

the face can vary from image to image. Thus, it is difficult to search for a fixed pattern (template) in the image. The presence of noise and occlusion makes the problem even more difficult. Detection of the exact face position in an observed image or image sequence has been approached in two ways. In the holistic approach, the face is determined as a whole unit. In the second, analytic approach, the face is detected by finding some important facial features first (e.g., the irises and the nostrils). The location of the features in correspondence with each other determines then the overall location of the face. Table 3.1 provides a classification of facial expression analysers according to the applied method.

	Reference	**View**	**Method**	**Comments**
Facial images				
Holistic approach	Huang	Frontal View	Canny edge detector PDM model fitting	No rigid head rotations
	Pantic	Dual view	Image histogram analysis Thresholding	Mounted camera on the subject's head
Analytic approach	Hara	Frontal view	Brightness distribution	No rigid head motions Real-time process
	Yoneyama	Frontal view	-	-
	Kimura	Frontal view	Integral projection Potential Net fitting	No rigid head rotation

Table 3.1: *Summary of the methods for automatic face detection.*

To represent the face as a whole unit (holistic approach), Huang and Huang [23] apply a Point Distribution Model (PDM). In order to achieve a correct placement of an initial PDM in an input image, Huang and Huang utilize a Canny edge detector to obtain a rough estimate of the face location in the image. The valley in pixel intensity that lies between the lips and the two symmetrical vertical edges representing the outer vertical boundaries of the face generate a rough estimate of the face location. The face should be without facial hair and glasses, no rigid head motion may be encountered and illumination variations must be linear for the system to work correctly.

Pantic and Rothkrantz [34] detect the face as a whole unit, too. As input to their system, they use dual-view facial images. To determine the vertical and horizontal outer boundaries of the head, they analyse the vertical and horizontal histogram of the frontal-view image. To localize the contour of the face, they use an algorithm based on the HSV colour model, which is similar to the algorithm based on the relative RGB model [41]. For the profile view image they apply a profile-detection algorithm, which represents a spatial approach to sampling the profile contour from a thresholded image. For thresholding of the input profile image, the "value" of the HSV model is exploited. No facial hair or glasses are allowed.

Kobayashi and Hara [26] apply an analytic approach to face detection. They are using a CCD camera in monochrome mode to obtain brightness distribution data of the human face. First, "base" brightness distribution was calculated as an average of brightness distribution data obtained from ten subjects. Then, the system extracts the position of the irises by applying a crosscorrelation technique on the "base" data and the currently examined data. Once the irises are identified, the overall location of the face is determined by using relative locations of the facial features in the face. The observed subject should face the camera while siting at approximately one metre distance in front of it.

Yoneyama et al. [42] use an analytic approach to face detection too. The outer corners of the eyes, the height of the eyes, and the height of the mouth are extracted in an automatic way. Once these features are identified, the size of the examined facial area is normalized and an 8×10 rectangular grid is placed over the image. It is not stated which method has been applied and no limitation of the used method has been reported by Yoneyama et al.

Kimura and Yachida [25] utilize a Potential Net for face representation. An input image is normalized first by using the centres of the eyes and

the centre of the mouth. This algorithm applies an integral projection method, which synthesizes the colour and the edge information. Then, the Potential Net is fitted to the normalized image to model the face and its movement. The face should be without facial hair and glasses and in a direct face-to-face position with the camera.

3.2 Facial Expression Data Extraction

After the presence of a face has been detected in the observed scene, the next step is to extract the information about the encountered facial expression in an automatic way. If the extraction cannot be performed automatically, a fully automatic facial expression analyser cannot be developed. Both, the applied face representation and the kind of input images affect the choice of the approach to facial expression information extraction.

In general, three types of face representation are mainly used in facial expression analysis: holistic (e.g., isodensity maps), analytic (e.g., deformable templates), and hybrid (e.g., analytic-to-holistic approach). The most common face representation techniques are listed in Table 3.2. Depending on the face model, a template-based or a feature-based method is applied for facial expression data extraction (see Table 3.3). Template-based methods fit a holistic face model to the input image. Feature-based methods localize the features of an analytic face model in the input image.

3.2.1 Template-Based Methods

As shown in Tables 3.2-3.3, a first category of techniques for AFEA from static images applies a holistic or a hybrid approach to face representation and a template-based method for facial expression information extraction from an input image. Edwards et al. [12] use a holistic face representation, which they refer to as the Active Appearance Model

Reference	Model
Holistic approach	
Edwards	AAM
Hong	Labeled graph
Huang	PDM
Padgett	Random block eigenvectors
Black	Optical flow (in facial regions)
Otsuka	Optical flow (in facial regions)
Analytic approach	
Hara	FCPs model and 13 vertical lines
Pantic	Dual-view point-based model
Zhao	Frontal-view point-based model
Cohn	Optical flow (facial points)
Hybrid approach	
Lyons	Fiducial grid & Gabor wavelets
Yoneyama	8×10 quadratic grid
Zhang	Fiducial points & Gabor wavelets
Essa	Optical flow (whole face)
Kimura	Potential Net
Wang	Labeled graph

Table 3.2: *Face models.*

(AAM). To build their model they used facial images that were manually labelled with 122 points localized around the facial features. To generate a statistical model of shape variation, Edwards et al. aligned all training images into a common coordinate frame and applied PCA to get a mean shape. To fit the AAM to an input image, they apply an AAM search algorithm. The method works with images of faces without facial hair and glasses, which are hand-labelled with the landmark points beforehand approximated with the proposed AAM.

Hong et al. [22] utilize a labelled graph to represent the face. Each node of the graph consists of an array, which is called "jet". Each component of a jet is the filter response of a certain Gabor wavelet extracted at a point of the input image. Hong et al. use wavelets of five different frequencies and eight different orientations. They defined two different labelled graphs, called General Face Knowledge (GFK): one is used to find the exact face location in an input facial image and the other one is used to localize the facial features. Then Hong et al. apply the PersonSpotter system [37]

and the method of elastic graph matching proposed by Wiskott [40] to fit the model-graph to a surface image. The dense model-graph seems very suitable for facial action coding based on the extracted deformations of the graph. However, this issue has not been discussed by Hong et al. As seen in Section 3.1, Huang and Huang [23] represent the face with a PDM, in which the mouth is included by approximating its contour with three parabolic curves. Since the proposed model is a combination of the PDM and a mouth template, it is arguably as close to a feature-based model as to a template-based model. Anyway, it can be classified as a holistic face model since the PDM models the face as a whole and interacts with the estimated face region of an input image as entire. Af-

Reference	**Method**	**Comment**
	Analysis from static facial images	
Template-based methods		
Edwards	A multivariate multiple regression for modelling the relationship between the AAM displacement and the image difference and in the recognition phase to match the AAM to the input image.	Direct frontal view Faces without facial hair, glasses Hand labelling of the images
Hong	Fitting a labelled graph to an input facial image by utilizing Wiskott's method of elastic graph matching.	Faces without facial hair, glasses Slightly rotated faces allowed Real-time process
Huang	Fitting the PDM by applying a gradient-descent-based shape parameters estimation; fitting 3 parabolas to the mouth by applying gradient-based edge detector	Direct frontal view Faces without facial hair, glasses No variation of the background
Yoneyama	Gradient-based optical flow algorithm for estimating an averaged optical flow in 80 20×20 pixels regions of the grid placed over a normalized image.	Direct frontal view Faces without facial hair, glasses Averaging the flow (drawback) Horizontal movement isn't modelled
Feature-based methods		
Hara	Extracting the brightness distribution data along the 13 vertical facial lines; CCD camera in monochrome mode used.	Direct frontal view Faces without facial hair, glasses Horizontal movement isn't modelled Real-time process
Pantic	Multiple feature detectors are applied per facial feature. From the localized contours of the prominent facial features the model features are extracted.	Dual view images Faces without facial hair, glasses 2 cameras mounted on user's head

Table 3.3: *Automatic facial expression data extraction techniques.*

ter an initial placement of the PDM in the input image, the method of Huang and Huang moves and deforms the entire PDM simultaneously. Here, a gradient-based shape parameters estimation, which minimizes the overall grey-level model fitness measure, is applied. The search for the mouth starts by defining an appropriate search region on basis of the fitted PDM. Successfulness of this method is strongly constrained.
Padgett and Cottrell [33] also use a holistic face representation, but they do not deal with facial expression information extraction in an automatic way. They made use of the facial emotion database assembled by Ekman and Friesen [15], [16], digitized 97 images of six basic emotional facial expressions, and scaled them so that the prominent facial features were located in the same image region. Then, in each image, the area around each eye was divided into two vertically overlapping 32×32 pixel blocks and the area around the mouth was divided into three horizontally overlapping 32×32 pixel blocks. PCA of 32×32 pixel blocks randomly taken over the entire image was applied in order to generate the eigenvectors. The input to a neural network used for emotional classification of an expression was the normalized projection of the seven extracted blocks on the top 15 principal components.

Yoneyama et al. [42] use a hybrid approach to face representation. They fit an 8×10 quadratic grid to a normalized facial image. Then, an averaged optical flow is calculated in each of the regions. The magnitude and the direction of the calculated optical flows are simplified to a ternary value magnitude in only the vertical direction. The information about a horizontal movement is excluded. Hence, the method will fail to recognize any facial appearance change that involves a horizontal movement of the facial features. The face should be without facial hair and glasses and no rigid head motion may be encountered for the method to work correctly.

Zhang et al. [43] use a hybrid approach to face representation, but do not deal with facial expression information extraction in an automatic way. They use 34 facial points for which a set of Gabor wavelet coeffi-

cients is extracted.

A similar face representation was recently used by Lyons et al. [29] for expression classification into the six basic plus "neutral" emotional categories. They used a fiducial grid of manually positioned 34 nodes on the 256×256 pixels images used in [43], but apply wavelets of five spatial frequencies and six angular orientations.

3.2.2 Feature-Based Methods

The second category of methods for automatic facial expression analysis from static images uses an analytic approach to face representation and a feature-based method for expression information extraction from an input image. In their work [26], Kobayashi and Hara utilize a CCD camera in monochrome mode to obtain a set of brightness distributions of 13 vertical lines crossing the FCPs. The range of the acquired brightness distributions is normalized to [0, 1] and these data are given further to a trained neural network for expression emotional classification. A shortcoming of the proposed face representation is that the facial appearance changes encountered in a horizontal direction cannot be modelled. The real-time system developed by Kobayashi and Hara works with on-line taken images of subjects with no facial hair or glasses facing the camera while siting at approximately one metre distance from it.

Pantic and Rothkrantz [34] are using a point-based model composed of two 2D facial views, the frontal and the side view. To localize facial features and then extract the model features in an input dual-view, Pantic and Rothkrantz apply multiple feature detectors for each prominent facial feature (eyebrows, eyes, nose, mouth, and profile). Then, the best of the acquired (redundant) results is chosen. This is done based on both, the knowledge about the facial anatomy (used to check the correctness of the result of a certain detector) and the confidence in the performance of a specific detector (assigned to it based on its testing results). The per-

formance of the detection scheme was tested on 496 dual views. Human observers in 89 percent approved when visually inspected the achieved localization of the facial features. The system cannot deal with minor inaccuracies of the extracted facial data and it deals merely with images of faces without facial hair or glasses.
Zhao et al. [45] also use a point-based frontal-view face model but do not deal with automatic facial expression data extraction.

3.3 Facial Expression Recognition

After the face and its appearance have been perceived, the next step of an automated expression analyser is to classify (identify, interpret) the facial expression conveyed by the face. A fundamental issue about the facial expression classification is to define a set of categories we want to deal with. A related issue is to devise mechanisms of categorization. Facial expressions can be classified in various ways: in terms of facial actions that cause an expression, in terms of some nonprototypic expressions such as "raised brows" or in terms of some prototypic expressions such as emotional expressions. Some of the systems perform both the mechanisms of categorization, based on a particular facial action or on a particular basic emotion.

The facial action coding system [16] is probably the most known study on facial activity. As described in Section 2.2.2, it is a system developed to facilitate objective measurement of facial activity for behavioural science investigations of the face. Automating FACS would make it widely accessible as a research tool in the behavioural science, which is furthermore the theoretical basis of multimodal/media user interfaces. This triggered researchers of computer vision field to take different approaches in handling the problem. Among the attempts to adapt FACS for coding automatically facial actions, we can mention Zhang and Ji [44] facial expression representation, based on dynamic Bayesian networks combined

with the facial action units.

Most of the studies on automated expression analysis perform an emotional classification. As indicated by Fridlund et al. [19], the most known and the most commonly used study on emotional classification of facial expressions is the cross-cultural study on existence of "universal categories of emotional expressions". Ekman defined six such categories, referred to as the "basic emotions" (see Section 2.2.1). In the past years, many questions arose around this study: are the basic emotional expressions indeed universal, or are they merely a stressing of the verbal communication and have no relation with an actual emotional state? Also, the six basic emotion categories are enough for classifying each facial expression able to be displayed on the face? Despite that, most of the studies on vision-based facial expression analysis rely on Ekman's emotional categorization of facial expressions.

Three more issues are related to facial expression classification in general. First, the classification mechanism may not depend on physiognomic variability of the observed person (the system should be capable of analysing any subject, male or female of any age and ethnicity). On the other hand, each person has their own maximal intensity of displaying a particular facial expression. Therefore, if the obtained classification is to be quantified (e.g., to achieve a quantified encoding of facial actions or a quantified emotional labelling of blended expressions), systems which can start with a generic expression classification and then adapt to a particular individual have an advantage. Second, it is important to realize that the interpretation of the body language is situation-dependent. However, the information about the context in which a facial expression appears is very difficult to obtain in an automatic way. This issue has not been handled by the currently existing systems. Finally, there is now a growing psychological research that argues that timing of facial expressions is a critical factor in the interpretation of expressions. For

the researchers of automated vision-based expression analysis, this suggests moving towards a real-time whole-face analysis of facial expression dynamics.

While the human mechanisms for face detection are very robust, the same is not the case for interpretation of facial expressions. It is often very difficult to determine the exact nature of the expression on a person's face. According to Bassili [4], a trained observer can correctly classify faces showing six basic emotions with an average of 87 percent. This ratio varies depending on several factors: the familiarity with the face, the familiarity with the personality of the observed person, the general experience with different types of expressions, the attention given to the face and the non-visual cues (e.g., the context in which an expression appears). It is interesting to note that the appearance of the upper face features plays a more important role in face interpretation as opposed to lower face features.

Independently of the used classification categories, the mechanism of classification applied by a particular surveyed expression analyser is either a template-based- or a neural-network-based- or a rule-based- classification method. Table 3.4 summarizes some of the most common methods for facial expression emotional classification.

If a template-based classification method is applied, the encountered facial expression is compared to the templates defined for each expression category. The best match decides the category of the shown expression. In general, it is difficult to achieve a template-based quantified recognition of a non-prototypic facial expression. There are infinitely a lot of combinations of different facial actions and their intensities that should be modelled with a finite set of templates. The problem becomes even more difficult due to the fact that everybody has his/her own maximal intensity of displaying a certain facial action. The better results are obtained by Lyons et al. [29], since the recognition rate is 92 percent for the

Reference	Method	#	Test cases	Accuracy
Analysis from static facial images				
Template-based methods				
Edwards	PCA based on Mahalanobis distance and LDA	7	200 images 25 subjects	74%
Hong	Personalised galleries and Elastic graph matching	7	>175 images 25 subjects	81%
Huang	2D emotion space (PCA) & minimum distance classifier	6	90 images 15 subjects	84.5%
Lyons	PCA and LDA of the labelled graph vectors	7	193 images 9 Japanese females	75-92%
Yoneyama	Two 14×14 Hopfield NNs with learning	4	-	-
Neural-network-based methods				
Hara	$234 \times 50 \times 6$ NN with backpropagation learning	6	90 images 15 subjects	85%
Padgett	$15 \times 10 \times 7$ NN with backpropagation learning	7	84 Ekman's photos	86%
Zhang	$646 \times 7 \times 7$ NN with RPROP propagation	7	213 images 9 Japanese females	90%
Zhao	$10 \times 10 \times 3$ NN with backpropagation learning	6	94 Ekman's photos	100%
Rule-based methods				
Pantic	Expert System rules	6	265 dual views 8 subjects	91%

Table 3.4: *Methods for facial expression emotional classification.*

familiar subjects and 75 percent in the case of unknown persons. They presented a Gabor wavelet-based method. The facial feature points sampled from a sparse grid covering on the face are represented by a set of Gabor filters and are then combined to form a single feature vector. The principle components of the feature vectors from training images are further analysed by linear discriminant analysis to form discriminant vectors. Finally, classification was performed by projecting the input vector of a test image along the discriminant vectors.

A second category of the surveyed methods for automatic facial expression analysis from static images applies a neural network for facial expression classification. Although the neural networks represent a "black box" approach and arguably could be classified as template based methods, they are classified as separate methods. This distinction is required because a typical neural network can perform a quantified facial expres-

sion categorization into multiple classes while, in general, the template-based methods cannot achieve such a performance. In a neural-network-based classification approach, a facial expression is classified according to the categorization process that the network "learned" during a training phase. A significant amount of research on spatial analysis for facial expression recognition has focused on using neural networks [33], [45], [43]. They differ mainly in their input facial data, which are either brightness distributions of feature regions, principle components of facial images, or even an entire face image. Most of the neural-network-based classification methods perform facial expression classification into a single category.

The rule-based classification methods classify the examined facial expression into the basic emotion categories based on the previously encoded facial actions. The prototypic expressions, which characterize the emotion categories, are first described in terms of facial actions. Then, the shown expression, described in terms of facial actions, is compared to the prototypic expressions defined for each of the emotion categories and classified in the optimal fitting one. Just one of the surveyed methods for automatic facial expression analysis from static images applies a rule-based approach to expressions classification. The method proposed by Pantic and Rothkrantz [34] achieves automatic facial action coding from an input facial dual-view in few steps. They used a dual-view face model (see Section 3.2.2) to extract facial features in order to reduce the ambiguities of face geometry. The extracted facial data is converted to a set of rule descriptors based on FACS. The classification of facial expressions is performed by comparing the AU-coded description of observed expression against the rule descriptors of six facial expressions. The average recognition rate was 92 percent for the upper face AUs and 86 percent for the lower face AUs.

In general, the existing expression analysers assign the examined ex-

pression to one of the basic emotion categories proposed by Ekman and Friesen [15]. This approach to expression classification has two main limitations. First, *pure* emotional expressions are seldom elicited. Most of the time, people show blends of emotional expressions. Therefore, classification of an expression into a single emotion category is not realistic. An automated facial expression analyser should realize quantified classification into multiple emotion categories. Second, it is not at all certain that all facial expressions displayed on the face can be classified under the six basic emotion categories. So even if an expression analyser performs a quantified expression classification into multiple basic emotion categories, it would probably not be capable of interpreting each and every encountered expression. In the present work, we try to tackle this issues. We use nine categories to classify facial expressions. In addition to the six basic expressions (happiness, sadness, anger, disgust, fear, surprise), we consider the neutral one and two new categories: "other", which should take in account all the expressions that cannot be tied to any of the main seven classes, and "don't know", to deal with the ambiguity of the face. Facial expressions are coded as suggested in the FACS. The relations of each facial expression to the corresponding combination of action units are derived from the work of Zhang and Ji [44]. Furthermore, we propose Discrete Choice Models (DCMs) for modelling expressions, basing on the previous work of Antonini and Sorci [1]. The logic behind the use of DCMs is to model the choice process representing the human observer labelling procedure.

CHAPTER

FOUR

Active Appearance Model and Discrete Choice Model

As we have discussed in Chapter 3, various techniques are adopted for automatic facial expression analysis, which differ in the approaches used for the main steps of the analysis. In our work we choose the Active Appearance Model (AAM) as the feature extractor method (see Section 4.1), since it has proved to be successful for coding and interpreting face images, also providing a useful basis for locating faces in images. Once a proper face representation has been defined, we propose Discrete Choice Models for expression modelling, as described in Section 4.2.

4.1 Active Facial Appearance Model

Faces are highly variable, deformable objects, and manifest very different appearances in images depending on pose, lighting, expression, and the identity of the person. Interpretation of such images requires the ability to understand this variability in order to extract useful information. Particularly suited to this task, the Active Appearance Models are a group of highly flexible deformable models introduced in the 1998 paper "Interpreting Face Images using Active Appearance Models" by Edwards, Taylor and Cootes [13]. This was a proposal and implementation of a statistical entity capable of capturing full appearance of faces – an ap-

pearance that can be faithfully described by the generic object shape[1] mapped with some overlaid textures. Such models express not only the variation of shape, but also pixel intensities that are vital for full reconstruction and synthesis of valid realistic model instances. In other words, by quoting Edwards [12], "the AAM contains a statistical, photo-realistic model of the shape and grey-level appearance of faces" .

The models are generated by combining a model of face shape variation with a model of the appearance variations of a shape-normalized face. Active appearance models learn what are valid shape and intensity variations from their training set. The training set consists of labelled images, where key landmark[2] points are marked on each example object. In our case, the object of interest is the face, so landmarks are placed around the main facial features (Fig. 4.1).

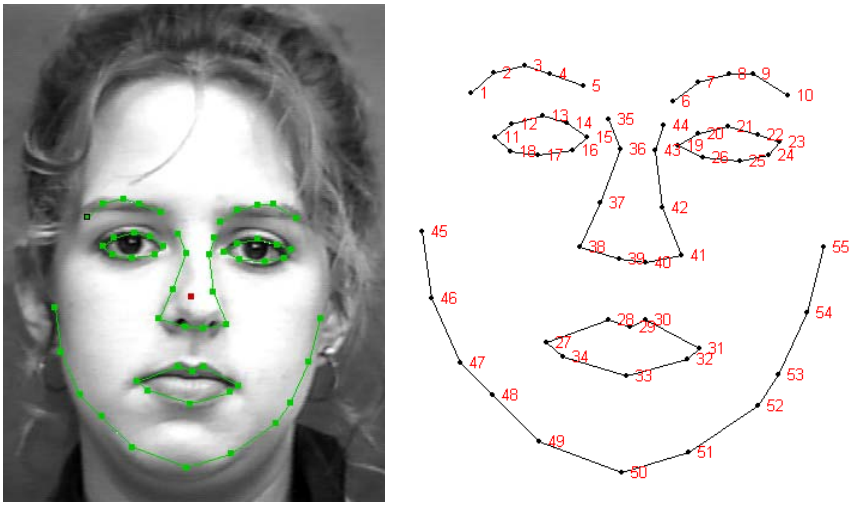


Figure 4.1: *Facial landmarks (55 points).*

Principal Component Analysis (PCA) is applied to the set of vectors describing the shapes in the training set, for building the statistical shape model. The labelled points, **s**, on a single object describe the shape of

[1] **Shape** is all the geometrical information that remains when location, scale and rotational effects are filtered out from an object.

[2] A **landmark** is a point of correspondence on each object that matches between and within populations.

that object. The vector $\mathbf{s}$ is brought into a common normalized frame – w.r.t. position, scale and rotation – to which all shapes are aligned. Any example can then be approximated by using:

$$\mathbf{s_i} = \bar{\mathbf{s}} + \mathbf{\Phi_s b_{si}} \tag{4.1}$$

where $\mathbf{s_i}$ is the synthesised shape, $\bar{\mathbf{s}}$ is the mean shape vector, $\mathbf{\Phi_s}$ is a set of orthogonal *modes of shape variation* and $\mathbf{b_{si}}$ is a vector of shape parameters.

After having computed the mean shape $\bar{\mathbf{s}}$ and aligned all the shapes from the training set by means of a Procrustes transformation (as described in [38]), it is possible to warp textures from the training set onto the mean shape $\bar{\mathbf{s}}$, in order to obtain shape-free patches. Similarly to the shape, after computing the mean shape-free texture $\bar{\mathbf{g}}$, all the textures in the training set can be normalized with respect to it by scaling and offset of luminance values. By applying PCA to the normalized data we obtain a linear model:

$$\mathbf{g_i} = \bar{\mathbf{g}} + \mathbf{\Phi_t b_{ti}} \tag{4.2}$$

where $\mathbf{g_i}$ is the shape-free texture, and, similarly as above, $\bar{\mathbf{g}}$ is the mean normalized grey-level vector, $\mathbf{\Phi_t}$ is a set of orthogonal *modes of intensity variation* and $\mathbf{b_{ti}}$ is a vector of grey-level parameters.

The shape and appearance of any example can thus be summarised by the vectors $\mathbf{b_{si}}$ and $\mathbf{b_{ti}}$. The unification of the presented shape and texture models into one complete appearance model is obtained by concatenating the vectors $\mathbf{b_{si}}$ and $\mathbf{b_{ti}}$ and learning the correlations between them by means of a further PCA. The combined statistical model is then

given by:

$$\mathbf{s_i} = \bar{\mathbf{s}} + \mathbf{Q_s c_i} \tag{4.3}$$

$$\mathbf{g_i} = \bar{\mathbf{g}} + \mathbf{Q_t c_i} \tag{4.4}$$

where $\mathbf{Q_s}$ and $\mathbf{Q_t}$ are the matrices describing the principal modes of the combined variations in the training set and $\mathbf{c_i}$ is the *appearance* parameters vector, allowing to control simultaneously both shape and texture. Note that the linear nature of the model allows us to express the shape and grey-levels directly as functions of $\mathbf{c_i}$. A face can be synthesised for a given $\mathbf{c_i}$ by generating the shape-free grey-level image from the vector $\mathbf{g_i}$, then warping it using the control points described by $\mathbf{s_i}$, so that the model points lie on the image points (see [11] for details). To complete the description of the face, the four *pose* parameters are also needed, namely $\mathbf{p} = (\alpha, \vartheta, t_x, t_y)$, representing scale, orientation and position, respectively.

The AAM method attempts thus to synthesise the complete appearance of the target image, by choosing parameters that minimise the difference between the target image and an image generated from the model.

Figure 4.2: *Optimization process for the AAM. Left: initial model. Middle: model after 2 iterations. Right: converged model.*

4.2 Discrete Choice Models

Discrete choice models have been recently introduced in the computer vision community by Antonini et al. [2], in the context of pedestrian modelling and tracking. DCMs are known in econometrics since the late 1950s. They are defined to describe and forecast the behaviour of people (*decision makers*) in choice situations, when the set of available alternatives is finite and discrete (*choice set*). In this context, the logic behind the use of DCMs is to model the choice process representing the human observer labelling procedure.

4.2.1 Modelling Assumptions

Dealing with human behaviour makes the system under consideration rather complex. In order to obtain *operational* models, namely models with parameters and variables that can be measured or estimated, some simplifying assumptions need to be made. A specific model will correspond to a specific set of assumptions about the decision maker, the alternatives, the attributes of alternatives.

The *decision maker* can be a person, a household, a firm, or any other decision-making unit, but he has to be an *individual*. This means that if we consider that a group of persons is the decision maker, we have to consider only the decision of the group as a whole. Because of its *disaggregate* nature, the model has to include the characteristics, or attributes, of the individual (e.g. age, gender, income, etc.), named *socio-economic characteristic*. Furthermore, the decision maker is assumed to be *rational*, in that he is supposed to perform a choice so as to maximize the utility he perceives from the alternatives. In the following we will refer to "decision maker" and "individual" interchangeably, as well as using "analyst" for "researcher". Moreover, we will use "he" for the decision maker and "she" for the researcher, in order to refer to both people in the same paragraph without ambiguity.

The *alternatives* might represent competing products, courses of action, or any other options or items over which choice must be made. To fit within a discrete choice framework, the set of alternatives needs to exhibit three characteristics. First, the alternatives must be mutually *exclusive* from the decision maker's perspective. The decision maker chooses only one alternative from the choice set. Second, the choice set must be *exhaustive*, in that all possible alternatives are included. The decision maker necessarily chooses one of the alternatives. Third, the number of alternatives must be *finite*. The researcher can count the alternatives and eventually be finished counting. The first and second criteria are not restrictive. Appropriate definition of alternatives can nearly always assure that the alternatives are mutually exclusive and the choice set is exhaustive, and the researcher often has several approaches for doing so. In contrast, the third condition is actually restrictive. This condition is the defining characteristic of discrete choice models and distinguishes their realm of application from that for regression models. With regression models, the dependent variable is continuous, which means that there is an infinite number of possible outcomes. When there is an infinite number of alternatives, discrete choice models cannot be applied.

Each alternative in the choice set must be characterized by a set of *attributes*. Similarly to the characterization of the decision-maker, the researcher has to identify the attributes of each alternative that are likely to affect the choice of the individual. Some attributes may be generic to all alternatives, and some may be specific to an alternative. An attribute is not necessarily a directly observed quantity: it can be any function of available data.

4.2.2 Derivation of Choice Probabilities

Discrete choice models are usually derived under an assumption of utility-maximizing behaviour by the decision maker. The *utility* is an index of

the attractiveness of an alternative. A decision maker, labelled n, faces a choice among J alternatives. The decision maker would obtain a certain level of utility (or profit) from each alternative. The utility that decision maker n obtains from alternative j is U_{nj}, $j = 1, ..., J$. He chooses the most attractive alternative, that is the one that provides the greatest utility. The behavioural model is therefore: choose alternative i if and only if $U_{ni} > U_{nj}, \forall j \neq i$.

Consider now the researcher: she does not observe the decision maker's utility. The researcher observes some attributes of the alternatives as faced by the decision maker, labelled $x_{nj}\, \forall j$, and some attributes of the decision maker, labelled s_n, and can specify a function that relates these observed factors to the decision maker's utility. The function is denoted $V_{nj} = V(x_{nj}, s_n)$, $\forall j$, and is often called *representative utility*.

Since there are aspects of utility that the analyst does not or cannot observe, $V_{nj} \neq U_{nj}$. Utility can be written as

$$U_{nj} = V_{nj} + \varepsilon_{nj},$$

where ε_{nj} captures the factors that affect utility but are not included in V_{nj}. The researcher does not know $\varepsilon_{nj}\, \forall j$ and therefore treats these terms as random. The joint density of the random vector $\varepsilon_n = \langle \varepsilon_{n1}, ..., \varepsilon_{nJ} \rangle$ is denoted $f(\varepsilon_n)$. With this density, the researcher can make probabilistic statements about the individual's choice.

Under the utility-maximization assumption, the probability that decision maker n chooses alternative i is

$$\begin{aligned} P_{ni} &= Prob(U_{ni} > U_{nj}, \forall j \neq i) \\ &= Prob(V_{ni} + \varepsilon_{ni} > V_{nj} + \varepsilon_{nj}, \forall j \neq i) \\ &= Prob(\varepsilon_{nj} - \varepsilon_{ni} < V_{ni} + V_{nj}, \forall j \neq i) \end{aligned} \tag{4.5}$$

This probability is a cumulative distribution, and it can be rewritten using $f(\varepsilon_n)$ as

$$P_{ni} = \int_{\varepsilon} I(\varepsilon_n < V_{ni} - V_{nj},\, j \neq i) f(\varepsilon_n) d\varepsilon_n, \tag{4.6}$$

where $\varepsilon_n = \varepsilon_{nj} - \varepsilon_{ni}$ and $I(.)$ is an indicator function which is equal to 1 when its argument is satisfied, zero otherwise. This is a multidimensional integral over the density of the unobserved portion of utility, $f(\varepsilon_n)$. Different discrete choice models are obtained from different specifications of this density, that is, from different assumptions about the distribution of the unobserved portion of utility. Between them, the Generalized Extreme Value (GEV) models are a family of models widely used in literature, since they usually provide a closed form solution for the choice probability integral (4.6).

GEV models constitute a large class of models, whose unifying attribute is that the unobserved portions of utility for all alternatives are jointly distributed as a generalized extreme value. This distribution allows for correlations over alternatives. The general expression of the GEV choice probability for a given individual to choose alternative i, given a choice set C with J alternatives, is as follows:

$$P_i = \frac{e^{V_i} G_i}{G(y_1, ..., y_J)} \tag{4.7}$$

where $y_i = e^{V_i}$ and $G_i = \frac{\partial G}{\partial y_i}$. For notational simplicity, we have omitted the subscript n denoting the decision-maker. The function G is called *generating function* and it captures the correlation patterns between the alternatives. Details about the mathematical properties of G are reported in [39], as well as algebra that obtains (4.7) from (4.6).

4.2.3 Multinomial Logit

Several GEV models can be derived from Equation 4.7, through different specifications of the generating function. In this work we use a Multi-

nomial Logit Model (MNL), which is largely the simplest and most used discrete choice model in literature. The logit model is obtained by assuming that each ε_{nj} is independently, identically distributed extreme value (Gumbel distribution[3]). In this case, the following G function is assumed, which implies no correlations between the alternatives:

$$G(y_1, ..., y_J) = \sum_{j \in C} y_j \tag{4.8}$$

Inserting this G and its first derivative G_i into (4.7), the resulting choice probability for the MNL is

$$P_{ni} = \frac{e^{V_{ni}}}{\sum_{j \in C} e^{V_{ni}}} \tag{4.9}$$

It is often reasonable to specify the observed part of utility to be linear in parameters with a constant: $V_{nj} = x'_{nj}\beta + \alpha_j \, \forall j$, where x_{nj} is a vector of variables that relate to alternative j as faced by individual n, β are coefficients of these variables, and α_j is a constant that is specific to alternative j . The *alternative-specific constant* (ASC) for an alternative captures the average effect on utility of all factors that are not included in the model. When alternative-specific constants are included, the unobserved portion of utility, ε_{nj}, has zero mean by construction. It is reasonable, therefore, to include a constant in V_{nj} for each alternative.

[3]The **Gumbel distribution** is a special case of the Fisher-Tippett distribution, where $\mu = 0$ and $\beta = 1$. The cumulative distribution function is

$$F(\varepsilon_{nj}) = \exp(-\exp(-\varepsilon_{nj}))$$

and the probability density function

$$f(\varepsilon_{nj}) = \exp(-\varepsilon_{nj}) \exp(-\exp(-\varepsilon_{nj})).$$

The variance of this distribution is $\pi^2/6$ and the mean is γ, the Euler-Mascheroni constant. Algebra that obtains (4.9) from Gumbel distribution can be found in [39].

If representative utility is specified to be linear in parameters, $V_{nj} = \beta' x_{nj}$, with $\alpha_j = 0$ in this case, the (4.9) becomes:

$$P_{ni} = \frac{e^{\beta' x_{ni}}}{\sum_{j \in C} e^{\beta' x_{nj}}} \tag{4.10}$$

where x_{nj} is a vector of observed variables relating to alternative j.

The logit probabilities exhibit several desirable properties. The first one is that $P_{ni} \in [0, 1]$, approaching to one when V_{ni} rises, and to zero when V_{ni} decreases, with $V_{nj}\, \forall j \neq i$ held constant. The second property is that the choice probabilities for all alternatives sum to one. Thirdly, McFadden [31] demonstrated that the log-likelihood function with choice probabilities given in (4.10) is globally concave in parameters β, which helps in numerical maximization procedures.

The relation of the logit probability to representative utility is sigmoid, or S-shaped, as shown in Figure 4.3.

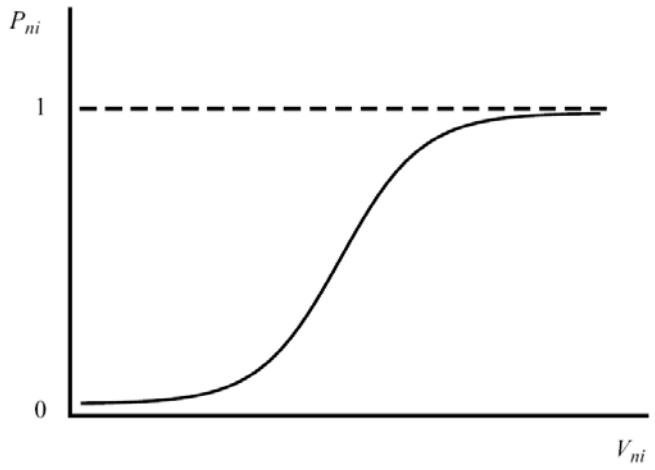


Figure 4.3: *Graph of logit curve.*

We can observe that when the representative utility of an alternative is very low compared with the other ones, a small increase in the utility of the alternative has little effect on the probability of its being chosen. Similarly, if one alternative is far superior to the others in observed attributes, a further increase in its representative utility has little effect

on the choice probability. On the contrary, when the probability is close to 0.5, small improvements in the representative utility of an alternative bring a large increase in the corresponding choice probability.

4.2.4 Specific models

As stated earlier, different choice models are derived under different specifications of the density of unobserved factors, $f(\varepsilon)$, like logit, GEV, probit, and mixed logit. A quick preview of these models is useful at this point, to show what distribution is assumed for each model, and what is the motivation for these different assumptions. A more exhaustive description of each of these DCMs can be found in [39].

We have already discussed *logit* in Section 4.2.3. Anyway, we should add some considerations about the fact that the model is derived under the assumption that ε_{ni} is iid extreme value for all i. The critical part of the assumption is that the unobserved factors are uncorrelated over alternatives, as well as having the same variance for all alternatives. This assumption of independence provides a very convenient form for the choice probability. However, it is restrictive, so it can be inappropriate in some situations. The development of other models has arisen largely to avoid the independence assumption within a logit.

Generalized extreme-value models are based, as the name implies, on a generalization of the extreme-value distribution. The generalization can take many forms, but the common element is that it allows correlation in unobserved factors over alternatives and collapses to the logit model when this correlation is zero. Depending on the type of GEV model, the correlations can be more or less flexible. For example, a comparatively simple GEV model places the alternatives into several groups called nests, with unobserved factors having the same correlation for all alternatives within a nest and no correlation for alternatives in different nests (*nested logit*). More complex forms allow essentially any pattern of correlation.

GEV models usually have closed forms for the choice probabilities.

Probits are based on the assumption that the unobserved factors are distributed jointly normal: $\varepsilon'_v = \langle \varepsilon_{n1}, \ldots, \varepsilon_{n1} \rangle \sim N(0, \Omega)$. With full covariance matrix Ω, any pattern of correlation and heteroskedasticity can be accommodated. The flexibility of the probit model in handling correlations over alternatives and time is its main advantage. Its only functional limitation arises from its reliance on the normal distribution.

Finally, *mixed logit* allows the unobserved factors to follow any distribution. The defining characteristic of a mixed logit is that the unobserved factors can be decomposed into a part that contains all the correlation and heteroskedasticity, and another part that is iid extreme value. The first part can follow any distribution, including non-normal distributions.

4.2.5 Some Considerations About DCMs

Several aspects of the behavioural decision process affect the specification and estimation of any discrete choice model. In particular, as stated by Train in [39], two aspects must be remembered: "Only differences in utility matter" and "The scale of utility is arbitrary". In the following we will explain what these statements mean and how they apply in models.

Consider the first statement. This means that the absolute level of utility is irrelevant to both the decision maker's behaviour and the researcher's model. In fact, if a constant k is added to the utility of all alternatives, the alternative with the highest utility does not change. So, if the most attractive alternative is i, the decision maker chooses the same alternative with U_{ni} as with $U_{ni} + k$ for any constant k.
The same thing happens if we consider the researcher's perspective. The choice probability is

$$P_{ni} = Prob(U_{ni} > U_{nj}, \forall j \neq i) = Prob(U_{ni} - U_{nj} > 0, \forall j \neq i)$$

which depends only on the difference in utility, not on its absolute level. When utility is decomposed into the observed and unobserved parts,

Equation (4.5) expresses the choice probability as

$$Prob(\varepsilon_{nj} - \varepsilon_{ni} < V_{ni} + V_{nj}, \forall j \neq i),$$

which also depends only on differences.

The main implication of this issue for the identification and specification of discrete choice models is that the only parameters that can be estimated (i.e., are identified) are those that capture differences across alternatives. Consider, for example, a model with alternative-specific constants. Any model with the same difference in constants will be equivalent, that is, will result in the same choice probabilities. This means that infinite values can be valid, so it is impossible to estimate the constants themselves, but only their difference. It is therefore necessary to normalize the absolute levels of the constants. A common approach to do it is to set one of the constants to zero. It is irrelevant which one is normalized: the other constants are interpreted as being relative to whichever one is set to zero. The same issue affects the way that socio-demographic variables enter a model. Attributes of the alternatives generally vary over alternatives. Attributes of the decision maker, instead, do not vary over alternatives. They can only enter the model if they are specified in ways that create differences in utility over alternatives.

Just as adding a constant to the utility of all alternatives does not change the decision maker's choice, neither does multiplying each alternative's utility by a constant. The alternative with the highest utility is the same no matter how utility is scaled. The model $U^0_{nj} = V_{nj} + \varepsilon_{nj} \forall j$ is equivalent to $U^1_{nj} = \lambda V_{nj} + \lambda \varepsilon_{nj} \forall j$ for any $\lambda > 0$. To take account of this fact, the researcher must normalize the scale of utility. The standard way to normalize the scale of utility is to normalize the variance of the error terms.

CHAPTER

FIVE

DCMs for Facial Expression Recognition

A typical automatic system for the recognition of facial expressions is based on a representation of the expression, learned from a training set of pre-selected meaningful features. In the learning process, an expert (or a group of experts) is asked to associate labels to the training samples. In this case, the expert should associate each image in the training set to one of the nine expressions (labels) we are dealing with. We consider an "expert" someone who has a *strong knowledge* of the problem, in order to ensure the correctness of what we are trying to learn. In the issue of expression evaluation, every single human can be considered as an expert. For this reason, a facial expressions evaluation survey (described in Section 5.1) has been proposed on the web, in order to obtain a set of images directly labelled by experts from all over the world. Once the labels have been assigned, we want to understand which factors affected the choice, regarding both image features and experts' characteristics. For this purpose, the most appropriate approach turned out to be a discrete choice analysis, since we face a choice situation with a finite and discrete set of alternatives (Section 5.2). The participant represents the decision maker. Hence, for simplicity, in the following we will use again the rule introduced in Section 4.2 of considering the participant as

masculine.

5.1 Facial Expressions Evaluation Survey

The facial expressions evaluation survey is born, in August 2006, in order to find a way to directly get experts' knowledge about the problem of expression recognition and to build a "common sense knowledge".

The ultimate aim of the survey is to collect a dataset created by a population of real human observers, from all around the world, doing different jobs, having different cultural backgrounds, ages and gender, belonging to different ethnic groups, doing the survey from different places (work, home, on travel, etc.). This heterogeneity in the respondent population will give us the opportunity to investigate (part of) the human factors which play different roles in the perception of human expressions. At the same time, we will be able to understand which facial parts are important and what their impact is on the expression recognition task performed by different people.

Finally, the analysis of the survey data will be able to provide insights for Human-Computer Interaction applications. Indeed, any prior model built on real data can be employed in order to improve the design of an automatic human expression recognition system.

5.1.1 Face Images

The images used in the survey comes from the Cohn-Kanade AU-Coded Facial Expression Database [24]. The database consists of expression sequences of subjects, starting from a neutral expression and ending most of the time in the peak of the facial expression. Subjects in the released portion of the Cohn-Kanade AU-Coded Facial Expression Database are 104 university students enrolled in introductory psychology classes. They ranged in age from 18 to 30 years. Sixty-five percent were female, fifteen percent were African-American, and three percent were Asian or Latino.

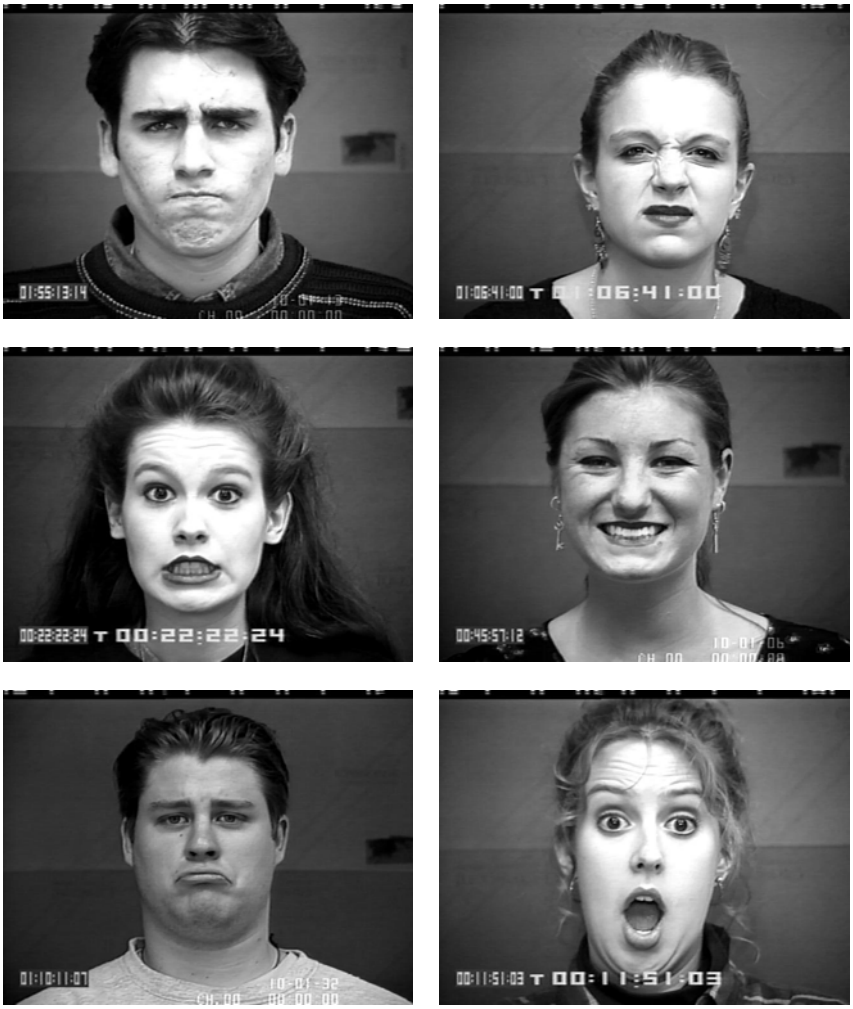


Figure 5.1: *Examples of faces in the Cohn-Kanade Database.*

The sequences are obtained with a Panasonic WV3230 camera connected to a Panasonic S-VHS AG-7500 video recorder with a Horita synchronized time-code generator. The camera was located directly in front of the sitting subject.

Subjects were instructed by an experimenter to perform a series of 23 facial displays that included single action units (e.g., AU 12, or lip corners pulled obliquely) and combinations of action units (e.g., AU 1+2, or inner and outer brows raised). Before performing each display, an experimenter described and modeled the desired display. Six of the displays were based on descriptions of prototypic emotions (i.e, happiness, anger, fear, disgust, sadness and surprise).

The subset of Cohn-Kanade Database used for the survey consists of 1274 images, extracted from the sequences of prototypic emotions of ten subjects (eight women and two men) who gave the consent for publication.

5.1.2 On-line Survey

The web page to participate in the survey is the following:

`http://lts5www.epfl.ch/face`

The survey is available in three languages (English, Italian and French), and the participant can choose the one he prefers. At the beginning of the survey and only once, the participant has to create a new account and insert a few personal information, as shown in Figure 5.2. The socio-economics fields are important in order to segment the labeller population based on different background knowledge, age, occupation and education. The ethnic group is relevant for us to investigate the choice behaviour of people when faced to images of individuals belonging to the same or

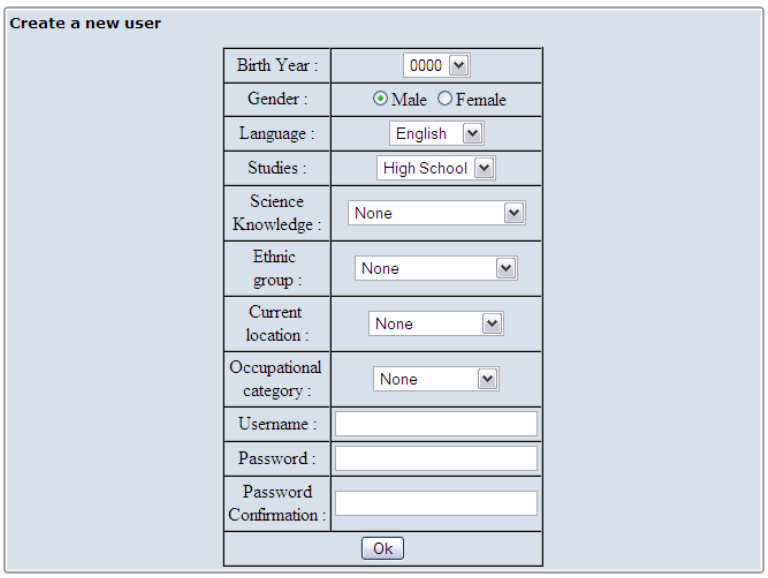


Figure 5.2: *Socio-economic form.*

to another ethnic group. The complete list and description of the socio-economics characteristics is reported in Table 5.1. The user can guarantee his own privacy choosing freely his own username and password. The data are treated confidentially and only for scientific purposes. Anyway, most of the fields include a "None" option for those responders who don't want to answer.

Once logged in his account, the participant has to specify the place where he is (home, work or other) and to choose the number of images he wants to annotate in the current survey. Then the participant can start the annotation process by clicking on the "Start the survey" button. The annotation process consists in associating an expression label to

Figure 5.3: *Image annotation interface.*

each image that will be proposed. The label must be chosen within a set of nine alternatives: Happiness, Surprise, Fear, Disgust, Sadness, Anger, Neutral, Other, I don't know. In the list of the available expressions we included, in addition to the seven prototypic emotions, the "I don't

Variable	**Description**
UserID	Unique identifier for each participant.
UserGender	1 if male, 0 otherwise
UserBirthDate	Age in years
UserOccupation	Occupation (00 = None, 01 = Medical, 02 = Educational, 03 = Management, 04 = Scientific, 05 = Engineering, 06 = Technical, 07 = Rural, 08 = Other)
UserFormation	Education (04 = High School, 05 = University, 06 = PhD, 07 = Other)
UserEthnic	Ethnic (00 = None, 01 = White, 02 = Black, 03 = Asian, 04 = Mixed White-Black, 05 = Mixed White-Asian, 06 = Mixed Asian-Black , 07 = Other)
UserRegion	Continent partecipant belongs to (00 = None, 01 = Africa, 02 = Antartica, 03 = Asia, 04 = Australia, 05 = Europe, 06 = North America, 07 = South America)
UserScienceKW	Participant scientific knowledge (00 = None, 02 = Behavioral Science, , 03 = Social Science, 04 = Computer Science, 05 = Cognitive Science, 06 = Otehr)
UserLanguage	Web Interface chosen language (01 = French, 02 = English, 03 = Italian)
UserLocation	Participant location (01 = Home, 02 = Work, 03 = Other)

Table 5.1: *Description of Participant Socio-Economic Variables.*

know" and "Other" options, to be used when the image seems ambiguous to the participant. A simple and intuitive interface has been designed in order to facilitate the labelling procedure (see Fig. 5.3). After the participant has selected one of the available options, he can click on the right arrow in order to validate the current choice and pass to the next image. The survey can be stopped whenever the participant wants by logging off and restarted from the first unlabelled image at his next login. At the end of the survey, the participant can validate the whole survey by clicking on the "Validate survey" button. Each participant can take part to the survey as many times as he wants.

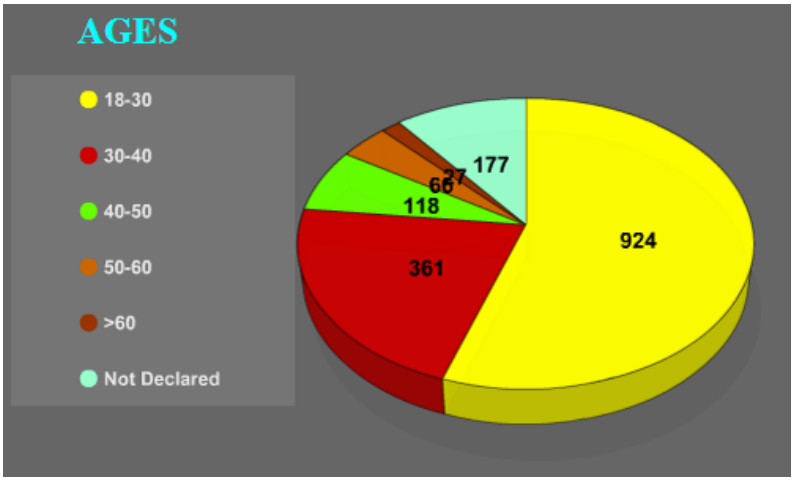
5.1.3 Collected Data

To date, 1700 participants have taken part in the survey and around 39000 images have been annotated. In Figures 5.4-5.5 we report some statistics on the participants, also available at

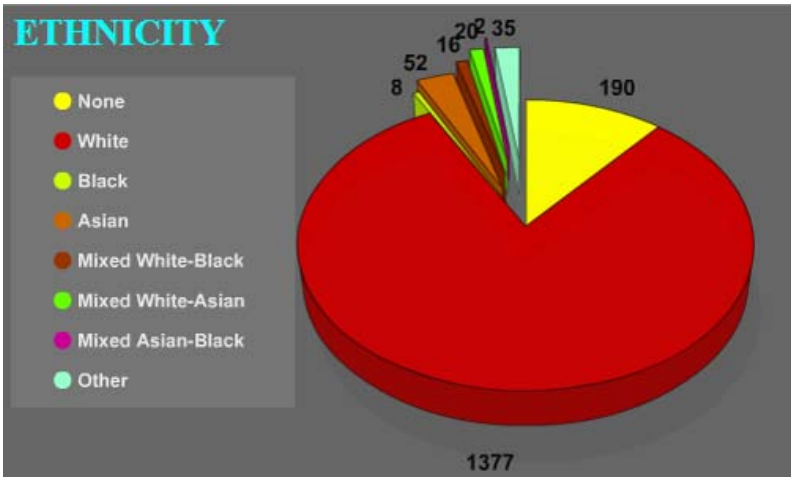
`http://itswww.epfl.ch/~sorci/SurveyStat.php`

We can observe that the majority of participants lives in Europe and the "White" group is the most numerous one. However, we have representatives from all the populated continents and from all the ethnic groups. Concerning participants' cultural background, almost half of the sample has a University Education and all the "Occupation" categories are quite well represented. Computer science and other not listed science branches are the two biggest groups for "Scientific Knowledge". Anyway, a good number of participants with social, behavioural and cognitive science background took part in the survey as well.

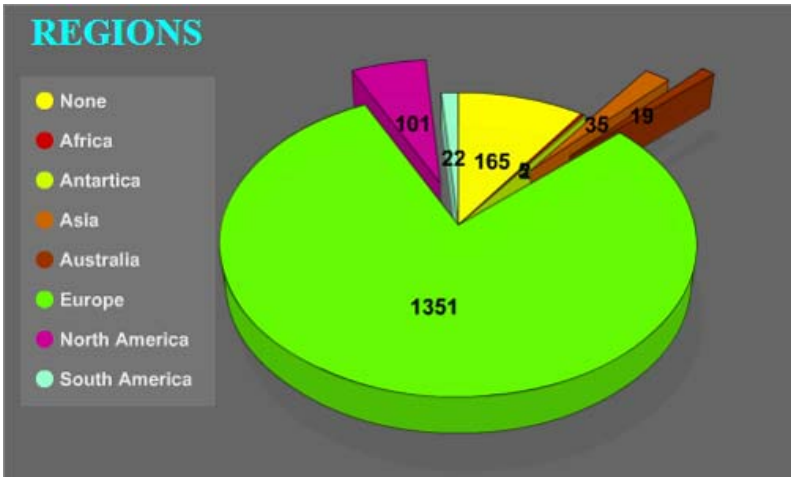
We can conclude that the analysed sample of population is fairly heterogeneous, because all the categories are well represented. The collected information is therefore enough to understand how and which of these human factors play a role in the perception of human expressions.



(a)

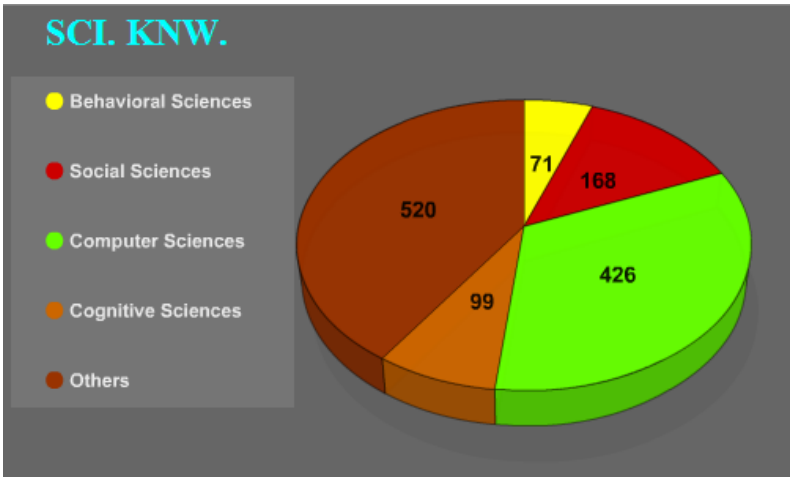


(b)

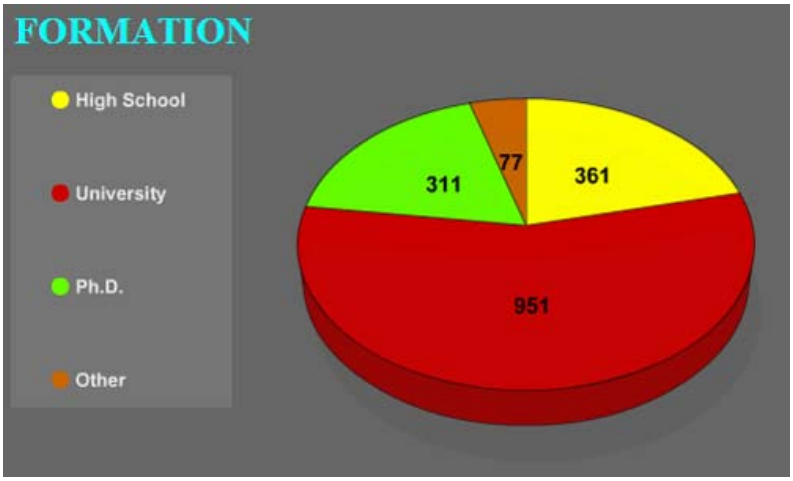


(c)

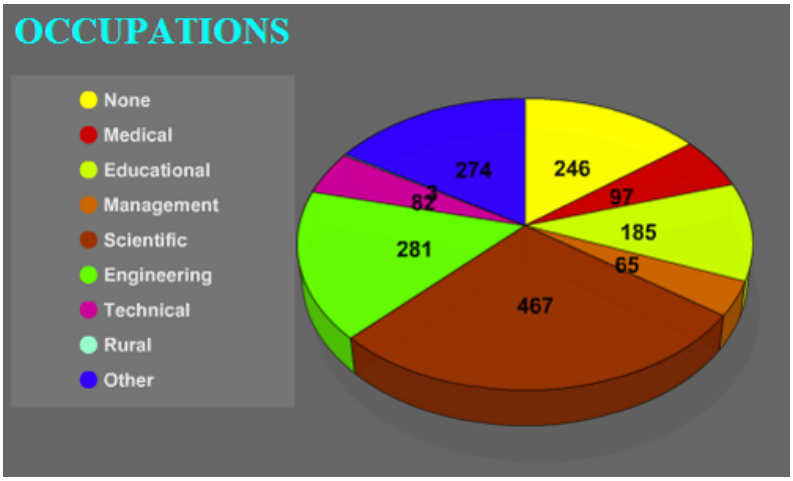
Figure 5.4: *Survey statistics: age, ethnic group, region.*



(a)



(b)



(c)

Figure 5.5: *Survey statistics: scientific knowledge, formation, occupation.*

5.2 Modelling Facial Expressions

Our approach to facial expression understanding relies on the linguistic descriptions of facial expressions from psychological view (the labels). We want to investigate the labelling procedure in order to identify the factors that people evaluate when they associate an emotional expression to a particular face.

We hypothesize that the choice depends on two kind of factors:

- Facial motion cues,
- Characteristics of the decision maker.

In order to demonstrate our assumption, a stepwise approach is adopted: we propose four discrete choice models, where new attributes are introduced at each step. In the first three models, only attributes relative to facial features are combined, while in the fourth one also the socio-economic features of the decision maker are included. An utility function is defined for each of the nine available expressions. The utility functions are specified using a linear-in-parameters form. The choice for a linear form is based purely on simplicity considerations, for reducing the number of parameters in the estimation process. The general form of the utilities is given by:

$$V_i = \alpha_i + \sum_{k=1}^{K} I_{ki}\beta_{ki}x_k \tag{5.1}$$

where $i = 1, ..., C$ with $C = 9$ is the number of expressions, K is the number of attributes included in the model, I_{ki} is an activation function equal to 1 if the k-th attribute is included in the utility for expression i and 0 otherwise and α_i is an alternative-specific constant. The α_i coefficients represent the average value of the unobserved part of the corresponding utility and one of them has to be normalized to 0, in order to be consistent with DCM theory (see [5]). In our case, we normalize with respect

to the neutral expression. Each attribute k in each utility i is weighted by an unknown deterministic coefficient, β_{ki}, that has to be estimated. In order to simplify the model, only the alternative-specific constant is included in the utility function for the "I don't know" alternative.

In the following paragraphs, we will illustrate in details all the proposed models, from the simplest to the more complex one, by explaining, at each step, the reasons for each included attribute. In the utility functions retained for each of the four models, only features corresponding to statistically significant parameters (t-test statistic against the zero value) are reported. Note that the proposed utility expressions are the result of a strong iterative process, where several hypothesis have been tested and validated.
Before building and estimating the models, we apply an outlier analysis to the data, in order to partially clean our dataset. Then, the whole dataset is split in two parts: the 80 percent of the observations is used for training and testing the models and the remaining 20 percent is used for validation.

5.2.1 Facial Motion Cues

Facial expressions represent a visible consequence of facial muscle activity. It is generally believed that at least the six basic expressions can be described linguistically using Ekman's AUs. Likewise, we adapt the AU-coded descriptions of facial expressions in the FACS to describe these expressions as well as the "Other" option. In Table 5.2, which is directly adapted from [16], we illustrate the facial AUs pertaining to the different expressions.

By drawing on the work of Zhang and Ji [44], we group AUs of facial expressions as primary AUs and auxiliary AUs. By the primary AUs, we mean those AUs or AU combinations that are strongly pertinent to one

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

Table 5.2: *A list of AUs related to six facial expressions.*

of the six expressions without ambiguities. In contrast, an auxiliary AU is the one that can be only additively combined with primary AUs to provide supplementary support to facial expression classification. Consequently, a facial expression contains primary AUs and auxiliary AUs. For example, AU9 (Nose Wrinkler) can be directly associated with an expression of disgust, while it is ambiguous to associate a single AU17 (Chin Raiser) with a disgust expression. When AU9 and AU17 appear simultaneously, the classification of this AU combination to a disgust expression then becomes more certain. Hence, AU9 is a primary AU of disgust, while AU17 is one of its auxiliary AUs.

Additionally, changes in facial transient features, such as wrinkles and furrows, also provide support cues to infer certain expressions. For example, a smiling face may lengthen and deepen the horizontal wrinkles on the eye outer canthi, while the vertical furrows between the eye brows tend to be intensive when a person expresses strong anger. The appearance of facial transient features is influenced by not only the interpersonal variation, but also by the age. Some of these transient features may become permanent due to age. We only consider the changes of these features as support evidence, which may partially contribute to the

identification of facial expressions. Table 5.3 gives a summary of primary AUs, auxiliary AUs and transient features associated with the six basic facial expressions.

Emotional	Primary Visual Cues					Auxiliary Visual Cues					
Category	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	

Table 5.3: *The association of six emotional expressions to AUs, AU combinations, and Transient Features.*

FACS is a human observer based system, designed to detect subtle changes in facial features. In order to automatically extract those features, we have to quantitatively code AUs into facial descriptors that can be extracted directly from the face image. Figure 5.6a presents facial geometrical relationships and furrow regions. Correct association between changes of the feature points and the corresponding AUs is crucial for accurate facial expression interpretation. Therefore, we manually establish the association of the AUs and the movements of the facial feature points as shown in Table 5.4, so that the facial visual changes are automatically measurable on imagery.

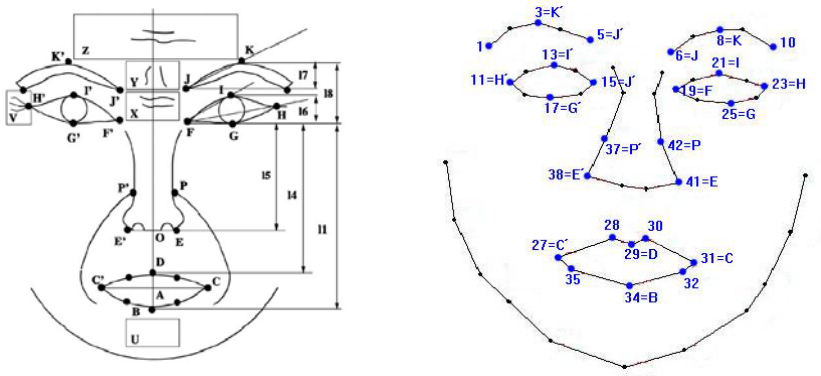
The AUs descriptors are thus obtained by measuring distances and angles between the appropriate points in the shape model of the face obtained from AAM, described in Section 4.1 (for example, $\overline{JF}$ is the distance between points 6 and 19). All the correspondences are reported in Table 5.5, where each measure is associated also with the name used in the utility functions.

By resting upon the relationships between expressions and facial motion cues shown in Table 5.3, we include in each utility function the action units related to the pertaining expression. In particular, each action unit

AUs	Facial Visual Cues	Visual Channel(s)
AU1	$\angle FHJ$, $\overline{JF}$ increased OR	
	$\overline{JF}$ increased, $l8$ nonincreased	Brow
AU2	$l8$ increased and $\overline{JF}$ nonincreased	
	furrow in $\square Z$ increased	Brow, Wrinkler
AU4	$l8$, $\overline{FJ}$, $\overline{JJ'}$, $\overline{FP}$, $\overline{F'P'}$ decreased,	
	$\angle HFI$ increased and	
	wrinkle in $\square Y$ increased	Brow, Wrinkler
AU5	$l6$, $\overline{JF}$ and $\overline{JJ'}$ increased	Lid
AU6	nasolabial furrow presence and	Nasolabial,
	wrinkle in $\square V$ increased	Wrinkler
AU7	$\angle HFI$ nonincreased and	
	$\angle HGF$ increased	Lid
AU9	wrinkle increased in $\square X$	
	nasolabial furrow presence OR	Wrinkler,
	$\overline{PF}$, $\overline{FJ}$ decreased	Nasolabial
AU10	$l4$ decreased and $\lvert\overline{FC} - \overline{F'C'}\rvert$	
	increased, nasolabial presence OR	Lip,
	$\overline{OD}$ decreased, $\overline{DB}$, $\overline{C'C}$ increased	Nasolabial
AU12	$\overline{FC}$, $\overline{F'C'}$ decreased,	
	$\overline{CC'}$ increased, $\overline{GI}$ nonincreased	LipCorner
AU15	$\overline{FC}$, $\overline{F'C'}$, $\overline{CC'}$ increased	LipCorner
AU16	$\overline{OD}$ non-change, $\overline{DB}$ decreased	Lip
AU17	$\overline{OB}$ decreased and	
	wrinkle in $\square U$ presence	Chin, Wrinkler
AU20	$\overline{CC'}$ increased and	
	$\overline{FC}$, $\overline{F'C'}$ nonchange	LipCorner
AU23	$\overline{DB}$, $\overline{CC'}$ decreased	Lip
AU24	$\overline{DB}$ decreased, $\overline{CC'}$ nonchange	Lip
AU25	$\overline{DB}$ increased, $\overline{DB} < T_1$,	
	$\overline{CC'}$ nonincreased	Mouth
AU26	$T_1 < \overline{DB} < T_2$, $\overline{CC'}$ nonincreased	Mouth
AU27	$\overline{DB} > T_2$, $\overline{CC'}$ nonincreased	Mouth

Note: T_1 and T_2 are predefined thresholds of the mouth vertical span, which are determined separately for each person in the neutral state.

Table 5.4: *Motion-Based feature descriptors for AUs.*



(a) (b)

Figure 5.6: *Masks for defining facial descriptors: (a) the geometrical relationship of facial feature points, where the rectangles represent the regions of furrows and wrinkles, and (b) the corresponding points on the face mask obtained with the AAM.*

is defined as the linear combination of the corresponding facial descriptors, according to Table 5.4.

We propose three different models. Each model is obtained by adding a new set of features to the previous one. In the first and most simple model only the primary AUs are included. Then, the complexity is increased by introducing firstly the auxiliary AUs and secondly the transient features.

Model Specification with Action Units

In the first model, expressions are described only by primary visual cues. 67 parameters have to be estimated. The general expression of the deterministic utility functions for each alternative in this model specification is the following:

$$V_i = \alpha_i + \sum_{m=1}^{M} I_{mi}\beta_{mi}prim_m \tag{5.2}$$

In the second model, instead, where also the auxiliary AUs are considered, the unknown parameters are 82, and the utility functions assume

Measures on mask 5.6a	Measures on mask 5.6b	Attribute name
$\overline{JJ'}$	6-5	brow_dist
$\overline{JF}$	6-19	browEye2
$l8$	8-25	browEye3
$\overline{GI} \equiv l6$	25-21	eye_height
$\overline{PF}$	42-19	eyeNose_dist
$\overline{FC}$	19-31	eyeMouth_dist
$l4$	25-29	eyeMouth_dist2
$\overline{OD}$	$(\frac{39+40}{2})$-29	mouthNose_dist
$\overline{OB}$	$(\frac{39+40}{2})$-33	mouthNose_dist2
$\overline{DB}$	29-33	mouth_height
$\overline{C'C}$	31-27	mouth_width
$\angle FHJ$	angle between 19, 23 and 6	eyeBrow_angle
$\angle HFI$	angle between 23, 19 and 21	eye_angle
$\angle HGF$	angle between 23, 25 and 19	eye_angle2

Table 5.5: *Correspondences between measures on masks 5.6a and 5.6b.*

the form:

$$V_i = \alpha_i + \sum_{m=1}^{M} I_{mi}\beta_{mi}prim_m + \sum_{n=1}^{N} I_{ni}\beta_{ni}aux_n \tag{5.3}$$

A summary of the attributes included in these two models is provided in Table 5.6. Each column corresponds to an expression, while the rows are the facial descriptors of the action units included in each utility. The symbol '$\star$' corresponds to the attributes $prim_m$ and the '$\bullet$' to aux_n.

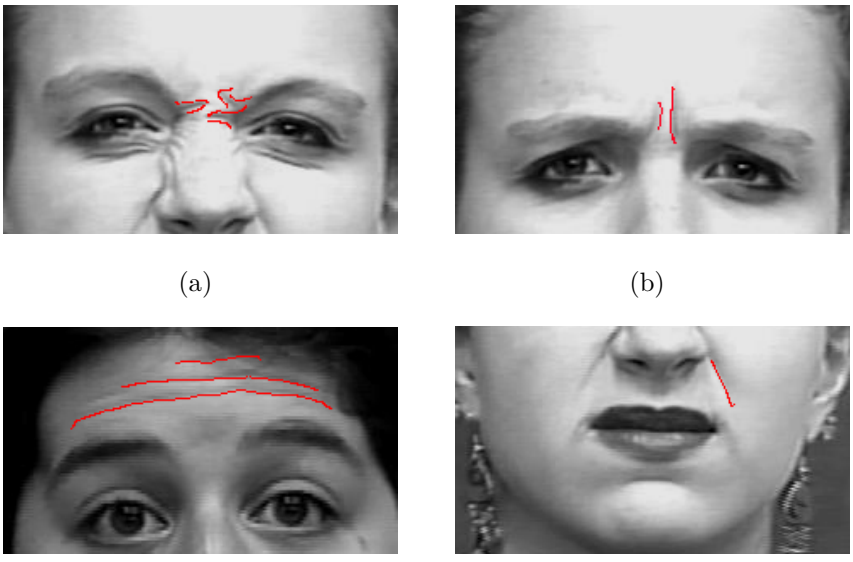
Model Specification with Transient Features

The activation of facial muscles also produces transient wrinkles and furrows perpendicular to the muscular motion direction in certain face regions. For example, raising the outer brows wrinkles up one's frontal eminence and raising the cheeks may deepen the nasolabial fold and de-

	Ang	**Disg**	**Fear**	**Happ**	**Oth**	**Sadn**	**Surp**
brow_dist	⋆		⋆			•	
browEye2_l	⋆				⋆	⋆	⋆
browEye2_r	⋆	⋆	⋆			⋆	⋆
browEye3_l	⋆						⋆
eye_angle_l	⋆		⋆			•	
eye_angle_r	⋆		⋆			•	
eye_angle2_l			⋆				
eye_angle2_r			⋆			•	
eyeBrow_angle_l			⋆		⋆	⋆	
eyeBrow_angle_r			⋆		⋆	⋆	⋆
eyeMouth_dist_l			⋆	⋆	⋆	⋆	
eyeMouth_dist_r			⋆	⋆	⋆	⋆	
eyeMouth_dist2_l		⋆			⋆		
eyeMouth_dist2_r		⋆			⋆		
eyeNose_dist_l	⋆	⋆	•		⋆	•	
eyeNose_dist_r	⋆	⋆	•		⋆	•	
eye_height_l			⋆	⋆			•
eye_height_r			⋆	⋆			•
mouth_height	⋆	⋆	•	•		•	⋆
mouth_width	⋆	⋆	⋆	⋆	⋆	⋆	
mouthNose_dist		⋆		•			
mouthNose_dist2	•					⋆	

Table 5.6: *Facial descriptors related to primary (⋆) and auxiliary (•) AUs included in each utility function. "Neutral" and "I don't know" are omitted, since we included only the ASC in their utilities.*

form its initial shape. While one's forehead, nasolabial region, and eye corners may be furrowed with age and become permanent facial features, more or less, facial muscle movement causes changes in their appearance, such as deepening or lengthening. The transient features can therefore provide additional visual cues to support the recognition of facial expressions. The regions of facial wrinkles and furrows are indicated by rectangles in Fig. 5.6a. The change of wrinkles in the region $\square X$ is directly related to AU9 (Nose Wrinkler), while others merely enhance the identification of AUs, e.g., furrows in the regions $\square Z$, $\square Y$, $\square V$, $\square U$ provide diagnostic information for the identification of AU2 (Outer Brow Raiser), AU4 (Brow Lowerer), AU6 (Cheek Raiser), and AU17 (Chin Raiser), respectively.



(a) (b)

(c) (d)

Figure 5.7: *Transient feature detection: (a) vertical furrows between brows, (b) horizontal wrinkles between eyes, (c) horizontal wrinkles on the forehead, and (d) nasolabial fold.*

The presence of furrows and wrinkles on an observed face image can be determined by edge feature analysis in the areas where transient features appear. We use firstly an edge detection with embedded confidence,

proposed by Meer and Georgescu [32], where the widely used three-step edge detection procedure: gradient estimation, nonmaxima suppression, hysteresis thresholding; is generalized to include the information provided by the confidence measure. Secondly, we check the direction of the extracted edges to filter the noise. In fact, as we can see in Figure 5.6a, wrinkles in regions $\square Z$ and $\square X$ should be mostly horizontal, and mostly vertical in region $\square Y$. If there are a few hairs on the forehead, this technique can still work since hairs are mostly represented by vertical edges while forehead wrinkles are mostly horizontal edges. Figure 5.7 shows examples of transient feature detection.

Then, we choose to consider the ratio between edge pixels (wrinkles) and background pixels (skin) to represent in the model wrinkles in regions $\square X$ and $\square Y$. Furthermore, we use two categorical variables for nasolabial fold and furrows on the forehead: these variables are equal to 1 if the corresponding furrows are present in the face, and zero otherwise.
The utility functions for the third model are therefore obtained by adding these attributes to the ones describing the action units, resulting in a final model with 93 unknown parameters:

$$V_i = \alpha_i + \sum_{m=1}^{M} I_{mi}\beta_{mi}prim_m + \sum_{n=1}^{N} I_{ni}\beta_{ni}aux_n + \sum_{l=1}^{L} I_{li}\beta_{li}tr_l \qquad (5.4)$$

Table 5.7 shows the transient features (tr_l) included in the utility function of each facial expression.

5.2.2 Socio-Economic Characteristics

The idea that the choice made by the decision maker depends also on some of his personal characteristics is one of the new aspects tackled in the present work. To identify the factors involved in the decision process, we analyse the decision makers' socio-economic features. All these attributes are represented in the utility functions by categorical variables.

	Ang	**Disg**	**Fear**	**Happ**	**Oth**	**Sadn**	**Surp**
browwrinkle_ratio	⋆	⋆			⋆	⋆	
forehead_presence	⋆		⋆		⋆		⋆
nasalwrinkle_ratio		⋆			⋆		
nasolabial		⋆					

Table 5.7: *Transient features included in each utility function. "Neutral" and "I don't know" are omitted, since we included only the ASC in their utilities.*

In other words, we define a variable for each subset (or group of subsets) within each considered socio-economic aspect. A socio-economic aspect (or category) is, for instance, the "Formation", and its subsets are "High School", "University", "PhD", "Other" (see Table 5.1). The categorical variable is equal to 1 if the participant belongs to the corresponding subset, and zero otherwise. For example, $reg5$ will be 1 for European decision makers and zero for all the others. For correctly estimating and interpreting the model, we need to define a reference subset for each socio-economic category. Where available, we choose the "None" option for this purpose. In the other cases, we consider "English" as the reference language and "Other" as the reference subset for formation and location.

In Table 5.8 we report socio-economic attributes included in the utility function of each expression in addition to facial descriptors. The utility functions for this model are defined as follows:

$$V_i = \alpha_i + \sum_{m=1}^{M} I_{mi}\beta_{mi} prim_m + \sum_{n=1}^{N} I_{ni}\beta_{ni} aux_n + \sum_{l=1}^{L} I_{li}\beta_{li} tr_l + \sum_{h=1}^{H} I_{hi}\beta_{hi} socec_h \tag{5.5}$$

In this complete model, where both attributes of the face and of the decision maker are considered, 177 parameters have to be estimated. The model reported here is the one with the most significant parameters.

	Ang	**Disg**	**Fear**	**Happ**	**Oth**	**Sadn**	**Surp**
age1		⋆	⋆		⋆	⋆	
age2		⋆		⋆		⋆	
eth1		⋆			⋆	⋆	
eth2	⋆				⋆		
eth3		⋆	⋆	⋆	⋆	⋆	
eth4	⋆	⋆	⋆		⋆		
form1		⋆	⋆	⋆		⋆	
form2			⋆	⋆	⋆		
gender		⋆			⋆	⋆	
lan1			⋆	⋆	⋆	⋆	⋆
lan3	⋆	⋆		⋆	⋆		⋆
loc1	⋆	⋆		⋆	⋆		⋆
loc2	⋆	⋆			⋆		
reg1		⋆					
reg3		⋆			⋆		
reg4	⋆	⋆	⋆	⋆	⋆		⋆
reg5		⋆	⋆		⋆		
reg6		⋆	⋆		⋆		
reg7	⋆	⋆			⋆		
science1				⋆	⋆		
science2						⋆	⋆
science3	⋆	⋆		⋆	⋆		⋆
science4				⋆	⋆	⋆	⋆
science5	⋆			⋆	⋆	⋆	

Table 5.8: *Socio-economic attributes included in each utility function. "Neutral" and "I don't know" are omitted, since we included only the ASC in their utilities.*

CHAPTER

SIX

Results

In this chapter we will analyse the performance of the proposed models. Section 6.1 will provide some details about BIOGEME, the tool used for estimating and simulating the proposed models. In Section 6.2, we will examine the output files obtained by estimating the models with BIOGEME. This analysis will demonstrate the goodness of each of the proposed models. Last, in Section 6.3, we will compare the performance of the different models, in order to identify the best one.

6.1 A tool for estimating DCMs

BIerlaire **O**ptimization toolbox for **GE**V **M**odel **E**stimation (BIOGEME) is a freeware designed for the estimation of Binary Logit, Multinomial Logit, Nested Logit models and more complex models in the GEV family as well as mixtures of these models (e.g. Mixed Logit). With BIOGEME come two additional utilities. BioRoute helps preparing the input files for BIOGEME in the context of a route choice analysis. BioSim is designed to perform simulations with a given model. All information relative to BIOGEME is maintained at

`roso.epfl.ch/biogeme`

6.1.1 BIOGEME

BIOGEME has been developed on Linux, but two Windows version are available: one to be executed in a terminal (DOS, Cygwin, or anything else), and another with a simple graphical user interface (GUI). The GUI is designed for teaching purposes, where simple models are estimated, so it is strongly recommended to use BIOGEME in a terminal for estimating complex models.

BIOGEME is invoked in a shell under Linux, in a DOS command window or a Cygwin command window under Windows using the following statement structure

```
biogeme  mymodel  sample.dat
```

If the name of the model is *mymodel*, BIOGEME reads the following files:

- a file containing the parameters controlling the behaviour of BIOGEME (e.g., the optimization algorithm to be used): `mymodel.par`,
- a file containing the model specification: `mymodel.mod`,
- a file containing the data: `sample.dat`,
- optionally a file containing the random numbers to use if estimation is based on simulation.

The model and data files are essential, while the parameter file in general does not need to be edited. In fact, it is created with default values when BIOGEME is invoked. In this case BIOGEME will use the "BIO" optimization algorithm, but five different algorithms are available: CFSQP, DONLP2, SOLVOPT, BIO and BIOMC. Note that it is possible that each of them produces different solutions. Usually, the discrepancies are small, and due to numerical differences and various stopping criteria. Also, none of them identifies a global maximum of the likelihood function.

Therefore, it may happen that one of them is caught in a local maxima, different from local maxima found by other algorithms. We managed to use CFSQP, since it is well-suited for estimating on huge data sets models with lots of parameters. CFSQP is a C implementation of the FSQP optimization algorithm developed by E.R. Panier, A.L. Tits, J.L. Zhou, and C.T. Lawrence [27].

BIOGEME automatically generates the following output files:

- a file reporting the results of the estimation: `mymodel.rep`,
- the same file in HTML format: `mymodel.html`,
- a file containing the specification of the estimated model, in the same format as the model specification file: `mymodel.res`,
- a file containing some descriptive statistics on the data: `mymodel.sta`;

and the following files to help understanding possible problems:

- a file containing messages produced by BIOGEME during the run: `mymodel.log`,
- a file containing the values of the parameters which have been actually used by BIOGEME: `parameters.out`,
- a file containing the data stored in BIOGEME to represent the model: `model.debug`,
- a file containing the specification of the model, as it has actually been understood by BIOGEME: `specFile.debug`.

To avoid any ambiguity, BIOGEME displays the filenames it has actually used for a specific run, for instance

```
BIOGEME Input files
===================
Parameters:                 default.par
Model specification:        mymodel.mod
Sample 1 :                  sample.dat
BIOGEME Output files
====================
Estimation results:         mymodel~3.rep
Estimation results (HTML):  mymodel~3.html
Result model spec.  file:   mymodel~2.res
Sample statistics:          mymodel~1.sta
BIOGEME Debug files
===================
Log file:                   mymodel.log
Parameters debug:           parameters.out
Model debug:                model.debug
Model spec.  file debug:    __specFile.debug
```

6.1.2 BioSim

The package BioSim is invoked exactly like BIOGEME, with the same input files:

```
biosim  mymodel  sample.dat
```

but instead of performing a parameter estimation, it performs a sample enumeration. Sample enumeration performed by BioSim produces correct predicted probabilities for all model versions as long as it is not in a panel data setting. The panel data setting requires a large set of choice probability calculations and this will not be available in BioSim in the very near future.

The file `mymodel.enu` contains the result of the sample enumeration. For each observation in the sample, the following results are provided:

1. the identifier of the choice actually reported in the sample file,

2. the name of the choice actually reported in the sample file,
3. the probability given by the model for the chosen alternative,
4. for each alternative, the utility given by the model,
5. for each alternative, the probability given by the model,
6. a list of simulated choice, based on Monte-Carlo simulation using the model.

6.2 Analysis of BIOGEME output files

Once we have defined each model, by deciding the attributes to include, we estimate them with BIOGEME. The estimation of unknown parameters is based on maximum likelihood. Details can be found in [6]. The algorithms used for maximization identify local maxima of the likelihood function. We performed various runs, with different starting points (a trivial model with all parameters to zero, and the estimated value of several intermediary models). They all converged to the same solution.

As described in the previous section, BIOGEME generates some output files, which contain information about the estimation process. In particular, we will now analyse the report file, containing the results of the maximum likelihood estimation of the model, in order to understand how well the model describes the studied behaviour. For space reasons, we report in Tables 6.1-6.4 only a subset of the estimated parameters for each model. The whole tables for all the models are available in Appendix A.

6.2.1 Model Specification with Primary AUs

In our first model we made the hypothesis that only primary action units affect the choice of the decision maker. For estimation purposes, we have normalized to zero the alternative “Neutral” and the estimated

coefficients are therefore interpreted relative to it. The estimation results for this model are shown in Table 6.1. The last three columns of the table contain, respectively, the estimated value of the β parameters, the associated standard error and the t-test. A sign $(*)$ is appended if the t-test fails, according to the specified threshold. Otherwise, the parameter is significantly different from zero at least at 95 percent.

MNL estimation				
Parameter number	Parameter name	Parameter estimate	Robust standard error	Robust t statistic
1	ASC_A	$-2.09E+00$	$+3.13E-01$	$-6.68E+00$
2	ASC_D	$-7.84E-01$	$+2.87E-01$	$-2.73E+00$
3	ASC_{DK}	$-2.31E+00$	$+3.72E-02$	$-6.21E+01$
4	ASC_F	$+1.93E+00$	$+1.11E+00$	$+1.74E+00*$
5	ASC_H	$+1.04E+00$	$+3.16E-01$	$+3.28E+00$
6	ASC_N	$0.00E+00$	fixed	
7	ASC_O	$-1.54E+00$	$+3.24E-01$	$-4.74E+00$
8	ASC_{SA}	$-3.00E+00$	$+3.08E-01$	$-9.74E+00$
9	ASC_{SU}	$-4.37E+00$	$+1.91E-01$	$-2.29E+01$
10	$\beta_{brow\_dist\_A}$	$-1.31E+01$	$+2.14E+00$	$-6.12E+00$
11	$\beta_{browEye2\_r\_SU}$	$+2.24E+01$	$+5.84E+00$	$+3.83E+00$
12	$\beta_{eyeBrow\_angle\_l\_F}$	$+7.09E+00$	$+5.93E-01$	$+1.20E+01$
13	$\beta_{eyeMouth\_dist\_r\_H}$	$-6.04E+01$	$+4.06E+00$	$-1.49E+01$
14	$\beta_{eyeMouth\_dist\_r\_SA}$	$+6.98E+01$	$+4.05E+00$	$+1.72E+01$
15	$\beta_{eyeNose\_dist\_l\_O}$	$+7.90E+01$	$+5.55E+00$	$+1.42E+01$
16	$\beta_{eyeNose\_dist\_r\_O}$	$-8.59E+01$	$+6.53E+00$	$-1.32E+01$
17	$\beta_{mouth\_height\_SU}$	$+4.46E+01$	$+1.37E+00$	$+3.26E+01$
18	$\beta_{mouth\_width\_D}$	$+2.70E+01$	$+1.86E+00$	$+1.46E+01$
Summary statistics				
Number of observations = 30514				
$\mathbf{L}(0) = -67046.1$				
$\mathbf{L}(\hat{\beta}) = -47143.9$				
$\bar{\rho}^2 = 0.2958$				

Table 6.1: *Estimation results for the MNL model with primary AUs.*

Given our specification, the negative sign of the alternative-specific constants of some alternatives can be interpreted as the decision maker prefers to choose the neutral expression with respect to them, all the rest remaining constant. On the contrary, the positive sign of ASC_H indicates a preference for the alternative "Happiness" with respect to "Neutral". With regard to the coefficients of facial descriptors, we can interpret them by pointing out that each action unit is defined as a combination of increasing or decreasing measures (see Tab. 5.4). Since larger values of facial descriptors with respect to the neutral expression imply larger utility values for the corresponding expressions, we can expect a positive coefficient for measures which should increase. Vice versa, a negative sign is expected for distances and angles that are supposed to be smaller in a particular expression than in the neutral face. By comparing the obtained results with Table 5.4, we can observe that almost all the β coefficients are consistent with the proposed interpretation. For example, the coefficient $\beta_{brow\_dist\_A}$ is negative and statistically different from zero, indicating that larger distance between brows for expression "Anger" with respect to "Neutral" implies a lower utility value. The positive value of the mouth height parameter $\beta_{mouth\_height\_SU}$ is intuitive: we often associate a vertically open mouth to an expression of surprise. Moreover, the coefficient $\beta_{eyeMouth\_dist\_r\_SA}$ has a positive sign, indicating that the mouth corner dips in "Sadness" with respect to "Neutral", while the coefficient $\beta_{eyeMouth\_dist\_r\_H}$ for the same measure in "Happiness" is negative, as expected.

It is interesting to notice that for some measures the coefficients have a different sign when considering the left or the right side of the face (i.e. $\beta_{eyeNose\_dist\_l\_O}$ and $\beta_{eyeNose\_dist\_r\_O}$). This could depend on the fact that the masks generated with the AAM don't perfectly match the real landmark points of the face. Another possible explanation is that the face itself is asymmetric, even in the neutral expression or just because muscles don't move in the same way on both the side of the face. In

fact, several studies [20] measured this movement asymmetry on posed expressions, showing that lateralised actions sometimes favour the left, and sometimes the right side.

6.2.2 Model Specification with Auxiliary AUs

In Table 6.2 we report the estimation results for this second model, where also the auxiliary AUs are included in the utility functions. Very few parameters are not statistically significant for the model. The estimated values for the alternative-specific constants show that, all the rest remaining constant, there is a preference in the choice of "Happiness" with respect to all the other alternatives. In fact, only ASC_H has a positive sign. We could speculate that happiness is easier to recognise than other expressions even when its distinguishing features are only slightly present on the face. However, there might also be other viable explanations, for example "Happiness" images may be more present than the others in the analysed database. The parameter ASC_F has changed sign, but now it is significant, while in the previous model it was not significantly different from zero.

The coefficients for facial descriptors related to the auxiliary AUs are significantly different from zero and their signs are consistent with the interpretation proposed also for primary AUs. We find that decreasing or not increasing measures have negative coefficients (e.g. $\beta_{brow\_dist\_SA}$), while the positive ones are for growing distances (e.g. $\beta_{mouth\_height\_F}$). It is important to observe that parameters related to primary AUs are still significant, and their signs and magnitudes have remained almost the same as in the previous model.

To test whether this model is better than the previous one we use the likelihood ratio test (see Ben-Akiva and Lerman [5]). The likelihood ratio test (LRT) is a statistical test of the goodness-of-fit between two models. A relatively more complex model (*unrestricted model*) is compared to

a simpler model (*restricted model*) to see if it fits a particular dataset significantly better. If so, the additional parameters of the more complex model are used in subsequent analysis. The LRT is only valid if used to compare hierarchically nested models. That is, the more complex model must differ from the simple model only by the addition of one or more

MNL estimation				
Parameter number	Parameter name	Parameter estimate	Robust standard error	Robust *t statistic*
1	ASC_A	$-2.69E+00$	$+3.69E-01$	$-7.29E+00$
2	ASC_D	$-1.06E+00$	$+2.92E-01$	$-3.62E+00$
3	ASC_{DK}	$-2.31E+00$	$+3.72E-02$	$-6.21E+01$
4	ASC_F	$-4.19E+00$	$+1.18E+00$	$-3.54E+00$
5	ASC_H	$+1.02E+00$	$+3.47E-01$	$+2.95E+00$
6	ASC_N	$0.00E+00$	fixed	
7	ASC_O	$-1.82E+00$	$+3.31E-01$	$-5.50E+00$
8	ASC_{SA}	$-3.17E+00$	$+5.31E-01$	$-5.96E+00$
9	ASC_{SU}	$-4.44E+00$	$+2.13E-01$	$-2.09E+01$
10	$\beta_{brow\_dist\_A}$	$-1.41E+01$	$+2.15E+00$	$-6.56E+00$
11	$\beta_{brow\_dist\_SA}$	$-2.60E+01$	$+2.22E+00$	$-1.17E+01$
12	$\beta_{eye\_height\_r\_SU}$	$+3.71E+01$	$+4.81E+00$	$+7.71E+00$
13	$\beta_{eyeMouth\_dist\_r\_H}$	$-7.65E+01$	$+4.25E+00$	$-1.80E+01$
14	$\beta_{eyeNose\_dist\_r\_F}$	$-8.96E+01$	$+1.14E+01$	$-7.89E+00$
15	$\beta_{eyeNose\_dist\_l\_O}$	$+9.28E+01$	$+5.74E+00$	$+1.62E+01$
16	$\beta_{mouth\_height\_F}$	$+4.30E+01$	$+1.61E+00$	$+2.67E+01$
17	$\beta_{mouth\_width\_D}$	$+3.16E+01$	$+1.94E+00$	$+1.63E+01$
Summary statistics				
Number of observations = 30514				
$\mathbf{L}(0) = -67046.1$				
$\mathbf{L}(\hat{\beta}) = -46109.2$				
$\bar{\rho}^2 = 0.3111$				

Table 6.2: *Estimation results for the MNL model with primary and auxiliary AUs.*

parameters. In this case, the unrestricted model is the one in which auxiliary visual clues are added to primary AUs. Introducing additional parameters will always result in a higher likelihood score. However, there comes a point when adding parameters is no longer justified in terms of significant improvement in fit of a model to a particular dataset. The LRT provides one objective criterion for selecting among possible models. The LRT begins with a comparison of the log-likelihood scores of the two models:

$$LR = -2(lnL_0 - lnL_1)$$

where L_0 and L_1 are, respectively, the maximum of the likelihood function under the null hypothesis (restricted model) and the maximum with that constraint relaxed (unrestricted model). This LRT statistic asymptotically follows a chi-square distribution. To determine if the difference in likelihood scores among the two models is statistically significant, we next must consider the degrees of freedom, d. In the LRT, degrees of freedom is equal to the number of additional parameters in the more complex model:

$$d = D_1 - D_0$$

where D_0 and D_1 are the number of parameters in the two different models. Using this information we can then determine the critical value of the test statistic from standard statistical tables. We reject the null hypothesis that restrictions are true, namely, that the model without auxiliary AUs fits better the dataset, if

$$-2(lnL_0 - lnL_1) > \chi^2_{((1-\alpha),d)} \tag{6.1}$$

with α, the level of significance. In this specific case, the degrees of freedom are $d = 82 - 67 = 15$ and using $\alpha = 0.01$ yields

$$-2(-47143.9 + 46109.2) = 2069.4 > 30.58$$

We can therefore reject the null hypothesis and conclude that the auxiliary AUs coefficients should be included in the model.

6.2.3 Model Specification with Transient Features

The estimation results for this model are reported in Table 6.3. Almost all the explanatory variables are statistically significant and have their expected signs. The learned parameters show important consistencies with the common reading of facial expressions in terms of facial component modifications, as seen in the previous paragraphs. Also coefficients for transient features are consistent with this interpretation. All these parameters have a positive sign, showing that the presence of transient wrinkles and furrows in the face has a positive impact in the corresponding utilities. For instance, positive $\beta_{forehead\_SU}$ and $\beta_{forehead\_A}$ reflect a preference for the "Surprise" and "Anger" alternatives with respect to "Neutral" when furrows are present on the forehead. Moreover, the higher value of the coefficient $\beta_{forehead\_SU}$ shows a greater impact of this feature on the utility of "Surprise" compared to "Anger". A similar consideration holds for $\beta_{nasolabial\_D}$, the coefficient for a categorical variable equal to 1 if the nasolabial furrow typical of "Disgust" appear in the face. As for the parameters $\beta_{naswrink\_D}$ and $\beta_{browwrink\_A}$, they are significantly different from zero. This indicates that the more wrinkles are present respectively in region $\square X$ and $\square Y$ of Fig. 5.6a, the larger is the impact of this parameters in the corresponding utilities.
The interpretation of the alternative-specific constants is analogous to the previous model specifications. They are all significantly different from zero, their signs have not changed and also their magnitudes are almost the same as before.

At this point, we can apply the likelihood ratio test for the model with transient feature (M_2) against the model with only primary and auxiliary

MNL estimation				
Parameter number	Parameter name	Parameter estimate	Robust standard error	Robust *t statistic*
1	ASC_A	$-2.30E+00$	$+3.76E-01$	$-6.13E+00$
2	ASC_D	$-1.66E+00$	$+2.85E-01$	$-5.84E+00$
3	ASC_{DK}	$-2.31E+00$	$+3.72E-02$	$-6.21E+01$
4	ASC_F	$-2.96E+00$	$+1.21E+00$	$-2.44E+00$
5	ASC_H	$+1.22E+00$	$+3.59E-01$	$+3.41E+00$
7	ASC_N	$0.00E+00$	fixed	
8	ASC_O	$-1.36E+00$	$+3.37E-01$	$-4.03E+00$
9	ASC_{SA}	$-3.15E+00$	$+5.35E-01$	$-5.89E+00$
10	ASC_{SU}	$-4.21E+00$	$+2.19E-01$	$-1.93E+01$
11	$\beta_{brow\_dist\_A}$	$-1.05E+01$	$+2.19E+00$	$-4.79E+00$
12	$\beta_{browEye2\_r\_D}$	$-5.90E+01$	$+2.96E+00$	$-1.99E+01$
13	$\beta_{eyeMouth\_dist\_r\_SA}$	$+6.08E+01$	$+4.83E+00$	$+1.26E+01$
14	$\beta_{mouth\_height\_F}$	$+4.29E+01$	$+1.64E+00$	$+2.62E+01$
15	$\beta_{mouth\_height\_SU}$	$+5.50E+01$	$+1.44E+00$	$+3.82E+01$
16	$\beta_{mouth\_width\_H}$	$+1.01E+02$	$+2.73E+00$	$+3.69E+01$
17	$\beta_{browwrink\_A}$	$+4.40E+00$	$+1.85E+00$	$+2.38E+00$
18	$\beta_{forehead\_A}$	$+1.64E-01$	$+8.74E-02$	$+1.88E+00$
19	$\beta_{forehead\_SU}$	$+5.41E-01$	$+6.81E-02$	$+7.94E+00$
20	$\beta_{nasolabial\_D}$	$+7.70E-01$	$+5.88E-02$	$+1.31E+01$
21	$\beta_{naswrink\_D}$	$+1.88E+01$	$+7.04E-01$	$+2.67E+01$
Summary statistics				
Number of observations = 30514				
$\mathbf{L}(0) = -67046.1$				
$\mathbf{L}(\hat{\beta}) = -45563.2$				
$\bar{\rho}^2 = 0.3190$				

Table 6.3: *Estimation results for the MNL model with primary AUs, auxiliary AUs and transient features.*

AUs (M_1). In this case, the null hypothesis is that all the coefficients for wrinkles and furrows descriptors are zero. As usual, $-2(lnL_1 - lnL_2)$ is chi-square distributed, with $d = 93 - 82 = 11$ degrees of freedom. If we consider again $\alpha = 0.01$, we have:

$$-2(-46109.2 + 45563.2) = 1092 > 24.73$$

We can therefore reject the null hypothesis, and conclude that adding variables for transient features improves the model.

6.2.4 Model Specification with Socio-Economic Attributes

In the last model we add some variables to describe the decision maker. Table 6.4 shows the estimation results for this model. The interpretation of the facial measures parameters and alternative specific constants made for the previous example holds in this case too.
The coefficients for the socio-economic variables have been estimated and the majority of them is significantly different from zero at 95 percent. We can observe that the parameter β_{gender_SA} in the "Sadness" alternative has a negative sign. It implies that men have lower probability than women to choose the alternative "Sadness" with respect to "Neutral", since the variable "Gender" is 1 if the decision maker is masculine. The positive sign of the age coefficient β_{age1_F} reflects a preference of younger individuals (18-40 years old) for the "Fear" alternative with respect to "Neutral". The coefficient related to the language of the decision maker is positive in the utility of "Disgust" for Italian speaking individuals. It reflects the ability of Italians in distinguishing between neutral expression and disgust better than English speaking participants.

To test if adding socio-economic variables improves the descriptiveness of the model, we use once again the likelihood ratio test. We compare the log-likelihood functions of this model and of the previous one, where only facial descriptors are included. In this case, the degrees of freedom

MNL estimation				
Parameter number	Parameter name	Parameter estimate	Robust standard error	Robust *t statistic*
1	ASC_A	$-2.32E+00$	$+3.92E-01$	$-5.92E+00$
2	ASC_D	$-1.95E+00$	$+3.23E-01$	$-6.03E+00$
3	ASC_{DK}	$-2.31E+00$	$+3.72E-02$	$-6.21E+01$
4	ASC_F	$-3.34E+00$	$+1.22E+00$	$-2.74E+00$
5	ASC_H	$+1.04E+00$	$+3.72E-01$	$+2.80E+00$
6	ASC_N	$0.00E+00$	fixed	
7	ASC_O	$-1.57E+00$	$+3.63E-01$	$-4.34E+00$
8	ASC_{SA}	$-3.06E+00$	$+5.45E-01$	$-5.62E+00$
9	ASC_{SU}	$-4.40E+00$	$+2.26E-01$	$-1.94E+01$
10	$\beta_{browwrink\_A}$	$+4.42E+00$	$+1.84E+00$	$+2.40E+00$
11	$\beta_{eyeNose\_dist\_r\_SA}$	$-1.38E+02$	$+7.50E+00$	$-1.83E+01$
12	$\beta_{mouth\_height\_SU}$	$+5.52E+01$	$+1.45E+00$	$+3.82E+01$
13	$\beta_{mouth\_width\_H}$	$+1.02E+02$	$+2.76E+00$	$+3.69E+01$
14	$\beta_{nasolabial\_D}$	$+7.68E-01$	$+5.91E-02$	$+1.30E+01$
15	β_{age1_F}	$+3.14E-01$	$+9.49E-02$	$+3.30E+00$
16	β_{form1_H}	$+2.18E-01$	$+8.57E-02$	$+2.55E+00$
17	β_{gender_SA}	$-2.26E-01$	$+4.41E-02$	$-5.12E+00$
18	β_{lan3_D}	$+1.67E-01$	$+6.21E-02$	$+2.69E+00$
19	β_{reg4_O}	$+1.47E+00$	$+2.08E-01$	$+7.08E+00$
20	β_{reg7_O}	$+5.41E-01$	$+2.26E-01$	$+2.39E+00$
21	$\beta_{science3\_A}$	$-1.52E-01$	$+5.90E-02$	$-2.58E+00$
Summary statistics				
Number of observations = 30514				
$\mathbf{L}(0) = -67046.1$				
$\mathbf{L}(\hat{\beta}) = -45285.4$				
$\bar{\rho}^2 = 0.3219$				

Table 6.4: *Estimation results for the MNL model with primary AUs, auxiliary AUs, transient features and socio-economic attributes.*

are $d = 177 - 93 = 84$. We look up in the chi-square table the critical value for the $\chi^2_{(0.99,90)}$, that is 124.1. Therefore, the condition to reject the null hypothesis is:

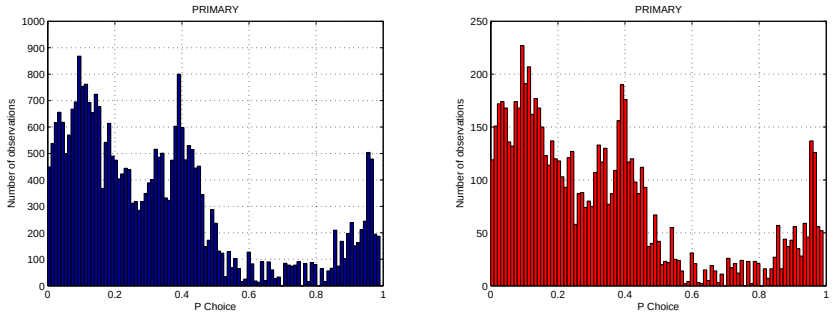
$$-2(-45563.2 + 45285.4) = 555.6 > 124.1$$

and it is satisfied. We can conclude that the unrestricted model, with both socio-economic characteristics and attributes describing the face, should be preferred to the models presented in the previous paragraphs.

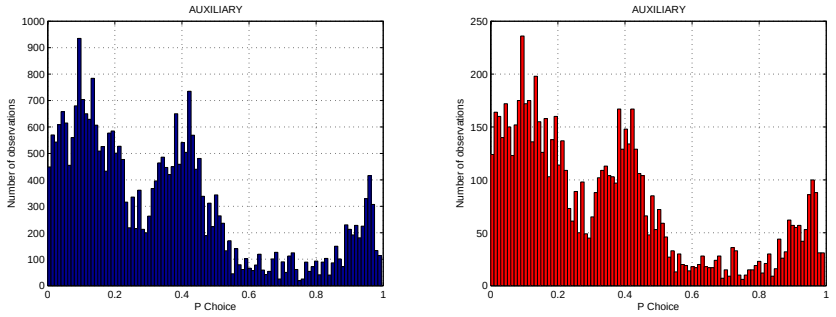
6.3 Performance of Forecasting

We demonstrated that the attributes we included in the proposed models provide significant information and affect the choice of the decision maker. Now we want to test how well these models can predict the chosen alternative given a set of attributes. To do this, we use BioSim (see Sec. 6.1), that can compute predicted probabilities for all the models estimated with BIOGEME. In fact, the estimated coefficients of a discrete choice model can be used to calculate the choice probability of each alternative for each observation in the sample. If we give the data file and the model file as inputs to BioSim, it generates an output file that contains, for each observation, the chosen alternative and the predicted probabilities for all the alternatives of the choice set.

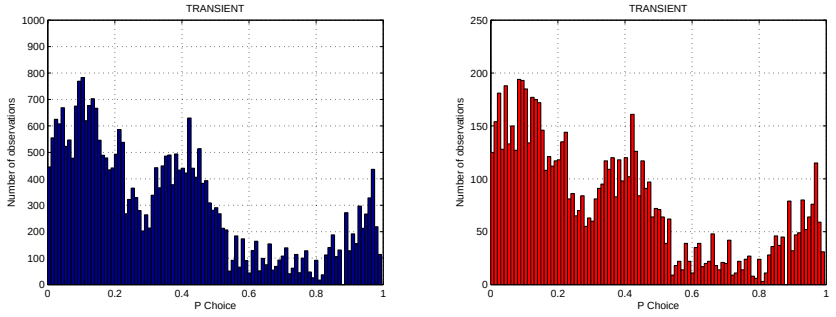
In Figure 6.1 we plot the predicted probabilities of the chosen alternative for each observation of the dataset, obtained with the four different models. In blue we report the probabilities calculated on the dataset used for estimating the model (about 31000 observations), and in red the ones for the validation set (about 8000 annotated images). To decide whether a particular choice is correctly reproduced, we compare its chosen alternative's probability with a threshold, namely, the probability of the worst case. The worst case is when all the alternatives have the same predicted probability, that is $p = \frac{1}{9} \approx 0.12$ for this particular problem.



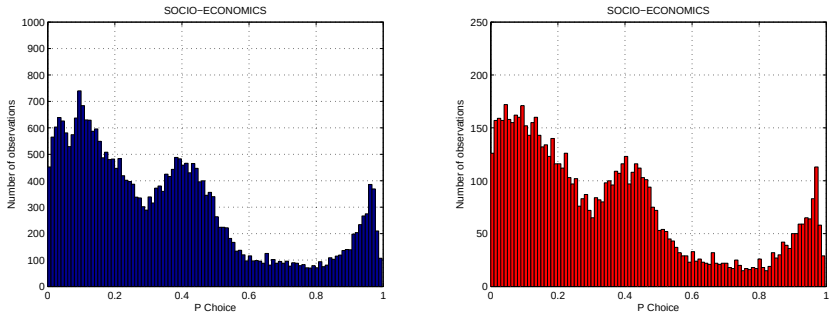
(a)



(b)



(c)



(d)

Figure 6.1: *Plots of choice probabilities: (a) Model with primary AUs, (b) Model with primary and auxiliary AUs, (c) Model with primary AUs, auxiliary AUs and transient features, (d) Model with facial descriptors and socio-economic characteristics.*

Training Set

Model	# observations	Correct predictions	Correct predictions (%)
Primary	30514	22749	74.55
Auxiliary	30514	23059	75.57
Transient	30514	23178	75.96
Socio-econ.	30514	23234	76.14

Validation Set

Model	# observations	Correct predictions	Correct predictions (%)
Primary	7629	5603	73.44
Auxiliary	7629	5679	74.44
Transient	7629	5724	75.03
Socio-econ.	7629	5749	75.36

Table 6.5: *Performance of the proposed Discrete Choice Models.*

In fact, this means that we have no information to forecast the choice. Therefore, if the probability for the chosen alternative is bigger than this threshold, we can say that the model correctly reproduces the behaviour of the decision makers.

As results from Tab. 6.5, all the four models show good performances. The percentage of correctly predicted observations rise from the first model (74.55% for the training set and 73.44% for new observations) to the last (76.14% for the blue plot and 75.36% for the red one). This means that adding attributes in the utility functions enhances the results. Moreover, it is worth noting that the shape of the plots is similar for the training and the validation set. Hence, we can conclude that our models exhibit good generalization performance.

CHAPTER

SEVEN

Conclusions and Future Works

The contribution of the presented work is twofold. First, we proposed a new method for facial expressions modelling, based on discrete choice analysis. Second, we included in the model the variation in the population of individuals performing the labelling task. Both the feature set and the modelling approach are motivated by the research of methods able to bring into the process the modeller prior knowledge.

Based on our experience, a subjective component biases the labelling process, requiring a detailed statistical analysis on collected data from an heterogeneous population. DCMs, coming from econometric, provide a strong statistical framework to include such an heterogeneity. The socio-economic features, describing the expert labellers, are introduced in the DCM as explanatory variables. We demonstrated that these attributes affect the choice of an expression label for a particular face. In fact, the fourth model, where these features have been included, turned out to be the best one, significantly better than the others. Besides this, also the employment of the FACS action units as facial descriptors proved to be suitable for discrete choice models, even if applied to static images.

The DCM modelling approach redefines facial expression classification as a discrete choice process, which well matches the human observer

labelling procedure. Prior knowledge can be included in the process customizing the utilities. The result of the DCM is a set of probabilities assigned to the alternatives, represented by the possible expressions. Our models generalize reasonably well. Their performances are, indeed, almost the same both if the models are applied to the training set of observations and to the validation set, containing new data.

From all these considerations, we can conclude that the DCM approach seems to be suitable for modelling facial expressions, showing good and encouraging performance.

However, this work represents one of the first attempts to apply discrete choice analysis for modelling facial expressions, so several means for improvement are possible. Since the strength of DCMs is the possibility of customising the utilities, some refinements can be made along these lines. First of all, socio-economic features entering in the choice process could be investigated more accurately by applying a segmentation. This means that, instead of introducing a parameter for each socio-economic attribute, we divide the population with respect to that feature. For example, we can study the behaviour of men and women by analysing the two groups separately. In this way, it is possible to get more information in order to define which parameters are really involved in the choice process. At the same time, we will reduce the complexity of the problem, by decreasing the number of the unknown quantities. The models can be improved also by adding other attributes, like descriptors of the dynamic evolution of facial expressions, that can make the modelling more robust. In the third place, other expressions of utility function can be used, for instance, non-linear combinations of parameters. However, this kind of models needs large computational resources, so only few parameters should be included in the model, in order to have a solvable problem.

It is important to stress the fact that a comparison with whatever

state of the art facial expression classifiers is meaningless. In fact, the fundamental hypothesis in a "learning by examples" framework is the use of training sets which have correct annotations and where each example must be associated to a unique class. In our case this hypothesis does not hold anymore. The main difference with a "classic" classification problem relies on the fact that every example can indeed belong to different classes, reflecting the heterogeneity in human perception of expressions.

APPENDIX

A

Estimation Results

A.1 MNL with Primary Action Units

MNL estimation				
Parameter name	Parameter estimate	Robust standard error	Robust t statistic	
ASC_A	-2.09E+00	3.13E-01	-6.68E+00	
ASC_D	-7.84E-01	2.87E-01	-2.73E+00	
ASC_{DK}	-2.31E+00	3.72E-02	-6.21E+01	
ASC_F	1.93E+00	1.11E+00	1.74E+00	*
ASC_H	1.04E+00	3.16E-01	3.28E+00	
ASC_N	0.00E+00	fixed		
ASC_O	-1.54E+00	3.24E-01	-4.74E+00	
ASC_{SA}	-3.00E+00	3.08E-01	-9.74E+00	
ASC_{SU}	-4.37E+00	1.91E-01	-2.29E+01	
$\beta_{brow\_dist\_A}$	-1.31E+01	2.14E+00	-6.12E+00	
$\beta_{brow\_dist\_F}$	-2.84E+01	4.70E+00	-6.04E+00	
$\beta_{browEye2\_A}$	-4.30E+01	5.22E+00	-8.24E+00	
$\beta_{browEye2\_O}$	1.90E+01	4.46E+00	4.25E+00	
$\beta_{browEye2\_SA}$	4.63E+01	5.42E+00	8.53E+00	
$\beta_{browEye2\_SU}$	-1.81E+01	4.22E+00	-4.28E+00	

$\beta_{browEye2\_r\_A}$	-6.99E+01	4.66E+00	-1.50E+01
$\beta_{browEye2\_r\_D}$	-8.23E+01	2.84E+00	-2.90E+01
$\beta_{browEye2\_r\_F}$	6.98E+01	1.17E+01	5.98E+00
$\beta_{browEye2\_r\_SA}$	-1.07E+02	5.93E+00	-1.81E+01
$\beta_{browEye2\_r\_SU}$	2.24E+01	5.84E+00	3.83E+00
$\beta_{browEye3\_l\_A}$	-1.47E+01	3.65E+00	-4.02E+00
$\beta_{browEye3\_l\_SU}$	1.33E+01	3.11E+00	4.26E+00
$\beta_{eye\_angle\_l\_A}$	-1.41E+00	3.20E-01	-4.41E+00
$\beta_{eye\_angle\_l\_F}$	5.14E+00	5.86E-01	8.77E+00
$\beta_{eye\_angle\_r\_A}$	3.59E+00	3.31E-01	1.09E+01
$\beta_{eye\_angle\_r\_F}$	2.47E+00	4.15E-01	5.96E+00
$\beta_{eye\_angle2\_l\_F}$	-1.09E+00	2.97E-01	-3.67E+00
$\beta_{eye\_angle2\_r\_F}$	-3.30E+00	3.16E-01	-1.04E+01
$\beta_{eye\_height\_l\_F}$	-3.93E+01	1.02E+01	-3.84E+00
$\beta_{eye\_height\_l\_H}$	-3.17E+01	5.92E+00	-5.35E+00
$\beta_{eye\_height\_r\_F}$	-1.35E+02	1.37E+01	-9.82E+00
$\beta_{eye\_height\_r\_H}$	3.72E+01	5.25E+00	7.08E+00
$\beta_{eyeBrow\_angle\_l\_F}$	7.09E+00	5.93E-01	1.20E+01
$\beta_{eyeBrow\_angle\_l\_O}$	-2.47E+00	3.58E-01	-6.90E+00
$\beta_{eyeBrow\_angle\_l\_SA}$	-4.96E+00	3.43E-01	-1.45E+01
$\beta_{eyeBrow\_angle\_r\_F}$	-6.56E+00	8.18E-01	-8.03E+00
$\beta_{eyeBrow\_angle\_r\_O}$	-1.77E+00	2.13E-01	-8.30E+00
$\beta_{eyeBrow\_angle\_r\_SA}$	6.29E+00	3.80E-01	1.65E+01
$\beta_{eyeBrow\_angle\_r\_SU}$	-2.04E+00	3.44E-01	-5.93E+00
$\beta_{eyeMouth\_dist\_l\_F}$	8.48E+01	8.57E+00	9.90E+00
$\beta_{eyeMouth\_dist\_l\_H}$	-1.45E+01	4.21E+00	-3.43E+00
$\beta_{eyeMouth\_dist\_l\_O}$	-2.97E+01	4.10E+00	-7.23E+00
$\beta_{eyeMouth\_dist\_l\_SA}$	-4.17E+01	4.21E+00	-9.90E+00
$\beta_{eyeMouth\_dist2\_r\_D}$	4.70E+01	2.46E+00	1.91E+01
$\beta_{eyeMouth\_dist2\_r\_O}$	1.12E+01	2.61E+00	4.28E+00
$\beta_{eyeMouth\_dist\_r\_F}$	-6.33E+01	8.22E+00	-7.71E+00
$\beta_{eyeMouth\_dist\_r\_H}$	-6.04E+01	4.06E+00	-1.49E+01

$\beta_{eyeMouth\_dist\_r\_O}$	3.00E+01	3.83E+00	7.85E+00
$\beta_{eyeMouth\_dist\_r\_SA}$	6.98E+01	4.05E+00	1.72E+01
$\beta_{eyeMouth\_dist2\_l\_D}$	-4.62E+01	2.96E+00	-1.56E+01
$\beta_{eyeMouth\_dist2\_l\_O}$	-1.30E+01	2.49E+00	-5.25E+00
$\beta_{eyeNose\_dist\_l\_A}$	-1.46E+01	6.53E+00	-2.24E+00
$\beta_{eyeNose\_dist\_l\_D}$	6.08E+01	6.21E+00	9.79E+00
$\beta_{eyeNose\_dist\_l\_O}$	7.90E+01	5.55E+00	1.42E+01
$\beta_{eyeNose\_dist\_r\_A}$	8.05E+01	8.08E+00	9.96E+00
$\beta_{eyeNose\_dist\_r\_D}$	-5.89E+01	7.09E+00	-8.30E+00
$\beta_{eyeNose\_dist\_r\_O}$	-8.59E+01	6.53E+00	-1.32E+01
$\beta_{mouth\_height\_A}$	-2.45E+01	2.08E+00	-1.18E+01
$\beta_{mouth\_height\_D}$	-9.15E+00	1.50E+00	-6.11E+00
$\beta_{mouth\_height\_SU}$	4.46E+01	1.37E+00	3.26E+01
$\beta_{mouth\_width\_A}$	2.06E+01	2.15E+00	9.59E+00
$\beta_{mouth\_width\_D}$	2.70E+01	1.86E+00	1.46E+01
$\beta_{mouth\_width\_F}$	1.60E+01	3.17E+00	5.06E+00
$\beta_{mouth\_width\_H}$	9.15E+01	1.69E+00	5.41E+01
$\beta_{mouth\_width\_O}$	1.80E+01	1.88E+00	9.57E+00
$\beta_{mouth\_width\_SA}$	3.99E+00	1.90E+00	2.10E+00
$\beta_{mouthNose\_dist\_D}$	-1.30E+01	3.72E+00	-3.51E+00
$\beta_{mouthNose\_dist2\_SA}$	-1.97E+01	1.32E+00	-1.50E+01

A.2 MNL with Auxiliary Action Units

MNL estimation				
Parameter name	Parameter estimate	Robust standard error	Robust t statistic	
ASC_A	-2.69E+00	3.69E-01	-7.29E+00	
ASC_D	-1.06E+00	2.92E-01	-3.62E+00	
ASC_{DK}	-2.31E+00	3.72E-02	-6.21E+01	
ASC_F	-4.19E+00	1.18E+00	-3.54E+00	
ASC_H	1.02E+00	3.47E-01	2.95E+00	
ASC_N	0.00E+00	fixed		
ASC_O	-1.82E+00	3.31E-01	-5.50E+00	
ASC_{SA}	-3.17E+00	5.31E-01	-5.96E+00	
ASC_{SU}	-4.44E+00	2.13E-01	-2.09E+01	
$\beta_{brow\_dist\_A}$	-1.41E+01	2.15E+00	-6.56E+00	
$\beta_{brow\_dist\_F}$	-3.88E+01	4.15E+00	-9.35E+00	
$\beta_{brow\_dist\_SA}$	-2.60E+01	2.22E+00	-1.17E+01	
$\beta_{browEye2\_l\_A}$	-3.59E+01	5.28E+00	-6.80E+00	
$\beta_{browEye2\_l\_O}$	2.40E+01	4.60E+00	5.23E+00	
$\beta_{browEye2\_l\_SA}$	-1.37E+00	6.22E+00	-2.20E-01	*
$\beta_{browEye2\_l\_SU}$	1.35E+00	4.64E+00	2.92E-01	*
$\beta_{browEye2\_r\_A}$	-7.57E+01	4.88E+00	-1.55E+01	
$\beta_{browEye2\_r\_D}$	-8.41E+01	2.88E+00	-2.92E+01	
$\beta_{browEye2\_r\_F}$	-5.31E+00	1.08E+01	-4.92E-01	*
$\beta_{browEye2\_r\_SA}$	-7.58E+01	7.31E+00	-1.04E+01	
$\beta_{browEye2\_r\_SU}$	5.86E+00	6.74E+00	8.70E-01	*
$\beta_{browEye3\_l\_A}$	-1.83E+01	3.72E+00	-4.91E+00	
$\beta_{browEye3\_l\_SU}$	1.74E+01	3.86E+00	4.52E+00	
$\beta_{eye\_angle\_l\_A}$	-1.24E+00	3.51E-01	-3.53E+00	
$\beta_{eye\_angle\_l\_F}$	5.72E+00	5.00E-01	1.14E+01	
$\beta_{eye\_angle\_l\_SA}$	4.40E+00	3.08E-01	1.43E+01	
$\beta_{eye\_angle\_r\_A}$	3.54E+00	3.77E-01	9.39E+00	
$\beta_{eye\_angle\_r\_F}$	7.59E-01	4.33E-01	1.75E+00	*

$\beta_{eye\_angle\_r\_SA}$	-4.89E+00	3.02E-01	-1.62E+01	
$\beta_{eye\_angle2\_l\_F}$	-1.57E+00	2.72E-01	-5.77E+00	
$\beta_{eye\_angle2\_r\_F}$	3.15E-01	3.35E-01	9.42E-01	*
$\beta_{eye\_angle2\_r\_SA}$	5.76E-01	1.38E-01	4.18E+00	
$\beta_{eye\_height\_l\_F}$	-9.58E+01	1.09E+01	-8.77E+00	
$\beta_{eye\_height\_l\_H}$	-3.17E+01	6.68E+00	-4.75E+00	
$\beta_{eye\_height\_l\_SU}$	-7.62E+01	7.53E+00	-1.01E+01	
$\beta_{eye\_height\_r\_F}$	-2.52E+01	1.39E+01	-1.82E+00	*
$\beta_{eye\_height\_r\_H}$	4.79E+01	5.38E+00	8.90E+00	
$\beta_{eye\_height\_r\_SU}$	3.71E+01	4.81E+00	7.71E+00	
$\beta_{eyeBrow\_angle\_l\_F}$	8.07E+00	5.74E-01	1.41E+01	
$\beta_{eyeBrow\_angle\_l\_O}$	-2.60E+00	3.68E-01	-7.08E+00	
$\beta_{eyeBrow\_angle\_l\_SA}$	-9.97E-01	4.49E-01	-2.22E+00	
$\beta_{eyeBrow\_angle\_r\_F}$	-2.65E+00	7.74E-01	-3.42E+00	
$\beta_{eyeBrow\_angle\_r\_O}$	-2.15E+00	2.17E-01	-9.89E+00	
$\beta_{eyeBrow\_angle\_r\_SA}$	6.22E+00	4.63E-01	1.34E+01	
$\beta_{eyeBrow\_angle\_r\_SU}$	-1.65E+00	3.62E-01	-4.55E+00	
$\beta_{eyeMouth\_dist\_l\_F}$	4.01E+01	7.45E+00	5.38E+00	
$\beta_{eyeMouth\_dist\_l\_H}$	-1.48E+01	4.24E+00	-3.48E+00	
$\beta_{eyeMouth\_dist\_l\_O}$	-2.70E+01	4.22E+00	-6.38E+00	
$\beta_{eyeMouth\_dist\_l\_SA}$	-2.63E+01	5.42E+00	-4.86E+00	
$\beta_{eyeMouth\_dist\_r\_F}$	-3.49E+01	7.23E+00	-4.82E+00	
$\beta_{eyeMouth\_dist\_r\_H}$	-7.65E+01	4.25E+00	-1.80E+01	
$\beta_{eyeMouth\_dist\_r\_O}$	3.07E+01	3.91E+00	7.84E+00	
$\beta_{eyeMouth\_dist\_r\_SA}$	6.20E+01	4.84E+00	1.28E+01	
$\beta_{eyeMouth\_dist2\_l\_D}$	-4.62E+01	2.97E+00	-1.56E+01	
$\beta_{eyeMouth\_dist2\_l\_O}$	-1.33E+01	2.52E+00	-5.28E+00	
$\beta_{eyeMouth\_dist2\_r\_D}$	4.41E+01	2.46E+00	1.79E+01	
$\beta_{eyeMouth\_dist2\_r\_O}$	1.03E+01	2.64E+00	3.89E+00	
$\beta_{eyeNose\_dist\_l\_A}$	7.35E+00	6.58E+00	1.12E+00	*
$\beta_{eyeNose\_dist\_l\_D}$	7.63E+01	6.33E+00	1.21E+01	
$\beta_{eyeNose\_dist\_l\_F}$	7.83E+01	8.99E+00	8.70E+00	

$\beta_{eyeNose\_dist\_l\_O}$	9.28E+01	5.74E+00	1.62E+01	
$\beta_{eyeNose\_dist\_l\_SA}$	1.15E+02	6.38E+00	1.80E+01	
$\beta_{eyeNose\_dist\_r\_A}$	5.23E+01	8.16E+00	6.40E+00	
$\beta_{eyeNose\_dist\_r\_D}$	-7.73E+01	7.27E+00	-1.06E+01	
$\beta_{eyeNose\_dist\_r\_F}$	-8.96E+01	1.14E+01	-7.89E+00	
$\beta_{eyeNose\_dist\_r\_O}$	-1.07E+02	6.84E+00	-1.56E+01	
$\beta_{eyeNose\_dist\_r\_SA}$	-1.33E+02	7.31E+00	-1.82E+01	
$\beta_{mouth\_height\_A}$	-3.09E+01	3.04E+00	-1.02E+01	
$\beta_{mouth\_height\_D}$	-6.93E+00	1.64E+00	-4.24E+00	
$\beta_{mouth\_height\_F}$	4.30E+01	1.61E+00	2.67E+01	
$\beta_{mouth\_height\_H}$	-2.16E+00	2.81E+00	-7.69E-01	*
$\beta_{mouth\_height\_SA}$	-8.66E+00	2.41E+00	-3.59E+00	
$\beta_{mouth\_height\_SU}$	5.54E+01	1.43E+00	3.88E+01	
$\beta_{mouth\_width\_A}$	2.36E+01	2.30E+00	1.02E+01	
$\beta_{mouth\_width\_D}$	3.16E+01	1.94E+00	1.63E+01	
$\beta_{mouth\_width\_F}$	2.07E+01	2.29E+00	9.05E+00	
$\beta_{mouth\_width\_H}$	1.02E+02	2.69E+00	3.78E+01	
$\beta_{mouth\_width\_O}$	2.03E+01	2.02E+00	1.00E+01	
$\beta_{mouth\_width\_SA}$	-5.06E+00	2.24E+00	-2.26E+00	
$\beta_{mouthNose\_dist\_D}$	-6.37E+00	3.65E+00	-1.74E+00	*
$\beta_{mouthNose\_dist\_H}$	3.05E+01	2.46E+00	1.24E+01	
$\beta_{mouthNose\_dist2\_A}$	1.10E+01	2.60E+00	4.22E+00	
$\beta_{mouthNose\_dist2\_SA}$	-1.49E+01	2.22E+00	-6.68E+00	

A.3 MNL with Transient Features

MNL estimation				
Parameter name	Parameter estimate	Robust standard error	Robust t statistic	
ASC_A	-2.30E+00	3.76E-01	-6.13E+00	
ASC_D	-1.66E+00	2.85E-01	-5.84E+00	
ASC_{DK}	-2.31E+00	3.72E-02	-6.21E+01	
ASC_F	-2.96E+00	1.21E+00	-2.44E+00	
ASC_H	1.22E+00	3.59E-01	3.41E+00	
ASC_N	0.00E+00			
ASC_O	-1.36E+00	3.37E-01	-4.03E+00	
ASC_{SA}	-3.15E+00	5.35E-01	-5.89E+00	
ASC_{SU}	-4.21E+00	2.19E-01	-1.93E+01	
$\beta_{brow\_dist\_A}$	-1.05E+01	2.19E+00	-4.79E+00	
$\beta_{brow\_dist\_F}$	-3.51E+01	4.12E+00	-8.53E+00	
$\beta_{brow\_dist\_SA}$	-2.44E+01	2.24E+00	-1.09E+01	
$\beta_{browEye2\_l\_A}$	-3.42E+01	5.27E+00	-6.49E+00	
$\beta_{browEye2\_l\_O}$	2.12E+01	4.67E+00	4.53E+00	
$\beta_{browEye2\_l\_SA}$	-4.56E+00	6.32E+00	-7.22E-01	*
$\beta_{browEye2\_l\_SU}$	2.98E+00	4.62E+00	6.44E-01	*
$\beta_{browEye3\_l\_A}$	-2.67E+01	3.99E+00	-6.70E+00	
$\beta_{browEye3\_l\_SU}$	1.36E+01	3.90E+00	3.48E+00	
$\beta_{browEye2\_r\_A}$	-7.63E+01	4.88E+00	-1.56E+01	
$\beta_{browEye2\_r\_D}$	-5.90E+01	2.96E+00	-1.99E+01	
$\beta_{browEye2\_r\_F}$	-1.50E+01	1.05E+01	-1.43E+00	*
$\beta_{browEye2\_r\_SA}$	-6.84E+01	7.33E+00	-9.33E+00	
$\beta_{browEye2\_r\_SU}$	-5.85E-01	6.87E+00	-8.51E-02	*
$\beta_{browwrink\_A}$	4.40E+00	1.85E+00	2.38E+00	
$\beta_{browwrink\_D}$	1.28E+01	1.87E+00	6.84E+00	
$\beta_{browwrink\_O}$	5.93E+00	1.57E+00	3.78E+00	
$\beta_{browwrink\_SA}$	2.19E+00	1.88E+00	1.16E+00	*
$\beta_{eye\_angle2\_l\_F}$	-1.58E+00	2.73E-01	-5.79E+00	

$\beta_{eye\_angle2\_r\_F}$	2.06E-01	3.30E-01	6.24E-01	*
$\beta_{eye\_angle2\_r\_SA}$	7.37E-01	1.44E-01	5.13E+00	
$\beta_{eye\_angle\_l\_A}$	-7.47E-01	3.58E-01	-2.09E+00	
$\beta_{eye\_angle\_l\_F}$	5.96E+00	5.15E-01	1.16E+01	
$\beta_{eye\_angle\_l\_SA}$	4.57E+00	3.15E-01	1.45E+01	
$\beta_{eye\_angle\_r\_A}$	3.50E+00	3.78E-01	9.25E+00	
$\beta_{eye\_angle\_r\_F}$	6.93E-01	4.38E-01	1.58E+00	*
$\beta_{eye\_angle\_r\_SA}$	-4.74E+00	3.06E-01	-1.55E+01	
$\beta_{eyeBrow\_angle\_l\_F}$	7.05E+00	5.71E-01	1.23E+01	
$\beta_{eyeBrow\_angle\_l\_O}$	-2.52E+00	3.82E-01	-6.59E+00	
$\beta_{eyeBrow\_angle\_l\_SA}$	-9.83E-01	4.80E-01	-2.05E+00	
$\beta_{eyeBrow\_angle\_r\_F}$	-1.75E+00	7.50E-01	-2.34E+00	
$\beta_{eyeBrow\_angle\_r\_O}$	-2.00E+00	2.29E-01	-8.76E+00	
$\beta_{eyeBrow\_angle\_r\_SA}$	5.65E+00	4.73E-01	1.19E+01	
$\beta_{eyeBrow\_angle\_r\_SU}$	-1.33E+00	3.73E-01	-3.56E+00	
$\beta_{eyeMouth\_dist2\_l\_D}$	-4.12E+01	3.17E+00	-1.30E+01	
$\beta_{eyeMouth\_dist2\_l\_O}$	-1.12E+01	2.62E+00	-4.28E+00	
$\beta_{eyeMouth\_dist\_l\_F}$	3.71E+01	7.46E+00	4.98E+00	
$\beta_{eyeMouth\_dist\_l\_H}$	-1.38E+01	4.24E+00	-3.26E+00	
$\beta_{eyeMouth\_dist\_l\_O}$	-2.97E+01	4.24E+00	-6.99E+00	
$\beta_{eyeMouth\_dist\_l\_SA}$	-2.81E+01	5.41E+00	-5.20E+00	
$\beta_{eyeMouth\_dist2\_r\_D}$	2.88E+01	2.63E+00	1.10E+01	
$\beta_{eyeMouth\_dist2\_r\_O}$	8.76E+00	2.77E+00	3.17E+00	
$\beta_{eyeMouth\_dist\_r\_F}$	-3.62E+01	7.26E+00	-4.98E+00	
$\beta_{eyeMouth\_dist\_r\_H}$	-7.76E+01	4.26E+00	-1.82E+01	
$\beta_{eyeMouth\_dist\_r\_O}$	3.18E+01	3.96E+00	8.02E+00	
$\beta_{eyeMouth\_dist\_r\_SA}$	6.08E+01	4.83E+00	1.26E+01	
$\beta_{eyeNose\_dist\_l\_A}$	5.10E+00	6.90E+00	7.40E-01	*
$\beta_{eyeNose\_dist\_l\_D}$	9.11E+01	6.25E+00	1.46E+01	
$\beta_{eyeNose\_dist\_l\_F}$	6.45E+01	8.86E+00	7.28E+00	
$\beta_{eyeNose\_dist\_l\_O}$	9.46E+01	5.97E+00	1.59E+01	
$\beta_{eyeNose\_dist\_l\_SA}$	1.19E+02	6.57E+00	1.80E+01	

$\beta_{eyeNose\_dist\_r\_A}$	5.22E+01	8.40E+00	6.21E+00	
$\beta_{eyeNose\_dist\_r\_D}$	-9.52E+01	7.18E+00	-1.33E+01	
$\beta_{eyeNose\_dist\_r\_F}$	-7.12E+01	1.13E+01	-6.28E+00	
$\beta_{eyeNose\_dist\_r\_O}$	-1.11E+02	7.01E+00	-1.58E+01	
$\beta_{eyeNose\_dist\_r\_SA}$	-1.37E+02	7.49E+00	-1.83E+01	
$\beta_{forehead\_A}$	1.64E-01	8.74E-02	1.88E+00	*
$\beta_{forehead\_F}$	8.92E-01	8.60E-02	1.04E+01	
$\beta_{forehead\_O}$	2.91E-01	6.57E-02	4.42E+00	
$\beta_{forehead\_SU}$	5.41E-01	6.81E-02	7.94E+00	
$\beta_{eye\_height\_l\_F}$	-8.68E+01	1.11E+01	-7.81E+00	
$\beta_{eye\_height\_l\_H}$	-3.25E+01	6.79E+00	-4.79E+00	
$\beta_{eye\_height\_l\_SU}$	-6.71E+01	7.57E+00	-8.86E+00	
$\beta_{mouth\_height\_A}$	-2.76E+01	3.08E+00	-8.95E+00	
$\beta_{mouth\_height\_D}$	-4.16E+00	1.72E+00	-2.41E+00	
$\beta_{mouth\_height\_F}$	4.29E+01	1.64E+00	2.62E+01	
$\beta_{mouth\_height\_H}$	-1.57E+00	2.83E+00	-5.54E-01	*
$\beta_{mouth\_height\_SA}$	-8.88E+00	2.44E+00	-3.64E+00	
$\beta_{mouth\_height\_SU}$	5.50E+01	1.44E+00	3.82E+01	
$\beta_{mouthNose\_dist2\_A}$	8.59E+00	2.62E+00	3.28E+00	
$\beta_{mouthNose\_dist2\_SA}$	-1.41E+01	2.26E+00	-6.25E+00	
$\beta_{mouthNose\_dist\_D}$	1.43E+01	3.63E+00	3.95E+00	
$\beta_{mouthNose\_dist\_H}$	3.12E+01	2.42E+00	1.28E+01	
$\beta_{mouth\_width\_A}$	2.38E+01	2.30E+00	1.03E+01	
$\beta_{mouth\_width\_D}$	2.99E+01	1.97E+00	1.51E+01	
$\beta_{mouth\_width\_F}$	1.90E+01	2.35E+00	8.09E+00	
$\beta_{mouth\_width\_H}$	1.01E+02	2.73E+00	3.69E+01	
$\beta_{mouth\_width\_O}$	1.91E+01	2.05E+00	9.29E+00	
$\beta_{mouth\_width\_SA}$	-5.40E+00	2.21E+00	-2.44E+00	
$\beta_{nasolabial\_D}$	7.70E-01	5.88E-02	1.31E+01	
$\beta_{naswrink\_D}$	1.88E+01	7.04E-01	2.67E+01	
$\beta_{naswrink\_O}$	3.77E+00	7.95E-01	4.74E+00	
$\beta_{eye\_height\_r\_F}$	-3.37E+01	1.40E+01	-2.40E+00	

$\beta_{eye\_height\_r\_H}$	4.71E+01	5.30E+00	8.87E+00
$\beta_{eye\_height\_r\_SU}$	3.22E+01	4.85E+00	6.63E+00

A.4 MNL with Socio-Economic Attributes

MNL estimation				
Parameter name	Parameter estimate	Robust standard error	Robust t statistic	
ASC_A	-2,32E+00	3,92E-01	-5,92E+00	
ASC_D	-1,95E+00	3,23E-01	-6,03E+00	
ASC_{DK}	-2,31E+00	3,72E-02	-6,21E+01	
ASC_F	-3,34E+00	1,22E+00	-2,74E+00	
ASC_H	1,04E+00	3,72E-01	2,80E+00	
ASC_N	0,00E+00	fixed		
ASC_O	-1,57E+00	3,63E-01	-4,34E+00	
ASC_{SA}	-3,06E+00	5,45E-01	-5,62E+00	
ASC_{SU}	-4,40E+00	2,26E-01	-1,94E+01	
β_{age1_D}	-3,79E-01	1,18E-01	-3,22E+00	
β_{age1_F}	3,14E-01	9,49E-02	3,30E+00	
β_{age1_O}	-2,01E-01	5,62E-02	-3,57E+00	
β_{age1_SA}	-1,67E-01	9,65E-02	-1,73E+00	*
β_{age2_D}	-2,27E-01	1,26E-01	-1,80E+00	*
β_{age2_H}	-1,79E-01	7,89E-02	-2,27E+00	
β_{age2_SA}	-2,30E-01	1,10E-01	-2,10E+00	
$\beta_{brow\_dist\_A}$	-1,06E+01	2,20E+00	-4,83E+00	
$\beta_{brow\_dist\_F}$	-3,50E+01	4,13E+00	-8,48E+00	
$\beta_{brow\_dist\_SA}$	-2,44E+01	2,25E+00	-1,09E+01	
$\beta_{browEye2\_l\_A}$	-3,42E+01	5,29E+00	-6,45E+00	
$\beta_{browEye2\_l\_O}$	2,10E+01	4,70E+00	4,48E+00	
$\beta_{browEye2\_l\_SA}$	-4,65E+00	6,32E+00	-7,36E-01	*
$\beta_{browEye2\_l\_SU}$	2,80E+00	4,63E+00	6,06E-01	*
$\beta_{browEye3\_l\_A}$	-2,67E+01	3,99E+00	-6,69E+00	
$\beta_{browEye3\_l\_SU}$	1,36E+01	3,91E+00	3,48E+00	
$\beta_{browEye2\_r\_A}$	-7,65E+01	4,90E+00	-1,56E+01	
$\beta_{browEye2\_r\_D}$	-5,93E+01	2,97E+00	-2,00E+01	
$\beta_{browEye2\_r\_F}$	-1,43E+01	1,05E+01	-1,36E+00	*

$\beta_{browEye2\_r\_SA}$	-6,83E+01	7,34E+00	-9,31E+00	
$\beta_{browEye2\_r\_SU}$	-1,03E+00	6,88E+00	-1,49E-01	*
$\beta_{browwrink\_A}$	4,42E+00	1,84E+00	2,40E+00	
$\beta_{browwrink\_D}$	1,30E+01	1,87E+00	6,92E+00	
$\beta_{browwrink\_O}$	6,25E+00	1,58E+00	3,95E+00	
$\beta_{browwrink\_SA}$	2,31E+00	1,89E+00	1,22E+00	*
β_{eth1_D}	-3,86E-01	1,28E-01	-3,02E+00	
β_{eth1_O}	-3,95E-01	1,16E-01	-3,40E+00	
β_{eth1_SA}	2,04E-01	7,80E-02	2,61E+00	
β_{eth2_A}	-1,19E+00	7,22E-01	-1,65E+00	*
β_{eth2_O}	-5,48E-01	4,06E-01	-1,35E+00	*
β_{eth3_D}	2,87E-01	1,91E-01	1,51E+00	*
β_{eth3_F}	-3,96E-01	2,55E-01	-1,55E+00	*
β_{eth3_H}	4,04E-01	1,49E-01	2,71E+00	
β_{eth3_O}	-5,19E-01	2,12E-01	-2,45E+00	
β_{eth3_SA}	2,03E-01	1,68E-01	1,21E+00	*
β_{eth4_A}	-2,11E-01	1,41E-01	-1,50E+00	*
β_{eth4_D}	-3,69E-01	1,76E-01	-2,09E+00	
β_{eth4_F}	-5,12E-01	2,02E-01	-2,54E+00	
β_{eth4_O}	-2,78E-01	1,59E-01	-1,75E+00	*
$\beta_{eye\_angle2\_l\_F}$	-1,57E+00	2,74E-01	-5,72E+00	
$\beta_{eye\_angle2\_r\_F}$	2,09E-01	3,32E-01	6,29E-01	*
$\beta_{eye\_angle2\_r\_SA}$	7,25E-01	1,44E-01	5,03E+00	
$\beta_{eye\_angle\_l\_A}$	-7,39E-01	3,59E-01	-2,06E+00	
$\beta_{eye\_angle\_l\_F}$	6,03E+00	5,22E-01	1,15E+01	
$\beta_{eye\_angle\_l\_SA}$	4,57E+00	3,16E-01	1,45E+01	
$\beta_{eye\_angle\_r\_A}$	3,47E+00	3,78E-01	9,18E+00	
$\beta_{eye\_angle\_r\_F}$	7,09E-01	4,42E-01	1,60E+00	*
$\beta_{eye\_angle\_r\_SA}$	-4,74E+00	3,07E-01	-1,54E+01	
$\beta_{eyeBrow\_angle\_l\_F}$	7,05E+00	5,73E-01	1,23E+01	
$\beta_{eyeBrow\_angle\_l\_O}$	-2,54E+00	3,84E-01	-6,62E+00	
$\beta_{eyeBrow\_angle\_l\_SA}$	-9,74E-01	4,81E-01	-2,02E+00	

$\beta_{eyeBrow\_angle\_r\_F}$	-1,77E+00	7,51E-01	-2,36E+00	
$\beta_{eyeBrow\_angle\_r\_O}$	-2,01E+00	2,30E-01	-8,76E+00	
$\beta_{eyeBrow\_angle\_r\_SA}$	5,64E+00	4,74E-01	1,19E+01	
$\beta_{eyeBrow\_angle\_r\_SU}$	-1,27E+00	3,74E-01	-3,41E+00	
$\beta_{eyeMouth\_dist\_l2\_D}$	-4,12E+01	3,17E+00	-1,30E+01	
$\beta_{eyeMouth\_dist\_l2\_O}$	-1,12E+01	2,63E+00	-4,26E+00	
$\beta_{eyeMouth\_dist\_l\_F}$	3,82E+01	7,49E+00	5,09E+00	
$\beta_{eyeMouth\_dist\_l\_H}$	-1,41E+01	4,24E+00	-3,32E+00	
$\beta_{eyeMouth\_dist\_l\_O}$	-2,99E+01	4,28E+00	-6,97E+00	
$\beta_{eyeMouth\_dist\_l\_SA}$	-2,81E+01	5,40E+00	-5,21E+00	
$\beta_{eyeMouth\_dist\_r2\_D}$	2,89E+01	2,63E+00	1,10E+01	
$\beta_{eyeMouth\_dist\_r2\_O}$	8,63E+00	2,77E+00	3,12E+00	
$\beta_{eyeMouth\_dist\_r\_F}$	-3,72E+01	7,30E+00	-5,09E+00	
$\beta_{eyeMouth\_dist\_r\_H}$	-7,79E+01	4,26E+00	-1,83E+01	
$\beta_{eyeMouth\_dist\_r\_O}$	3,22E+01	4,00E+00	8,05E+00	
$\beta_{eyeMouth\_dist\_r\_SA}$	6,09E+01	4,83E+00	1,26E+01	
$\beta_{eyeNose\_dist\_l\_A}$	6,00E+00	6,92E+00	8,67E-01	*
$\beta_{eyeNose\_dist\_l\_D}$	9,16E+01	6,27E+00	1,46E+01	
$\beta_{eyeNose\_dist\_l\_F}$	6,42E+01	8,91E+00	7,20E+00	
$\beta_{eyeNose\_dist\_l\_O}$	9,46E+01	5,99E+00	1,58E+01	
$\beta_{eyeNose\_dist\_l\_SA}$	1,19E+02	6,57E+00	1,81E+01	
$\beta_{eyeNose\_dist\_r\_A}$	5,14E+01	8,43E+00	6,10E+00	
$\beta_{eyeNose\_dist\_r\_D}$	-9,59E+01	7,21E+00	-1,33E+01	
$\beta_{eyeNose\_dist\_r\_F}$	-7,15E+01	1,14E+01	-6,26E+00	
$\beta_{eyeNose\_dist\_r\_O}$	-1,11E+02	7,04E+00	-1,57E+01	
$\beta_{eyeNose\_dist\_r\_SA}$	-1,38E+02	7,50E+00	-1,83E+01	
$\beta_{forehead\_A}$	1,65E-01	8,74E-02	1,89E+00	*
$\beta_{forehead\_F}$	8,99E-01	8,61E-02	1,04E+01	
$\beta_{forehead\_O}$	2,87E-01	6,61E-02	4,34E+00	
$\beta_{forehead\_SU}$	5,47E-01	6,83E-02	8,00E+00	
β_{form1_D}	9,71E-02	6,84E-02	1,42E+00	*
β_{form1_F}	4,61E-01	1,09E-01	4,23E+00	

β_{form1_H}	2,18E-01	8,57E-02	2,55E+00	
β_{form1_SA}	2,53E-01	6,30E-02	4,02E+00	
β_{form2_F}	3,21E-01	8,21E-02	3,91E+00	
β_{form2_H}	1,17E-01	6,36E-02	1,84E+00	*
β_{form2_O}	-5,97E-02	4,43E-02	-1,35E+00	*
β_{gender_D}	-6,11E-02	4,92E-02	-1,24E+00	*
β_{gender_O}	-1,31E-01	4,41E-02	-2,98E+00	
β_{gender_SA}	-2,26E-01	4,41E-02	-5,12E+00	
β_{lan1_F}	-1,25E-01	6,71E-02	-1,86E+00	*
β_{lan1_H}	-1,26E-01	5,69E-02	-2,21E+00	
β_{lan1_O}	1,81E-01	5,37E-02	3,36E+00	
β_{lan1_SA}	-1,11E-01	4,41E-02	-2,53E+00	
β_{lan1_SU}	2,51E-01	5,61E-02	4,46E+00	
β_{lan3_A}	-2,43E-01	6,79E-02	-3,58E+00	
β_{lan3_D}	1,67E-01	6,21E-02	2,69E+00	
β_{lan3_H}	-3,13E-01	8,22E-02	-3,81E+00	
β_{lan3_O}	2,87E-01	6,61E-02	4,34E+00	
β_{lan3_SU}	4,04E-01	7,15E-02	5,65E+00	
$\beta_{eye\_height\_I\_F}$	-8,73E+01	1,12E+01	-7,82E+00	
$\beta_{eye\_height\_I\_H}$	-3,28E+01	6,80E+00	-4,83E+00	
$\beta_{eye\_height\_I\_SU}$	-6,77E+01	7,59E+00	-8,92E+00	
β_{loc1_A}	1,79E-01	1,12E-01	1,60E+00	*
β_{loc1_D}	1,82E-01	1,07E-01	1,70E+00	*
β_{loc1_H}	1,48E-01	5,26E-02	2,81E+00	
β_{loc1_O}	2,95E-01	1,01E-01	2,92E+00	
β_{loc1_SU}	6,84E-02	4,80E-02	1,43E+00	*
β_{loc2_A}	1,61E-01	1,13E-01	1,42E+00	*
β_{loc2_D}	1,13E-01	1,08E-01	1,04E+00	*
β_{loc2_O}	1,58E-01	1,03E-01	1,53E+00	*
$\beta_{mouth\_height\_A}$	-2,76E+01	3,09E+00	-8,93E+00	
$\beta_{mouth\_height\_D}$	-4,20E+00	1,72E+00	-2,43E+00	
$\beta_{mouth\_height\_F}$	4,30E+01	1,64E+00	2,62E+01	

$\beta_{mouth\_height\_H}$	-1,64E+00	2,84E+00	-5,79E-01	*
$\beta_{mouth\_height\_SA}$	-9,05E+00	2,45E+00	-3,70E+00	
$\beta_{mouth\_height\_SU}$	5,52E+01	1,45E+00	3,82E+01	
$\beta_{mouthNose\_dist2\_A}$	8,62E+00	2,63E+00	3,28E+00	
$\beta_{mouthNose\_dist2\_SA}$	-1,41E+01	2,27E+00	-6,20E+00	
$\beta_{mouthNose\_dist\_D}$	1,43E+01	3,64E+00	3,92E+00	
$\beta_{mouthNose\_dist\_H}$	3,12E+01	2,41E+00	1,29E+01	
$\beta_{mouth\_width\_A}$	2,38E+01	2,31E+00	1,03E+01	
$\beta_{mouth\_width\_D}$	3,01E+01	1,97E+00	1,52E+01	
$\beta_{mouth\_width\_F}$	1,92E+01	2,37E+00	8,12E+00	
$\beta_{mouth\_width\_H}$	1,02E+02	2,76E+00	3,69E+01	
$\beta_{mouth\_width\_O}$	1,92E+01	2,06E+00	9,31E+00	
$\beta_{mouth\_width\_SA}$	-5,10E+00	2,21E+00	-2,31E+00	
$\beta_{nasolabial\_D}$	7,68E-01	5,91E-02	1,30E+01	
$\beta_{naswrink\_D}$	1,89E+01	7,08E-01	2,67E+01	
$\beta_{naswrink\_O}$	3,82E+00	7,96E-01	4,79E+00	
β_{reg1_D}	1,63E+00	4,29E-01	3,79E+00	
β_{reg3_D}	7,61E-01	2,69E-01	2,82E+00	
β_{reg3_O}	6,67E-01	2,88E-01	2,31E+00	
β_{reg4_A}	5,30E-01	1,78E-01	2,99E+00	
β_{reg4_D}	1,02E+00	2,57E-01	3,97E+00	
β_{reg4_F}	4,50E-01	2,70E-01	1,67E+00	*
β_{reg4_H}	3,00E-01	2,18E-01	1,38E+00	*
β_{reg4_O}	1,48E+00	2,08E-01	7,08E+00	
β_{reg4_SU}	-4,62E-01	2,62E-01	-1,77E+00	*
β_{reg5_D}	8,63E-01	1,76E-01	4,91E+00	
β_{reg5_F}	-1,78E-01	1,24E-01	-1,44E+00	*
β_{reg5_O}	6,77E-01	1,39E-01	4,88E+00	
β_{reg6_D}	7,85E-01	1,98E-01	3,96E+00	
β_{reg6_F}	-1,89E-01	1,77E-01	-1,07E+00	*
β_{reg6_O}	7,46E-01	1,63E-01	4,58E+00	
β_{reg7_A}	-3,42E-01	2,09E-01	-1,64E+00	*

β_{reg7_D}	8,51E-01	2,57E-01	3,31E+00	
β_{reg7_O}	5,41E-01	2,26E-01	2,39E+00	
$\beta_{eye\_height\_r\_F}$	-3,41E+01	1,41E+01	-2,42E+00	
$\beta_{eye\_height\_r\_H}$	4,78E+01	5,31E+00	9,01E+00	
$\beta_{eye\_height\_r\_SU}$	3,21E+01	4,86E+00	6,60E+00	
$\beta_{science1\_H}$	2,63E-01	1,15E-01	2,29E+00	
$\beta_{science1\_O}$	-1,46E-01	9,21E-02	-1,59E+00	*
$\beta_{science2\_SA}$	1,48E-01	7,17E-02	2,06E+00	
$\beta_{science2\_SU}$	-1,47E-01	7,91E-02	-1,86E+00	*
$\beta_{science3\_A}$	-1,52E-01	5,90E-02	-2,58E+00	
$\beta_{science3\_D}$	-1,40E-01	5,89E-02	-2,38E+00	
$\beta_{science3\_H}$	1,21E-01	7,12E-02	1,70E+00	*
$\beta_{science3\_O}$	-4,94E-01	6,43E-02	-7,68E+00	
$\beta_{science3\_SU}$	-8,38E-02	5,73E-02	-1,46E+00	*
$\beta_{science4\_H}$	1,64E-01	1,05E-01	1,56E+00	*
$\beta_{science4\_O}$	-3,01E-01	9,46E-02	-3,18E+00	
$\beta_{science4\_SA}$	1,38E-01	9,53E-02	1,45E+00	*
$\beta_{science4\_SU}$	-2,23E-01	1,14E-01	-1,95E+00	*
$\beta_{science5\_A}$	-1,81E-01	5,41E-02	-3,34E+00	
$\beta_{science5\_H}$	6,47E-02	6,56E-02	9,87E-01	*
$\beta_{science5\_O}$	-1,63E-01	5,05E-02	-3,22E+00	
$\beta_{science5\_SA}$	7,64E-02	5,08E-02	1,51E+00	*

BIBLIOGRAPHY

[1] G. Antonini, M. Sorci, M. Bierlaire and J.P. Thiran (2006), "Discrete Choice Models for Static Facial Expression Recognition", in *Advanced Concepts for Intelligent Vision Systems, 8th International Conference, ACIVS 2006*, Berlin, Springer-Verlag, vol. 4179, pp. 710-721.

[2] G. Antonini, S. Venegas, M. Bierlaire and J.P. Thiran (2006), "Behavioral Priors for Detection and Tracking of Pedestrians in Video Sequences", *International Journal of Computer Vision* , Hingham, MA, USA, Kluwer Academic Publishers, vol. 69, no. 2, pp. 159-180.

[3] G. Antonini, C. Gioia, E. Frejinger and M. Thémans (2006), *Discrete Choice Analysis: Predicting Demand and Market Shares - Case Study Workbook*.

[4] J.N. Bassili (1978), "Facial Motion in the Perception of Faces and of Emotional Expression", *Journal of Experimental Psychology. Human Perception and Performance*, vol. 4, no. 3, pp. 373-379.

[5] , M.E. Ben-Akiva and S.R. Lerman (1985), *Discrete Choice Analysis: Theory and Application to Travel Demand*, Cambridge, UK, MIT Press.

[6] M. Bierlaire (2003), "BIOGEME: A Free Package for the Estimation of Discrete Choice Models", in *Proceedings of the 3rd Swiss Transportation Research Conference, Ascona, Switzerland*.

[7] M. Bierlaire (2005), *An Introduction to BIOGEME Version 1.4*, biogeme.epfl.ch.

[8] N.G. Blurton-Jones (1971), "Criteria for Use in Describing Facial Expressions in Children", *Human Biology; an International Record of Research*, vol. 43, no. 3, pp. 365-413.

[9] C. Darwin (1872), *The Expression of the Emotions in Man and Animals*, London, UK, John Murray.

[10] I.L. Dryden and K.V. Mardia (1998), *Statistical Shape Analysis*, New York, USA, John Wiley & Sons.

[11] G.J. Edwards, C.J. Taylor and T. Cootes (1997), "Learning to Identify and Track Faces in Image Sequences", in *BMVC '97: Proceedings of the British Machine Vision Conference, University of Essex, UK*, British Machine Vision Association, pp. 130-139.

[12] G.J. Edwards, C.J. Taylor and T.F. Cootes (1998a), "Face Recognition Using Active Appearance Models", in *ECCV '98: Proceedings of the 5th European Conference on Computer Vision*, London, UK, Springer-Verlag, vol. 2, pp. 581-595.

[13] G.J. Edwards, C.J. Taylor and T. Cootes (1998b), "Interpreting Face Images using Active Appearance Models", in *FG '98: Proceedings of the 3rd International Conference on Face and Gesture Recognition, Nara, Japan*, Washington, DC, USA, IEEE Computer Society, pp. 300-305.

[14] P. Ekman and W.V. Friesen, (1971), "Constants across Cultures in the Face and Emotion", *Journal of Personality and Social Psychology*, vol. 17, no. 2, pp. 124-129.

[15] P. Ekman and W.V. Friesen (1975), *Unmasking the Face*, Englewood Cliffs, New Jersey, Prentice Hall.

[16] P. Ekman and W.V. Friesen (1978), *Facial Action Coding System (FACS): Manual*, Palo Alto, Consulting Psychologists Press.

[17] P. Ekman, W.V. Friesen and J.C. Hager (2002), *Facial Action Coding System (FACS): Investigator's Guide*, Salt Lake City, Utah, A Human Face.

[18] B. Fasel and J. Luettin (2003), "Automatic facial expression analysis: A survey", *Pattern Recognition*, vol. 36, no. 1, pp. 259-275.

[19] A.J. Fridlund, P. Ekman and H. Oster (1987), "Facial Expressions of Emotion: Review Literature 1970-1983", in A.W. Siegman and S. Feldstein, eds., *Nonverbal Behavior and Communication*, Hillsdale, NJ, Lawrence Erlbaum Assoc., pp. 143-224.

[20] J.C. Hager (1999), *Asymmetries in Facial Actions*, <http://face-and-emotion.com/dataface/dissertation/frontpage.html>.

[21] C.H. Hjortsjo (1970), *Man's Face and Mimic Language*, Sweden, Lund, Student-litteratur.

[22] H. Hong, H. Neven and C. von der Malsburg (1998), "Online Facial Expression Recognition based on Personalized Gallery", in *FG '98: Proceedings of the 3rd International Conference on Automatic Face and Gesture Recognition, Nara, Japan*, Washington, DC, USA, IEEE Computer Society, pp. 354-359.

[23] C.L. Huang and Y.M. Huang (1997), "Facial Expression Recognition Using Model-Based Feature Extraction and Action Parameters Classification", *Journal of Visual Communication and Image Representation*, vol. 8, no. 3, September 1997, pp. 278-290.

[24] T. Kanade, J.F. Cohn and Y.L. Tian (2000), "Comprehensive Database for Facial Expression Analysis", in *FG '00: Proceedings of the 4th IEEE International Conference on Automatic Face and Gesture Recognition, Grenoble, France*, Washington, DC, USA, IEEE Computer Society, pp. 46-53.

[25] S. Kimura and M. Yachida (1997), "Facial Expression Recognition and Its Degree Estimation", in *CVPR '97: Proceedings of the 1997 Conference on Computer Vision and Pattern Recognition*, Washington, DC, USA, IEEE Computer Society, pp. 295-300.

[26] H. Kobayashi and F. Hara (1997), "Facial Interaction between Animated 3D Face Robot and Human Beings", in *SMC '97: Proceedings of the*

IEEE International Conference on Systems, Man, Cybernetics, vol. 4, pp. 3732-3737.

[27] C. Lawrence, J. Zhou and A. Tits (1997), *User's Guide for CFSQP Version 2.5: A C Code for Solving (Large Scale) Constrained Nonlinear (Minimax) Optimization Problems, Generating Iterates Satisfying All Inequality Constraints*, Technical Report TR-94-16r1, Institute for Systems Research, University of Maryland, College Park, Maryland.

[28] Y. Liu, K.L. Schmidt, J.F. Cohn and R.L. Weaver (2003), "Facial Asymmetry Quantification for Expression Invariant Human Identification", *Computer Vision and Image Understanding Journal*, vol. 91, no. 1/2, pp. 138-159.

[29] M.J. Lyons, J. Budynek and S. Akamatsu (1999), "Automatic Classification of Single Facial Images", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 12, pp. 1357-1362.

[30] K. Mase and A. Pentland (1991), "Recognition of Facial Expression From Optical Flow", *IEICE Transactions*, vol. E74, no. 10, (October), pp. 3474-3483.

[31] D. McFadden (1978), "Modelling the Choice of Residential Location", in A. Karlqvist, L. Lundqvist, F. Snickars, and J. Weibull, eds., *Spatial Interaction Theory and Residential Location*, Amsterdam, North-Holland, pp. 75-96.

[32] P. Meer and B. Georgescu (2001), "Edge Detection with Embedded Confidence", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 12, (December), pp. 1351-1365.

[33] C. Padgett and G.W. Cottrell (1997), "Representing Face Images for Emotion Classification", in *NIPS: Advances in Neural Information Processing Systems, Denver, CO, USA, 1996*, Cambridge, UK, MIT Press, vol. 9, pp. 894-900.

[34] M. Pantic and L.J.M. Rothkrantz (2000a), "Expert System for Automatic Analysis of Facial Expression", *Image and Vision Computing Journal*, vol. 18, no. 11, pp. 881-905.

[35] M. Pantic and L.J.M. Rothkrantz (2000b), "Automatic Analysis of Facial Expressions: The State of the Art", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 12, (December), pp. 1424-1445.

[36] M. Sorci, G. Antonini and J.P. Thiran (2007), *Facial Expression Evaluation Survey*, Technical report TR-ITS-2007-07, Ecole Polytechnique Federale de Lausanne, Lausanne, Switzerland.

[37] J. Steffens, E. Elagin and H. Neven (1998), "PersonSpotter - Fast and Robust System for Human Detection, Tracking and Recognition", in *FG '98: Proceedings of the 3rd International Conference on Face and Gesture Recognition, Nara, Japan*, Washington, DC, USA, IEEE Computer Society, pp. 516-521.

[38] M.B. Stegmann and D.D. Gomez (2002), *A Brief Introduction to Statistical Shape Analysis*, Technical Report, Informatics and Mathematical Modelling, Technical Univ. of Denmark, DTU.

[39] K. Train (2003), *Discrete Choice Methods with Simulation*, Cambridge, UK, Cambridge University Press.

[40] L. Wiskott (1995), *Labelled Graphs and Dynamic Link Matching for Face Recognition and Scene Analysis*, PhD thesis, Ruhr-Universitat Bochum, vol. 53 of Reihe Physik, Thun, Frankfurt am Main, Germany, Verlag Harri Deutsch.

[41] J. Yang and A. Waibel (1996), "A Real-time Face Tracker", in *WACV '96: Proceedings of the 3rd IEEE Workshop on Applications of Computer Vision*, Washington, DC, USA, IEEE Computer Society, pp. 142-147.

[42] M. Yoneyama, Y. Iwano, A. Ohtake and K. Shirai (1997), "Facial Expressions Recognition Using Discrete Hopfield Neural Networks", in *ICIP '97:*

Proceedings of the 1997 International Conference on Image Processing, Washington, DC, USA, IEEE Computer Society, vol. 1, pp. 117-120.

[43] Z. Zhang, M. Lyons, M. Schuster and S. Akamatsu (1998), "Comparison between Geometry-Based and Gabor Wavelets-Based Facial Expression Recognition Using Multi-Layer Perceptron", in *FG '98: Proceedings of the 3rd International Conference on Face and Gesture Recognition, Nara, Japan*, Washington, DC, USA, IEEE Computer Society, pp. 454-459.

[44] Y. Zhang and Q. Ji (2005), "Active and Dynamic Information Fusion for Facial Expression Understanding from Image Sequences", *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 27, no. 5, pp. 699-714.

[45] J. Zhao and G. Kearney (1996), "Classifying Facial Emotions by Backpropagation Neural Networks with Fuzzy Inputs", in *Proceedings of the International Conference on Neural Information Processing*, vol. 1, pp. 454-457.