Modelling human perception of facial expressions

M. Sorci [*] T. Robin [†] J. Cruz [†] M. Bierlaire [†]
J.-P. Thiran [*] G. Antonini [‡]

July 6, 2009

Report TRANSP-OR 090706
Transport and Mobility Laboratory
School of Architecture, Civil and Environmental Engineering
Ecole Polytechnique Fédérale de Lausanne
transp-or.epfl.ch

[*]LTS5, Signal Processing Laboratory, Ecole Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland, {matteo.sorci, jp.thiran}@epfl.ch

[†]Transport and Mobility Laboratory, Ecole Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland, {michel.bierlaire, thomas.robin, javier.cruz}@epfl.ch

[‡]IBM Zürich research laboratory, IBM, CH-8803 Rüschlikon, Switzerland, {GAN}@zurich.ibm.com

Abstract

Facial expression recognition by human observers is affected by subjective components. Indeed there is no ground truth. We have developped Discrete Choice Models to capture the human perception of facial expressions. In a first step, the static case is treated, that is modelling perception of facial images. Image information is extracted using a computer vision tool called Active Appearance model (AAM). DCMs attributes are based on the Facial Action Coding System (FACS), Expressions Descriptive Units (EDU) and outputs of AAM. Some behavioral data have been collected using an internet survey, where respondents are asked to label facial images from the Cohn-Kanade database with expressions. Different models were estimated by likelihood maximization using the obtained data. In a second step, the proposed static discrete choice framework is extended to the dynamic case, which considers facial video instead of images. The model theory is described and another internet survey is currently conducted in order to obtain expressions labels on videos. In this second internet survey, videos come from the Cohn-Kanade database and the Facial Expressions and Emotions Database (FEED).

1 Introduction

Facial expressions are one of the most visual method to convey emotions and one of the most powerful means used by human beings to relate to each other. In order to move towards real interacting human-computer systems, where algorithms written by humans should be able to capture, mimic and reproduce human perceptions, facial expressions play surely a central role. One of the key issues to consider in building such systems is the definition of facial expression measurements to study and quantify facial behaviour. The two major approaches in psychological research are *message and sign judgement* (Cohn, 2006). The task of message judgement is the inference of the displayed facial behaviour, in terms of inferred emotion. As indicated by Cohn, 2006, among the different descriptors those of Ekman, 1992 have been largely used in the recent past. Ekman proposed the use of the 6 basic emotions (happiness, surprise, fear, disgust, sadness and anger) that are universally displayed and recognized from facial expressions (Keltner, 2000). In sign judgement approaches the displayed facial behaviour is described by facial movements. Among the various methods the

Facial Action Coding Systems (FACS) (Ekman and Friesen, 1978, Ekman et al., 2002) is the most comprehensive and widely used. The FACS is a human-observed based system designed to detect subtle changes in facial features and associates facial expression changes with actions of the muscles that produce them. Thus, a nasolabial furrow, running down from the nostrils outward beyond the corners of the lips, can be judged as "sadness" in a message-judgement and as a facial movement that raises the cheeks in a sign-judgement approach. In other terms, while message judgement is all about interpretation, sign judgement attempts to be objective.
In this work we focus and propose an automatic approach belonging to the the family of message judgement based system. The dominant challenge in building such an automatic system, even if narrowed down to the facial expression perception task of message judgement, arises from the fact that such a perception (performed by human beings in the real world) is subjective and strongly related to contextual information.
A typical automatic facial expressions recognition system (Tian et al., 2003, M. and Bartlett, 2007, Fasel and Luettin, 2003) is based on a representation of each expression, learned from a training set of pre-selected meaningful features. In the learning process, an expert is asked to associate labels to training samples. An expert should be someone having a *strong knowledge* of the problem, in order to ensure the correctness of what we are trying to reproduce.
Three important questions arise from this fundamental hypothesis of "learning by examples" technique:

- Can one expert be representative of humans'perception?
- How to get and use the experts'strong knowledge?
- How to represent the visual information used by the experts?

The outstanding human ability to identify individual human faces has long been of major interest to cognitive scientists, neuropsychologists, and neuroscientists (Diamond and Carey, 1986, Carey, 1992, Moses et al., 1996). Whereas 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, 1978, 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).

Whereas sign judgement systems are completely insensitive to context and familiarity with the face, the message based ones are strongly influenced by them. This consideration leads to the answer to the first question: in a message based framework the judgement of one human is not enough to reproduce and capture the different behaviours of humans. In support of this last statement and in order to answer to the second question, the data collected by a web-based static facial expression evaluation survey, developed by the authors (Sorci et al., 2007) and described in Section 3, shows the need for a model capable of taking into account the heterogeneity in human's perception of facial expressions. Figure 1 shows two images of the survey and the histograms of the 33 participants that have annotated them. These are two typical examples of how heterogeneity (Figures 1(a)-1(b)) and homogeneity (Figures 1(c)-1(d)) can both be present in human's judgement. Concerning the last question, most recent attempts in the representation of visual information for facial expression have focused on reproducing the set of rule descriptors suggested by the FACS system. Based on this system, a facial expression can be linguistically described in terms of measures that can be extracted from the face. These measures can be considered as the mathematical representations of local facial features. In the last decade, works on psychophysics and cognitive psychology (Farah et al., 1998, Schwaninger et al., 2002, Cabeza and Kato, 2000, Meulders et al., 2005) have shown that face recognition and perception of emotions rely on featural[1] and configural[2] information. Human's visual perception of a face involves the processing of both local facial measures and their holistic spatial layout. The implication of these findings is that an automatic system, aiming at interpreting faces, needs to extract and make use of these two sources of information as well.

The objective of this work is to propose novel models to describe and reproduce the evaluation of humans, considered as an heterogeneous population, facing the task of labelling static facial expressions. The labelling process is a decision making process where individuals choose a categorized expres-

[1]facial featural features represent local measures of facial components

[2]facial configural features represent the holistic spatial layout of facial components

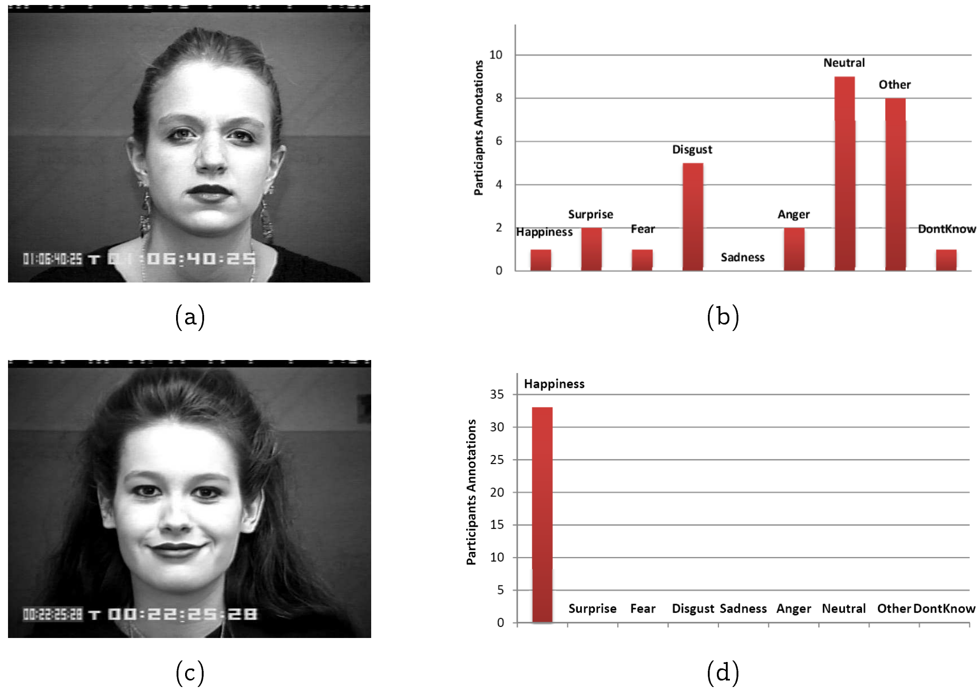


Figure 1: Examples of heterogeneous and homogeneous judgements in the data collected by the survey.a-b)Image of an ambiguous expression and histogram of participants annotations;c-d)Image of a happiness expression, unanimously perceived by the participants.

sion among a set of 9 different options: *happiness, surprise, fear, disgust, sadness, anger, neutral, other* and *I don't know*.

Discrete Choice Models (DCM) (Ben-Akiva and Lerman, 1985, Manski, 1977, Manski and McFadden, 1981) well fit our needs and they represent a reasonable and theoretically grounded modelling framework. DCMs are econometric models designed to forecast the behaviour of individuals in choice situations, when the set of available alternatives is finite and discrete. Our idea is to approach the decision making process through the *rational behaviour paradigm*, representing the logic behind the DCMs and well matching the evaluation process of the human observer. Three main factors will lead us in the development of a good model: 1)a strong a priori knowledge of the problem; 2)realistic annotations from an heterogeneous population of humans; 3)a reliable set of features. The contributions of this work can be summarized as follows:

- we propose the use of discrete choice models for modelling the human perception of static facial expression;
- we develop 3 models of increased complexity;
- we show how measures extrapolated by the FACS can be combined with two new sets of features to complete the characterization of each expression and improve the descriptiveness of the model;
- we have extended the discrete choice framework for static facial expression perception to a dynamic version, which consists in considering videos instead of images.

The remainder of the paper is organized as follows: in the next section, we present an overview of the existing works and identify the limitations and differences with ours. In Section 3, we describe the facial expression survey we have developed to provide the data used in this work. Section 4 introduce the methodological framework, while Section 5 details the feature sets used in our model and the associated methods. In Sections 6,7 and 8 we describe, respectively, the model specification, the estimation of the related parameters and the extension to dynamic facial expression recognition. We end in Section 9 with discussions and conclusions.

2 Previous Work

The current research on facial expression analysis is mostly oriented in two main directions: recognition of prototypic emotional expression and recognition of facial action units. The first aims to a categorical representation of the six universal basic emotions. The second does not attempt to give an interpretation of the expression, but it focuses on the detection of atomic facial signals. The interpretation can be delegated to higher order decision making.

The two approaches are strictly related to the two main streams in psychological research: message and sign judgement. Most of the available literature on both approaches proposes a three step procedure in order to make the problem operational: face detection, facial features extraction and facial changes recognition (prototypic emotions or action units).

Face detection is a problem studied since the very begining times of computer vision. It consists of determining all the regions of the scene under analysis that contain a face. In order to achieve that, a wide variety of works can be found on the literature (Pentland et al., 1994, Rowley et al., 1998, Sung and Poggio, 1998, Schneiderman and Kanade, 2000) but probably the most commonly used nowadays is tha face detector introduced by Viola and Jones, 2004. This detector is based on a cascade of classifiers trained with the AdaBoost algorithm (Freund and Schapire, 1997) and the use of the integral image, which makes the method able to run in real-time. A survey on the topic can be found in (Yang et al., 2002) or in Chapter 8 of Medioni and Kang, 2004.

Once faces are detected, features from these faces need to be extracted. These features can be divided into geometric features and appearance features. Geometric features are featural descriptors of the face that represent it in terms of shape and locations of the main facial components (mouth, eyes, nose, etc.). Some recent examples of geometric features extraction can be found in Hu et al., 2004, Pantic and Patras, 2006 or Valstar and Pantic, 2007. With respect to appearance features, they are configural or featural descriptors of the face that represent it in terms of facial texture, including wrinkles, bulges and furrows. Some recent examples of these techniques can be found in Ye et al., 2004, Chang et al., 2004 or Bartlett et al., 2006. Hybrid techniques can also be found in the literature, as for example the approach of Zhang and Ji, 2005, that uses 26 landmarks around the main facial components as well as the transient features, like wrinkles and furrows.

Finally, in the third step, all the information extracted from the face has to be associated with a facial expression, or an action unit, by means of a decision or classification rule. A wide variety of approaches can be found on the literature using a broad range of machine learning techniques: Neural Networks (NN) (Zhang et al., 1998, Padgett and Cottrell, 1998, li Tian et al., 2001, li Tian et al., 2002), Bayesian classifiers (Cohen et al., 2003), Linear Discriminant Analysis (LDA) (Abboud and Davoine, 2004), Hidden Markov Models (HMM) (Cohen et al., 2003) or Support Vector Machines (SVM) (Valstar and Pantic, 2007), for mentioning some of them. Recently, the authors introduced in Antonini et al., 2006 the use of Discrete Choice Models (DCM) for static facial expression classification.

2.1 Limitations of Previous Approaches

Current works on facial expression understanding, in our view, suffer from the following shortcomings:

1. The main paradigm of standard classification approaches, in the context of message judgement frameworks, consists in associating any two examples having the same features to the same corresponding class. One of the main assumptions is that facial expression labels, reported in the training set, represent the true expressions. As underlined by the example in Figure 1, this assumption does not hold in modelling human's perception static facial expression. Indeed, facial expressions are ambiguous and different people might perceive differently the same expression. This fact is even more accentuated in a static context, where the lack of transitions between following expressions deprives the observer of an important source of information. A probabilistic approach is more suitable in this case.

2. Another limitation of most previous approaches, concerns the inability to interpret knowledge acquired by the systems. In other words, their black-box nature prevent any interpretation about the relations between the inputs and outputs of the model. For the same reason, it is also impossible to gain any understanding of the problem at hand or to incorporate human expertise to simplify, accelerate and improve the modelling process.

3. The integration of featural and configural facial features provides crucial cues in the human interpretation of an expression. Besides the work of Zhang and Ji, 2005, more complex hybrid system have not been investigated rigorously by the existing works.

To overcome the above limitations, we propose the use of DCMs and the introduction of new sets of features. The proposed probabilistic approach allows to:

- model the possible ambiguities in human perception of static facial expressions;

- enable the analyst to exploit her knowledge of the problem;

- improve the descriptiveness of a face by introducing a more complete set of featural and configural features.

3 Data collection

Construction of a good database of facial expressions requires time and training of subjects. Only a few of such databases are available, such as the Cohn-Kanade Database (Kanade et al., 2000), JAFFE (Lyons et al., 1998) and most recently the MMI database (Pantic et al., 2005). The images used in the survey come from the Cohn-Kanade Database.

3.1 Cohn-Kanade database

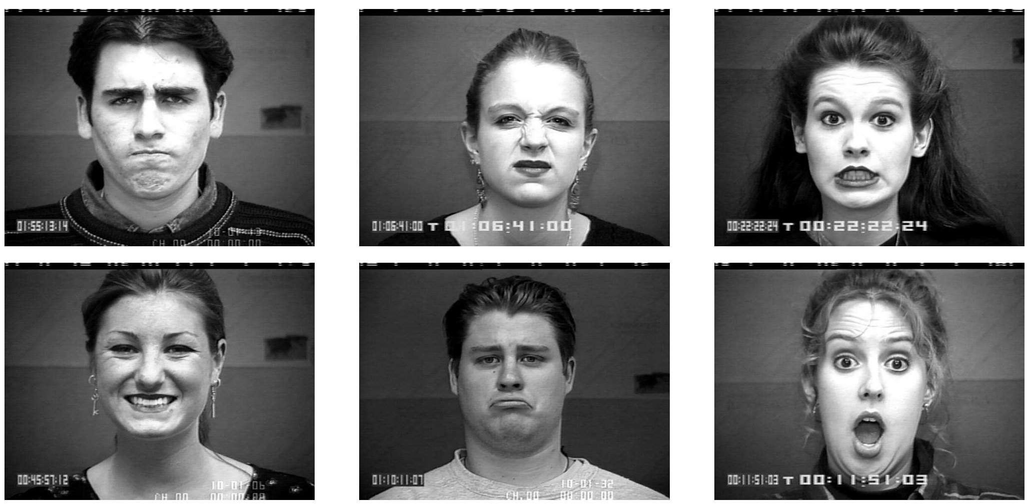


Figure 2: *Examples of faces in the Cohn-Kanade Database.*

The Cohn-Kanade Database consists of image sequences of expressions, starting from a neutral expression and ending most of the time in the peak of the facial expression. The 104 subjects of the database are university students enrolled in introductory psychology classes. They ranged in age from 18 to 30 years. 65 percent were female, 15 percent were African-American, and three percent were Asian or Latino. Subjects were instructed by an experimenter to perform a series of 23 facial displays. Six of the displays were based on descriptions of prototypic emotions (i.e, happiness, anger, fear, disgust, sadness and surprise). Before performing each display, an experimenter described and modelled the desired display.

3.2 Facial expressions evaluation survey

In August 2006, Sorci et al., 2007 published the internet facial expressions evaluation survey in order to find a way to directly get humans' perception of facial expressions (`http://lts5www.epfl.ch/face`). The aim of the survey is to collect a dataset created by a sample 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.). The images used in the survey comes from the Cohn-Kanade Database.Over the 104 subjects in the database, only 11 of them gave the consent for publication. The subset of the Cohn-Kanade Database used in this survey consists of the 1271 images of these 11 subjects (9 women and 2 men). The annotation

(a) (b) .

Figure 3: On-line survey interface: a)Socio-economic form;b)Image annotation interface

process consists in associating an expression label (among a set of available human expressions) to each image that is presented to the survey's participant. A simple and intuitive interface has been designed in order to facilitate the annotation process 3. For each image in the group the participant has to choose one of the following options: happiness, surprise, fear, anger, disgust, sadness, "I don't know" and "Other". The last two options have been introduced in order to deal with images particularly ambiguous to the participant. In addition, these two options make the set exhaustive, in the sense that they permit to cover the whole range of human expressions. We should remind that in this work we deal with static perception of human expressions and with frames randomly chosen from small video sequences displaying the whole dynamic of the performed expression. The lack of temporal factor, in the labelling process, makes the annotation task difficult and subjective in some cases.

4 Discrete choice analysis: a behavioural modelling framework

Discrete choice models are known in econometrics since the late 50's. They are designed to describe the behavior of people in choice situations, when the set of available alternatives is finite and discrete (choice set). They are based on the concept of *utility maximization* in economics, where the decision maker is assumed to be *rational*, performing a choice in order to maximize the utilities she perceives from the alternatives. The alternatives are supposed to be mutually exclusive and collectively exhaustive, while the rationality of the decision maker implies transitive and coherent preferences. The utility is a *latent* construct, which is not directly observed by the modeler, and is treated as a random variable. The discrete choice paradigm matches well the labelling assignment process of the participants in the survey. This approach can be interpreted as an attempt to model the decision process performed by an hypothetical human observer during the labelling procedure for the facial expressions. Given a population of N individuals, the (random) utility function U_{in} perceived by individual n from alternative i, given a choice set C_n, is defined as follows:

$$U_{in} = V_{in} + \varepsilon_{in} \tag{1}$$

It is composed by the sum of a deterministic term V_{in}, capturing the systematic behaviour (features extracted from a face), and a random term ε_{in}, capturing the uncertainty. This random term captures unobserved attributes, unobserved individual characteristics, measurement errors and instrumental variables. We actually do not observe the real values of the utilities as perceived by the participant. Under the utility maximization assumption, the output of the model is represented by the choice probability that individual n will choose alternative i, given the choice set C_n. It is given by:

$$P_n(i|C_n) = P_n(U_{in} \geq U_{jn}, \forall j \in C_n) = \int_{\varepsilon_n} I(\varepsilon_n < V_{in} - V_{jn}, \forall j \in C_n, j \neq i) f(\varepsilon_n) d\varepsilon_n \qquad (2)$$

where $\varepsilon_n = \varepsilon_{jn} - \varepsilon_{in}$ and $I(.)$ is an indicator function which is equal to 1 when its argument is satisfied, zero otherwise. In this paper we use a Multinomial Logit Model (MNL), which is largely the simplest and most used discrete choice model in literature. The MNL choice probability is given by the following expression

$$P_n(i|C_n) = \frac{e^{\mu V_{in}}}{\sum\limits_{j \in C_n} e^{\mu V_{jn}}} \qquad (3)$$

In this work the choice set C_n is represented by the 9 survey alternatives ("happiness", "surprise", "fear", "disgust", "sadness", "anger", "neutral", "other" and "I don't know").

5 Explanatory variables

The survey provides the raw data capturing the participants perception of facial expressions. This raw data consists on a set of facial expressions images (the Cohn-Kanade images) and the set of participants choices. In order to exploit the information coming from both sources we need to identify and represent the facial visual cues describing an expression. The Facial Action Coding Systems (FACS) (Ekman and Friesen, 1978) represents the leading standard for measuring facial expressions in behavioural science. The main measures suggested by this human observer system represent a valid starting point in the quest of variables characterizing the different expressions. In the rest of the paragraph we detail the set of explanatory variables induced by the FACS and we introduce two new and complementary sets of visual measures aiming at improving the descriptiveness of each expression. Figure 4 schematically shows the image pre-processing steps necessary to

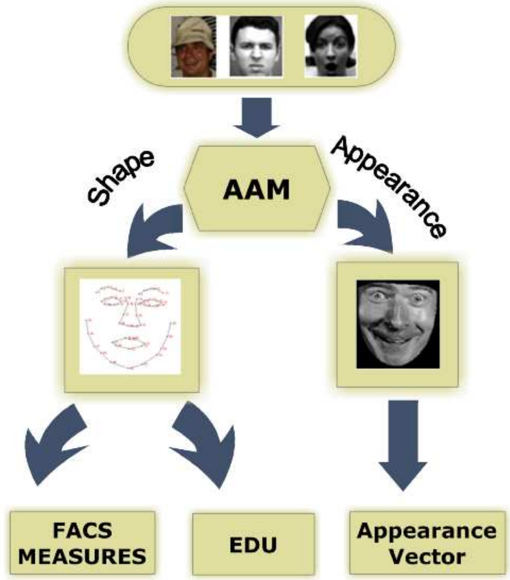
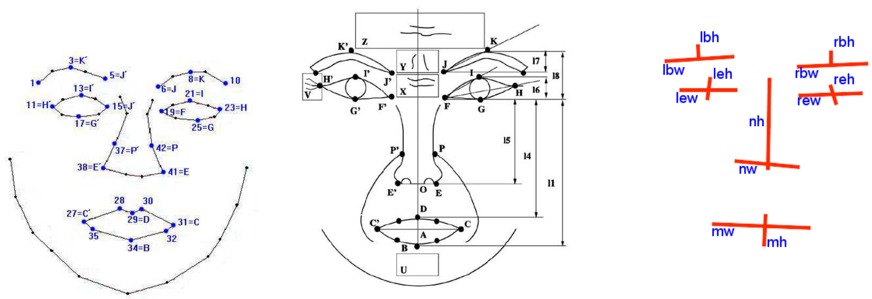


Figure 4: Schema of the image processing steps that lead to the extraction of the 3 sets of explanatory variables.

compute these 3 sets of explanatory variables. For that purpose, the AAM representation of the face, described in Section 5.1, is applied to the available 1271 images. The shape description of the face (Figure 5(a)) is used for computing both measures coming from the FACS (detailed in Section 5.2) and the new set of configural measures(Section 5.3), called Expression Descriptive Unit(EDU), complementing Ekman's ones. Since both holistic features and local features are important from the human perceptual point of view (Schwaninger, 2003, Cabeza and Kato, 2000, Wallraven et al., 2005, Bicego et al., 2007), a third set of measures representing the appearance of the face has been introduced(Section 5.4).

5.1 Active Appearance model

The Active Appearance Model (AAM) is a statistical method for matching a combined model of shape and texture to unseen faces. The combination of a model of shape variation with a model of texture variation generates a statistical appearance model. The model relies on a set of annotated images. A training set of images is annotated by putting a group of landmark



(a) (b) (c)

Figure 5: a) Facial landmarks (55 points); b) the geometrical relationship of facial feature points, where the rectangles represent the regions of furrows and wrinkles; c) Featural descriptors used in the definition of the EDUs;

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

Table 1: The association of six emotional expressions to AUs, AU combinations, and Transient Features

points around the main facial features, marked in each example. The shape is represented by a vector s brought into a common normalized frame -w.r.t. position, scale and rotation- to which all shapes are aligned. After having computed the mean shape $\bar{s}$ and aligned all the shapes from the training set by means of a Procrustes transformation (I.L. and K.V., 1998), it is possible to warp textures from the training set onto the mean shape $\bar{s}$, in order to obtain shape-free patches. Similarly to the shape, after computing the mean shape-free texture $\bar{g}$, all the textures in the training set can be normalized with respect to it by scaling and offset of luminance values. PCA is applied to build the statistical shape and textures models:

$$s_i = \bar{s} + \Phi_s b_{si} \quad \text{and} \quad g_i = \bar{g} + \Phi_t b_{ti} \tag{4}$$

FACS Measures	Measures on mask 5(a)	Explanatory Variables
$\overline{JJ'}$	Dist(P5, P6)	EV_1^F
$\overline{JF}$	Dist(P6, P19)	EV_2^F
$\overline{J'F'}$	Dist(P5, P15)	EV_3^F
$\overline{KG} \equiv l8$	Dist(P8, P25)	EV_4^F
$\overline{K'G'}$	Dist(P3, P17)	EV_5^F
$\overline{GI} \equiv l6$	Dist(P25, P21)	EV_6^F
$\overline{G'I'}$	Dist(P13, P17)	EV_7^F
$\overline{PF}$	Dist(P19, P42)	EV_8^F
$\overline{P'F'}$	Dist(P15, P37)	EV_9^F
$\overline{FC}$	Dist(P19, P31)	EV_{10}^F
$\overline{F'C'}$	Dist(P15, P27)	EV_{11}^F
$\overline{FD} \equiv l4$	Dist(P25, P29)	EV_{12}^F
$\overline{F'D}$	Dist(P17, P29)	EV_{13}^F
$\overline{OD}$	$\mathrm{Dist}(\left(\frac{P39+P40}{2}\right), P29)$	EV_{14}^F
$\overline{OB}$	$\mathrm{Dist}(\left(\frac{39+40}{2}\right), 33)$	EV_{15}^F
$\overline{DB}$	Dist(P29, P33)	EV_{16}^F
$\overline{C'C}$	Dist(P27, P31)	EV_{17}^F
$\measuredangle FHJ$	$\measuredangle$P19P23P6	EV_{18}^F
$\measuredangle F'H'J'$	$\measuredangle$P15P11P5	EV_{19}^F
$\measuredangle HFI$	$\measuredangle$P23P19P21	EV_{20}^F
$\measuredangle H'F'I'$	$\measuredangle$P11P15P13	EV_{21}^F
$\measuredangle HGF$	$\measuredangle$P23P25P19	EV_{22}^F
$\measuredangle H'G'F'$	$\measuredangle$P15P17P11	EV_{23}^F
Nose Wrinkles 6(a)	Presence Detection	EV_{24}^F
Eyes Wrinkles 6(b)	Presence Detection	EV_{25}^F
Forehead Wrinkles 6(c)	Presence Detection	EV_{26}^F
Nasolabial Fold 6(d)	Presence Detection	EV_{27}^F

Table 2: *Correspondences between measures on masks 5(b) and 5(a)*

where s_i and g_i are, respectively, the synthesized shape and shape-free texture, Φ_s and Φ_t are the matrices describing the modes of variation derived from the training set, b_{si} and b_{ti} the vectors controlling the synthesized shape and shape-free texture. The unification of the presented shape and texture models into one complete appearance model is obtained by concatenating the vectors b_{si} and b_{ti} by means of normalizing matrix $\mathbf{W}_s$:

$$\mathbf{b}_i = \begin{pmatrix} \mathbf{W}_s \mathbf{b}_{si} \\ \mathbf{b}_{ti} \end{pmatrix} \tag{5}$$

and learning the correlations between them by means of a further PCA.

$$\mathbf{b_i} = \Phi_\mathbf{c}\mathbf{c_i} \tag{6}$$

where $\Phi_\mathbf{c}$ are the eigenvectors and $\mathbf{c_i}$ is a vector of appearance parameters allowing to simultaneously control both shape and texture.
The statistical model is then given by:

$$\mathrm{s_i} = \bar{\mathrm{s}} + \mathbf{Q_s}\mathbf{c_i} \quad and \quad \mathrm{g_i} = \bar{\mathrm{g}} + \mathbf{Q_t}\mathbf{c_i} \tag{7}$$

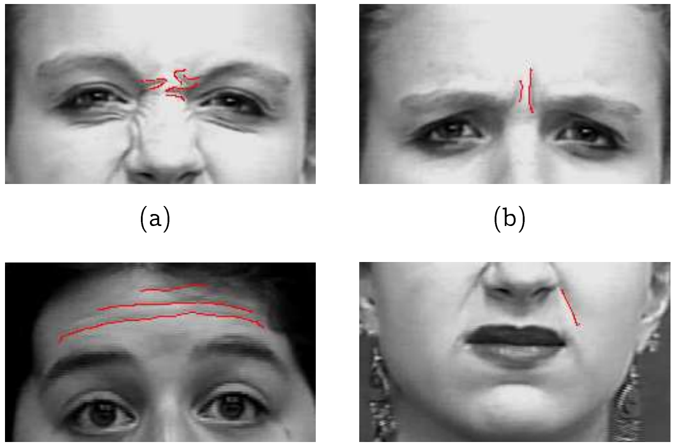
where Q_s and Q_t are the matrices describing the principal modes of the combined variations in the training set. Fixing the parameters c_i we derive the shape and the shape-free texture vectors using equations (7). A full reconstruction is given by warping the generated texture into the generated shape. In order to allow pose displacement of the model, other parameters must be added to the appearance parameters c_i: the pose parameters p_i. The matching of the appearance model to a target face can be treated as an optimization problem, minimizing the difference between the synthesized model image and the target face (Stegmann, 2000, Cootes et al., 2001, Cootes and Taylor, 2004, Matthews and Baker, 2004).

5.2 Measures from the FACS

Facial expressions represent a visible consequence of facial muscle and autonomic nervous system actions. Ekman and Friesen, 1978 propose the Facial Action Coding System (FACS) in order to measure all visible movements. Ideally, FACS would differentiate every change in muscular action, but it is limited to what a user can reliably discriminate. 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. FACS measurement units are called *Action Units*(AUs) and represent the muscular activity that produces momentary changes in facial appearance. A facial expression is indeed the combination of AUs. In particular, there are six basic emotions (happiness, anger, disgust, fear, surprise and sadness) that Keltner, 2000 postulated as having a distinctive content together with a unique facial expression. Based on the EMFACS (Friesen and Ekman, 1983) the 6 basic expressions can be described linguistically using

Ekman's AUs. Likewise, we adapt the AU-coded descriptions of facial expressions in the EMFACS in order to describe these 6 expressions. Table 1, which is directly adapted from Friesen and Ekman, 1983 and Friesen and Ekman, 1984, illustrates the facial AUs pertaining to the different expressions. By drawing on the work of Zhang and Ji, 2005, we group AUs of facial expressions as primary AUs and auxiliary AUs. The primary AUs refer to those AUs or combinations of AUs that univocally describe one of the 6 expressions. The auxiliary AUs provide an additional support to the expression characterization. This additional support can come from transient features, such as wrinkles and furrows, or from nontransient features, such as measures among facial components. In order to transform the AUs in a set of quantitatively measures Zhang and Ji translate these appearance changes descriptors in a set of geometrical relationships of some facial feature points, showed in Figure 5(b), and linguistically reported by Zhang and Ji, 2005. We use the shape mask, provided by the AAM, to measure the set of angles and distances detailed in Table 2. In the computation of these measures we need to take into account that there exists a large variance in the morphology of human faces. In order to deal with these differences a shape normalization is required. The AAM framework establishes *a coordinate reference* to which all the shapes are aligned by filtering out location, scale and rotational effects. The use of the alignment procedure on the detected masks ensures the computation of consistent measures.

On completion of the FACS system visual cues, we describe here the transient features and the measures used to quantify them. Transient wrinkles and furrows are the result of facial muscles movements. These movements produce small ridges in certain face regions. The regions of facial wrinkles and furrows are indicated by rectangles in Figure 5(b) and by the curves starting from P and P' for the nasolabial furrows. The change of wrinkles in the region X is directly related to AU9 (Nose Wrinkler). The furrows in the regions Z, Y, V, U provide diagnostic information for the identification of AU1 (Inner Brow Raiser), AU2 (Outer Brow Raiser), AU4 (Brow Lowerer), AU6 (Cheek Raiser), and AU17 (Chin Raiser), respectively. In order to detect these features, the edge detection with embedded confidence, proposed by Meer, Dec 2001, is used. The detection is successively refined by analysing the direction of the extracted edge. Referring to Figure 5(b), wrinkles in regions Z and X should be mostly horizontal while those in



(a) (b) (c) (d)

Figure 6: Transient feature detection: (a) vertical furrows between brows, (b) horizontal wrinkles between eyes, (c) horizontal wrinkles on the forehead, and (d) nasolabial fold.

region Y mostly vertical. Figure 6 shows examples of transient feature detection. The ratio between edge pixels (wrinkles) and background pixels (skin) is used to measure and detect the presence of wrinkles in regions X, Y and Z.
For the nasolabial furrows, the areas of interest are those reported in Figure 7(a). These regions, as well as all the other transient areas, are automatically detected using the AAM landmarks. Figure 7 shows the 4 possible configurations for the nasolabial region: nasolabial furrow absence, nasolabial furrow due to cheek raising Figures 7(b)-7(c) (AU6), nasolabial furrow due to nose wrinkling or upper lip raising Figure 7(d) (AU9,AU10). If the analysis of the longest connected edge in the 2 nasolabial regions (Figure 7(a)) reveals the presence of furrows, then the extracted curve is approximated by a quadratic equation: $y = ax^2 + bx + c$. The approximated curve is obtained by fitting the set of nasolabial furrow's pixels to y using the least-square method, similarly to Zhang and Ji, 2005. The a coefficient represents the curvature of the nasolabial fold. According to its value we can detect and encode the presence of the nasolabial furrows as follows: $a > 0$, as shown in Figure 7(b), which contributes to AU6 and to happiness-like expressions; $a < 0$ and the vertex $x = -b/2a$ is a pixel belonging to the detected furrow, as indicated by the red curve in Figure 7(c).

This instance is again connected to AU6; $a < 0$ and it has no vertex, as shown in Figure 7(d). This case is a support evidence to AU 9 and AU 10 and so to disgust-like expressions. The measures concerning regions V and U are discarded for two main reasons : 1)the related wrinkles are not always detectable in subjects; 2)they are redundant, since strictly linked to wrinkles and furrows in the retained regions.

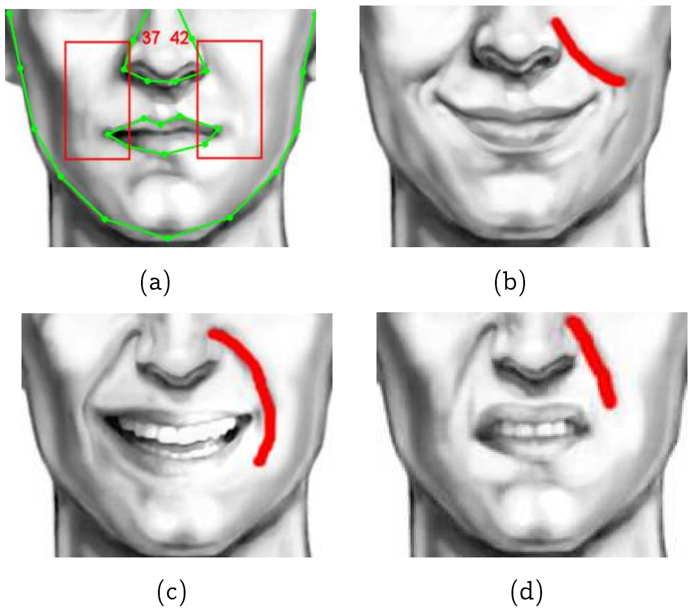


Figure 7: Nasolabial furrows possible scenarios: (a) nasolabial furrows absence and the two monitored regions around landmarks 37 and 42; (b) nasolabial furrows curve paramterized by $a = 0$ and associated to AU6; (c) nasolabial furrows curve characterized by $a < 0$ and $x = -b/2a$,associated to AU6; (d) nasolabial furrows curve characterized by $a < 0$ and $x \notin$ the visible curve, associated to AU9 and AU10.

5.3 Expressions Descriptive Units (EDU)

In the visual perception community there is a general agreement on the fact that face recognition is the result of two main sources of information: featural coming from individual facial features (mouth, nose, etc.) and configural related to the facial layout and configuration of the previous features (Farah et al., 1998, Cabeza and Kato, 2000). The measures extrapolated by the FACS give information about isolated components in a face, providing a featural contribution to face representation. According to the hypothesis of configural encoding, the spatial relationships between facial components provide additional sources of information in the analysis of facial expressions. In order to exploit the combination of these two useful sources we have decided to add a group of measures encoding the interactions among the featural descriptors showed in Figure 5(c). For that purpose we define to use the set of measures, called Expression Descriptive Unit (EDU), reported in Table 3 and introduced by the authors in (Antonini et al., 2006). The first 5 EDUs represent, respectively, the eccentricity of eyes, left and right eyebrows, mouth and nose. The EDUs from 7 to 9 represent the eyes interactions with mouth and nose, while the 10th EDU is the nose-mouth relational unit. The last 4 EDUs relate the eyebrows to mouth and nose. The EDUs can be intuitively interpreted. For example, in a face displaying a surprise expression, the eyes and the mouth are usually opened and this can be captured by EDU7 ($eye_{\text{height}}/\text{mouth}_{\text{height}}$).

EDU Measures	Measures definition	Explanatory Variables
EDU1	$\frac{lew+rew}{leh+reh}$	EV^E_{28}
EDU2	$\frac{lbw}{lbh}$	EV^E_{29}
EDU3	$\frac{rbw}{rbh}$	EV^E_{30}
EDU4	$\frac{mw}{mh}$	EV^E_{31}
EDU5	$\frac{nh}{nw}$	EV^E_{32}
EDU6	$\frac{lew}{mw}$	EV^E_{33}
EDU7	$\frac{leh}{mh}$	EV^E_{34}
EDU8	$\frac{leh+reh}{lbh+rbh}$	EV^E_{35}
EDU9	$\frac{lew}{nw}$	EV^E_{36}
EDU10	$\frac{nw}{mw}$	EV^E_{37}
EDU11	$\frac{EDU2}{EDU4}$	EV^E_{38}
EDU12	$\frac{EDU3}{EDU4}$	EV^E_{39}
EDU13	$\frac{EDU2}{EDU10}$	EV^E_{40}
EDU14	$\frac{EDU3}{EDU14}$	EV^E_{41}

Table 3: Expressions Descriptive Units

Figure 8: Examples of synthesized faces obtained varying the first c parameter from the mean face ($\pm 3\mathrm{std}$).

5.4 Appearance vector(c)

FACS and EDU provide measures of local facial features or areas that are prone to change with facial expressions, but they do not provide a description of a face as a global entity. This information can be obtained considering the appearance vector c matching the face in the processed image. Figure 8 shows the effect of varying the first appearance model parameter, showing changes in identity and expression.

6 Models specification

In this paragraph we focus on the deterministic part V_i of the random utility function (see Eq. (1)). Any alternative i can be described in terms of a combination of a certain number of attributes EV_i reflecting reasonable hypotheses about the effects of these variables on the corresponding utility. We propose three models of increasing complexity.

$$\begin{aligned} V_j = ASC_j \quad & + \textstyle\sum_{k=1}^{K_F} I_{kj}^F \beta_{kj}^F EV_k^F && \textit{FACS Model} \\ & + \textstyle\sum_{h=1}^{K_E} I_{hj}^E \beta_{hj}^E EV_h^E && \textit{FACS + EDU Model} \\ & + \textstyle\sum_{l=1}^{K_C} I_{kl}^C \beta_{lj}^C EV_l^C && \textit{FACS + EDU +C Model} \end{aligned} \tag{8}$$

where $j \in \{$"happiness", "surprise", "fear", "disgust", "sadness", "anger", "neutral", "other", "I don't know"$\}$, $\{F, E, C\}$ refer respectively to the FACS, EDUs and the appearance parameters c, $EV_{\{k,h,l\}}^{\{F,E,C\}}$ refers to $\{k, h, l\}$-th explanatory variable of one of the used sets, $K_{\{F,E,C\}}$ is the total number of the explanatory variables for each set, $I_{kj}^{\{F,E,C\}}$ is an indicator function equal to 1 if the k-th explanatory variable is included in the utility for the alternative j and 0 otherwise, $\beta_{kj}^{\{F,E,C\}}$ is the weight for the k-th EV in alternative j and ASC_j is an alternative specific constant. The ASC_j coefficients represent the average value of the unobserved part of the corresponding utility and they are added in each utility. For the model to be identified, one of the constant must be normalized to zero. In our case the neutral alternative is considered as the reference alternative and its ASC is set to zero. In addition neutral is a "by default expression", it corresponds to a fully relaxation of the facial muscles. Indeed features of a neutral face are supposed to be at their basic level. Consequently in the developed DCMs, the deterministic utility associated to the neutral expression is fixed to zero. Concerning the "Don't know" alternative, it has been introduced in the survey in order to avoid collecting noise. In the models, its corresponding utility contains only an ASC because no clear causal effect can be identified. This is not the case for the "Other" alternative, which represents a set of expressions. Principal features are introduced in its deterministic utility, according to principal AUs. Different models utilities specifications are presented in table 6 in Appendix A. The first version of the systematic utility functions (FACS Model, in Eq. (8)), for the proposed MNL model, includes the explanatory variables associated with the local measures defined in the AU. In the second step the local interactions between facial features provided by the EDUs are also included, $FACS + EDU$ *Model* in Eq. (8). In the last model the c appearance parameters, encoding global measures about the face, are finally added to the two previous sets of measures, *Model FEC* in eq.8. The 5 first c parameters, that capture the 75% of the total variance in the AAM training set, are introduced in the utility functions using alternative specific parameters.

7 Model estimation

The models introduced in the previous section have been estimated using the free Biogeme package (Bierlaire, 2003) using maximum likelihood estimation. In Table 4 we report the final coefficients estimates for some β for the three models. In the first half of the table, each row relates each particular β for a specific model to its estimated coefficient and its associated t-statistic values. The second half of the table shows summary statistics for the entire estimation run for each of the three models.
The sign of the parameters are consistent with the common reading of facial expressions in terms of facial component modifications. In Table 4, we report a subset of β_{ki} estimates. A parameter is considered significant if

(a) Expressions decreasing C5 (b) Neutral

(c) Expressions increasing C5

Figure 9: Example of the effect of variation of the c5 value. Increasing this parameter (leaving unchanged the others) we move towards a happiness-like expression, whereas an anger-like face corresponds to values of c5 smaller than the reference one.

the norm of the t-test against 0 is bigger than 1.96, representing the 95% of significance.

β^{F}_{17H} represents the coefficient of the mouth width measure in the happiness expression. It is a FACS parameter and it is included in all the specifications. Its positive value shows a positive impact on the respective utility. This means that an increase of the mouth width with respect to the neutral expression (the reference one in our model) corresponds to higher utilities for the happiness alternative. The β^{F}_{17H} estimate is inline with the FACS expectations for the happiness expression. The first row in Table 1 describes the FACS happiness encoding in terms of the primary action units 6 and 12. During an AU12 a stretching of the mouth's lip corners is expected. This corresponds indeed to an increase of the measure $\overline{CC'}$ associated to the estimated parameter β^{F}_{17H} and representing the mouth width.

β^{FE}_{31SU} is the parameter related to EDU4 (Table 3) describing the mouth eccentricity in the surprise alternative. Its positive sign explains the expected behaviour of the mouth in subjects performing a surprise expression, where the mouth movement leads to a lower mouth's height and a higher mouth's width, with respect to the reference alternative.

The third parameter β^{FEC}_{46A} is the coefficient related to the fifth appearance parameters c for the anger utility. The bigger this coefficient is the more negative is the impact on the anger utility. We can visually interpret this result by looking at Figure 9. Considering the neutral *c5* value as the reference value, we can notice how increasing this parameter (leaving unchanged the others) we move towards a happiness-like expression, whereas an anger-like face corresponds to values of *c5* smaller than the reference one.

The statistics concerning the goodness of fit for the three different models are reported in the second half of Table 4. It can be observed that for the second model the fitting is better than for the first one (higher log-likelihood and $\overline{\rho}^2$) and the same for the third model with respect to the second one. The proposed models have been built in a nested way. This means that the first model is a restricted version of the second one and the latest a restriction of the third one. In this case, a *likelihood ratio test* (Ben-Akiva and Lerman, 1985) can be used to verify if the additional variables of the unrestricted model add a significant explanatory power to the model and compensate for the degrees of freedom used by the fuller specification. The null hypothesis for this test states that the restricted and unrestricted models are equivalent. The statistic to compute the test is

$$-2(\mathcal{L}(\widehat{\beta_R}) - \mathcal{L}(\widehat{\beta_U})) \sim \chi^2_{K_U - K_R} \tag{9}$$

where K_i is the number of parameters of the model i and χ^2_j is a χ^2 distribution with j degrees of freedom. Usually, a significance level of 95% is taken, and then the null hypothesis is rejected if the test value is above the threshold provided by the χ^2 distribution corresponding to the j degrees of freedom. The results for this test are reported in Table 5. The performed tests refer to the two possible *(restriced,unrestricted)* models couples. The first test shows that the inclusion of new parameters makes the unrestricted FE model significantly different from its restricted counterpart, the F model. This result justifies the second test comparing the most complex model (FEC) with its restricted version (FE), showing that the model considering the whole set of 3 different explanatory variables can be considered and retained as the final model that best fit our data.

F MODEL			FE MODEL			FEC MODEL		
β^{F}_{ki}	**estimate**	***t* test 0**	β^{FE}_{ki}	**estimate**	***t* test 0**	β^{FEC}_{ki}	**estimate**	***t* test 0**
β^{F}_{17H}	+ 103	+ 56.81	β^{FE}_{17H}	+ 34	+ 4.98	β^{FEC}_{17H}	+ 105	+ 37.67
			β^{FE}_{31SU}	+ 8.12	+ 48.3	β^{FEC}_{31SU}	+ 6.89	+ 39.59
						β^{FEC}_{46A}	- 9.67	- 11.13
β^{F}_{17H}=mouth width Happiness, β^{FE}_{31SU}=EDU4 Surprise, β^{FEC}_{46A}=C5 Anger								
Sample size = 38110			Sample size = 38110			Sample size = 38110		
Nb. of estimated parameters = 93			Nb. of estimated parameters = 120			Nb. of estimated parameters = 139		
Null log-likelihood = - 83736.229			Null log-likelihood = - 83736.229			Null log-likelihood = - 83736.229		
Final log-likelihood = - 57072.872			Final log-likelihood = - 55027.381			Final log-likelihood = - 53474.271		
Likelihood ratio test = 53326.712			Likelihood ratio test = 57417.695			Likelihood ratio test = 60523.915		
$\bar{\rho}^2$ = 0.317			$\bar{\rho}^2$ = 0.341			$\bar{\rho}^2$ = 0.360		

Table 4: Estimation results for the FACS, FACS+EDU, FACS+EDU+C models

Performed test	Degrees of freedom	Test value	χ^2 Threshold
F vs FE	27	4090.98	40.11
FE vs FEC	19	3106.22	30.14

Table 5: Summary of the different performed likelihood ratio tests

8 Extension to dynamic facial expression recognition

The Discrete Choice framework used for static facial expression recognition is extended in order to consider face video sequences instead of images. An internet survey similar to the one described in section 3.2 is currently conducted for collecting expressions labels on face video sequences. It is available at *http://transp-or2.epfl.ch/videosurvey/*. Two video databases are used, the Cohn-Kanade database (Kanade et al., 2000) (also used in the static case), and the Facial Expressions and Emotions Database (Wallhoff, 2004). The dynamic model is inspired by car line changing models (Choudhury, 2007) and is a direct application of discrete choice model with latent segmentation (Walker, 2001). We hypothesise that the respondent expression perception evolves when watching the video. In addition we consider that the influence of the video frames on the respondent perception is varying depending on their dynamic. Considering perception evolving at each frame is not realistic. Indeed frames transition is too fast as frame rate is 25 per second, consequently a perception evolution time step is defined equal to one second. The sequence is therefore sampled selecting the first frame of each group of 25 frames. Features for each frame group are then the features of its first frame. By extension in the following we call a group of frames, a frame.

The dynamic facial expression recognition model consists of a combination of two DCMs. A perception state, corresponding to the respondent facial expression perception, is associated to each time step. A first DCM is used to quantify this perception, whose choice set is composed of the nine expressions used in the static case. This is similar to the static model. The second DCM quantifies the frame influences on the respondent observed facial expression choice. The choice set in this case is composed of the frames of the labelled video, which makes that the choice set varies from one video to another. Note that both models are based on latent concepts, indeed the respondent instantaneous perception and the frames influences are not observed. Only the video expression choice is observed.

The probability for respondent n to choose the expression i when watching the frame t of the video sequence o is written $P_n(i|t,o)$ (first DCM). Then, the probability for the respondent n to make her expression choice when watching the frame t of the video sequence o is $P_n(t|o)$ (second DCM). The two DCMs are linked by the probability for the respondent n to label the video o with the expression i, called $P_n(i|o)$. This relation can be expressed as

$$P_n(i|o) = \sum_{t=1}^{T_o} P_n(i|t,o) P_n(t|o), \tag{10}$$

T_o being the video duration in seconds.

As shown for the static model, $P_n(i|t,o)$ is quite universal, in the sense that for the moment no clear socio-economic characteristic seems to interact with the expression perception. We expect that this is not the case for $P_n t|o$ which should strongly depend on the respondent n. Indeed the frame dynamic perception depends on the current respondent attention. This leads to take into account the panel data effect. ξ_n is defined as a random term specific to the respondent n. So equation 10 can be transformed as

$$P_n(i|o,\xi_n) = \sum_{t=1}^{T_o} P_n(i|t,o) P_n(t|o,\xi_n). \tag{11}$$

In order to obtain a closed form of $P_n(i|o,\xi_n)$, we need to integrate on ξ_n. By default ξ_n is supposed to be normally distributed $N(0,\sigma)$. $f(\xi)$ is the probability density distribution of ξ_n, and O_n is the number of observations associated to the respondent n. By integration we obtain $P_n(i|o)$

$$\prod_{o=1}^{O_n} P_n(i|o) = \int \prod_{o=1}^{O_n} \sum_{t=1}^{T_o} P_n(i|t,o) P_n(t|o,\xi_n) f(\xi) d\xi. \tag{12}$$

Theoretically $P_n(i|t,o)$ can be of any DCM type, such as multivariate extreme value (MEV), or mixture of logit models. But as mentioned before, the model is designed exactly for the same purpose than the static model, so in a first time a simple logit model will be used, and the utility specification will be near from the one proposed in the static model version. In a second step, utilities will take into account the perception memory effect. Concerning $P_n(t|o,\xi_n)$, it is a mixture of logit models, due to the panel data effect term. We prefer to use a quite simple model form, such as mixture of logit models, and not mixtures of MEV models, because the correlation between frames is difficult to define. Indeed the frames number vary from one video to another. The utility specification has to contain attributes which reflect the frame dynamics, such as derivatives of the attributes used in the first DCM. The idea to use a simple correlation structure is also motivated by the fact that both models are estimated jointly by likelihood maximization, as a classical DCM. Indeed the combination of such models can imply high non linearities in the likelihood function, and the optimization algorithm has to deal with such difficulties. If we call β the parameters vector we want to estimate, the likelihood $l(\beta)$ has the following form

$$l(\beta) = \prod_{n=1}^{N} \left(\prod_{o=1}^{O_n} P_n(i|t,o,\beta) \right). \tag{13}$$

By mixing equation 12 and equation 13 we obtain

$$l(\beta) = \prod_{n=1}^{N} \left(\int \prod_{o=1}^{O_n} \sum_{t=1}^{T_o} P_n(i/t,o,\beta) P_n(t/o,\xi_n,\beta) f(\xi) d\xi \right). \tag{14}$$

But for numerical reasons, the logarithm of the likelihood function,

$$L(\beta) = \sum_{n=1}^{N} \log\left(\int \prod_{o=1}^{O_n} \sum_{t=1}^{T_o} P_n(i|t,o,\beta) P_n(t|o,\xi_n,\beta) f(\xi) d\xi \right), \tag{15}$$

is used instead of $l(\beta)$ during the estimation process. An extension of the biogeme software (Bierlaire, 2003) will be implemented to estimate such models, the optimization toolbox remaining the same.

We conclude this section by underlying the fact that the model specification will depend on the number of observations provided by the internet video survey. Indeed nowadays the data base contains 500 observations. This little number constrains the number of alternative specific parameters in the perception model to be reduced, compared to the static model version.

9 Conclusion and discussion

We have proposed a new method for facial expressions modelling, based on discrete choice analysis. The data of the facial evaluation survey suggested that a subjective component biases the labelling process, requiring a detailed statistical analysis on the collected data. DCM paradigm well matches the human observer labelling procedure, allowing to capture and model the subjective perception of the choice makers. In the static case, we showed how to improve the descriptiveness of the model by sequentially introducing complementary set of features. The estimation of the three proposed models has shown the correctness of the chosen sets of features, revealing the best fitting behaviour of the third and most complex model.

This work represents one of the first attempts to apply discrete choice analysis for modelling facial expressions. Several means of improvement are possible. First, a deeper understanding of the choice process can be achieved by exploring the personal characteristics of the decision-maker. The heterogeneity in the respondent population of the survey will allow the investigation and the interpretation of these human factors. For that purpose, the socio-economic features can be analysed and introduced in the utility functions as categorical variables. This analysis would overcome another shortcoming of previous approaches where humans are usually modelled as *invariants* and not as *individuals*. While modelling invariants is fundamental for most machine learning or patterns recognition problems, in perception it is also important to ask how people are different. A further

investigation of the parameters involved in the decision-maker's choice process can be obtained by applying a segmentation of the population. This means that, instead of introducing a parameter for each socio-economic attribute, the population is divided with respect to that feature. For example, the behaviour of men and women can be explored by analysing the two groups separately.
Secondly, other families of discrete choice models can be used. As described in Section 4, the utility of each alternative is a random variable containing a systematic random part. Different assumptions about the random term give rise to different models. The MNL models assume no correlations between alternatives. This hypothesis can be relaxed, by considering Nested (Daly and Zachary, 1978) and Cross-Nested (Bierlaire, 2006) models.
Finally the static discrete choice framework has been extended to the dynamic case. A model composed of 2 discrete choice sub-models is proposed, one of them being similar to the model used in the static version, the other one measuring the influence of each video frame. The dynamic model is an adaptation of a DCM with latent segmentation proposed by Walker, 2001.

References

Abboud, D. and Davoine, F. (2004). Appearance factorization based facial expression recognition and synthesis., *ICPR (4)*, pp. 163–166.

Antonini, G., Sorci, M., Bierlaire, M. and Thiran, J. (2006). Discrete choice models for static facial expression recognition, *in* J. Blanc-Talon, W. Philips, D. Popescu and P. Scheunders (eds), *8th International Conference on Advanced Concepts for Intelligent Vision Systems*, Vol. 4179 of *Lecture Notes in Computer Science*, Springer Berlin / Heidelberg, Berlin, pp. 710–721. ISBN: 978-3-540-44630-9.

Bartlett, M. S., Littlewort, G., Frank, M., Lainscsek1, C., Fasel, I. and Movellan, J. (2006). Fully automatic facial action recognition in spontaneous behavior, *FGR '06: Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition*, IEEE Computer Society, Washington, DC, USA, pp. 223–230.

Bassili, J. N. (1978). Facial motion in the perception of faces and of emotional expression., *J Exp Psychol Hum Percept Perform* **4**(3): 373–379.

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

Bicego, M., Salah, A. A., Grosso, E., Tistarelli, M. and Akarun, L. (2007). Generalization in holistic versus analytic processing of faces, *ICIAP '07: Proceedings of the 14th International Conference on Image Analysis and Processing*, IEEE Computer Society, Washington, DC, USA, pp. 235–240.

Bierlaire, M. (2003). BIOGEME: a free package for the estimation of discrete choice models, *Proceedings of the 3rd Swiss Transportation Research Conference*, Ascona, Switzerland.

Bierlaire, M. (2006). A theoretical analysis of the cross-nested logit model, *Annals of operations research* **144**(1): 287–300.

Cabeza, R. and Kato, T. (2000). Features are also important: Contributions of featural and configural processing to face recognition, *Psychological Science* **11**: 429–433.

Carey, S. (1992). Becoming a face expert, *Philosophical Transactions of the Royal Society of London, B* **335**: 95–103.

Chang, Y., Hu, C. and Turk, M. (2004). Probabilistic expression analysis on manifolds, *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition CVPR 2004*, Vol. 2, pp. II–520–II–527.

Choudhury, C. F. (2007). *Model Driving Decisions with Latent Plans*, PhD thesis, Massachusetts institue of technology.

Cohen, I., Sebe, N., Chen, L., Garg, A. and Huang, T. S. (2003). Facial expression recognition from video sequences: Temporal and static modeling, *Computer Vision and Image Understanding* (10): 160–187.

Cohn, J. F. (2006). Foundations of human computing: facial expression and emotion, *ICMI '06: Proceedings of the 8th international conference on Multimodal interfaces*, ACM, New York, NY, USA, pp. 233–238.

Cootes, T. F., Edwards, G. J. and Taylor, C. J. (2001). Active appearance models, *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **23**: 681–685.

Cootes, T. and Taylor, C. (2004). Statistical models of appearance for computer vision.

Daly, A. J. and Zachary, S. (1978). Improved multiple choice, *in* D. A. Hensher and M. Q. Dalvi (eds), *Determinants of travel demand*, Saxon House, Sussex.

Diamond, R. and Carey, S. (1986). Why faces are and are not special: An effect of expertise, *Journal of Experimental Psychology: General* **115**(2): 107–117.

Ekman, P. (1992). An argument for basic emotions, *Cognition & Emotion* **6**(3): 169–200.

Ekman, P., Friesen, W. and Hager, J. (2002). *Facial action coding system*, Research Nexus, Network Research Information, Salt Lake City,UT.

Ekman, P. and Friesen, W. V. (1978). *Facial Action Coding System Investigator's Guide*, Consulting Psycologist Press, Palo Alto, CA.

Farah, M. J., Wilson, K. D., Drain, M. and Tanaka, J. N. (1998). What is "special" about face perception?, *Psychol Rev* **105**(3): 482–498.

Fasel, B. and Luettin, J. (2003). Automatic facial expression analysis: A survey, *Pattern Recognition* **36**(1): 259–275.

Freund and Schapire (1997). A decision-theoretic generalization of on-line learning and an application to boosting, *Journal of Computer and System Sciences* **55**.

Friesen, W. and Ekman, P. (1983). Emfacs-7: Emotional facial action coding system. University of California at San Francisco.

Friesen, W. V. and Ekman, P. (1984). Emotional facial action coding system.

Hu, C., Chang, Y., Feris, R. and Turk, M. (2004). Manifold based analysis of facial expression, *CVPRW '04: Proceedings of the 2004 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'04) Volume 5*, IEEE Computer Society, Washington, DC, USA, p. 81.

I.L., D. and K.V., M. (1998). *Statistical Shape Analysis*, John Wiley & Sons, New York.

Kanade, T., Cohn, J. and Tian, Y. L. (2000). Comprehensive database for facial expression analysis, *Proceedings of the 4th IEEE International Conference on Automatic Face and Gesture Recognition (FG'00)*, pp. 46 – 53.

Keltner, D. Ekman, P. (2000). Facial expression of emotion, *Handbooks of emotions*, M.Lewis & J.M.Havilland, pp. 236–249.

li Tian, Y., Kanade, T. and Cohn, J. F. (2001). Recognizing action units for facial expression analysis, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **23**: 97–115.

li Tian, Y., Kanade, T. and Cohn, J. F. (2002). Evaluation of gabor-wavelet-based facial action unit recognition in image sequences of increasing complexity, *Proc. Fifth IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 229–234.

Lyons, M., Akamatsu, S., Kamachi, M. and Gyoba, J. (1998). Coding facial expressions with gabor wavelets, *FG '98: Proceedings of the 3rd. International Conference on Face & Gesture Recognition*, IEEE Computer Society, Washington, DC, USA, p. 200.

M., P. and Bartlett, M. (2007). Machine analysis of facial expressions, *in* K. D. . M. Grgic (ed.), *Face Recognition*, Vienna, Austria: I-Tech Education and Publishing, chapter 20, pp. 377–416.

Manski, C. (1977). The structure of random utility models, *Theory and Decision* **8**: 229–254.

Manski, C. F. and McFadden, D. (1981). *Econometric models of probabilistic choice*, in C.F. Manski and D. McFadden, editors, Structural Analysis of Discrete Data with Econometric Applications, MIT Press, Cambridge, 198-272.

Matthews, I. and Baker, S. (2004). Active appearance models revisited, *International Journal of Computer Vision* **60**(1): 135–164.

Medioni, G. and Kang, S.-B. (2004). *Emerging Topics in Computer Vision*, Prentice Hall PTR, Upper Saddle River, NJ, USA.

Meer, P.; Georgescu, B. (Dec 2001). Edge detection with embedded confidence, *Transactions on Pattern Analysis and Machine Intelligence* **23**(12): 1351–1365.

Meulders, M., Boeck, P. D., Mechelen, I. V. and Gelman, A. (2005). Probabilistic feature analysis of facial perception of emotions, *Journal Of The Royal Statistical Society Series C* **54**(4): 781–793.

Moses, Y., Ullman, S. and Edelman, S. (1996). Generalization to novel images in upright and inverted faces, *Perception* **25**: 443–462.

Padgett, C. and Cottrell, G. W. (1998). A simple neural network models categorical perception of facial expressions, *In Proceedings of the Twentieth Annual Cognitive Science Conference*, Erlbaum, pp. 806–807.

Pantic, M. and Patras, I. (2006). Dynamics of facial expression: recognition of facial actions and their temporal segments from face profile image sequences, *IEEE Transactions on Systems, Man, and Cybernetics, Part B* **36**(2): 433–449.

Pantic, M., Valstar, M. F., Rademaker, R. and Maat, L. (2005). Web-based database for facial expression analysis, *IEEE International Conference on Multimedia and Expo (ICME)*, IEEE, pp. 317–321.

Pentland, A., Moghaddam, B. and Starner, T. (1994). View-based and modular eigenspaces for face recognition, *Proceedings of the Fourth International Conference on Computer Vision*, pp. 84–91.

Rowley, H., Baluja, S. and Kanade, T. (1998). Neural network-based face detection, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **20**(1): 23–38.

Schneiderman, H. and Kanade, T. (2000). A statistical model for 3d object detection applied to faces and cars., *in* IEEE (ed.), *IEEE Conference on Computer Vision and Pattern Recognition*.

Schwaninger, A. (2003). *Perception and representation of faces*, PhD thesis, Universität Zürich.

Schwaninger, A., Lobmaier, J. S. and Collishaw, S. M. (2002). Role of featural and configural information in familiar and unfamiliar face recognition, *BMCV '02: Proceedings of the Second International Workshop on Biologically Motivated Computer Vision*, Springer-Verlag, London, UK, pp. 643–650.

Sorci, M., Antonini, G., Thiran, J.-P. and Bierlaire, M. (2007). Facial Expressions Evaluation Survey. ITS.

Stegmann, M. B. (2000). *Active appearance models: Theory, extensions and cases*, Master's thesis, Informatics and Mathematical Modelling, Technical University of Denmark, DTU, Richard Petersens Plads, Building 321, DK-2800 Kgs. Lyngby.

Sung, K. K. and Poggio, T. (1998). Example-based learning for view-based human face detection, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **20**(1): 39–51.

Tian, Y.-L., Kanade, T. and Cohn, J. (2003). Facial expression analysis, *in* S. L. . A. Jain (ed.), *Handbook of face recognition*, Springer.

Valstar, M. F. and Pantic, M. (2007). Combined support vector machines and hidden markov models for modeling facial action temporal dynamics, *ICCV-HCI*, pp. 118–127.

Viola, P. and Jones, M. (2004). Robust real-time face detection, *International Journal of Computer Vision* **57**(2): 137–154.

Walker, J. L. (2001). *Extended Discrete Choice Models: Integrated Framework, Flexible Error Structures, and Latent Variables*, PhD thesis, Massachusetts Institute of Technology.

Wallhoff, F. (2004). Fgnet-facial expression and emotion database, *Technical report*, Technische Universitt Mnchen.
URL: *http://www.mmk.ei.tum.de/ waf/fgnet/feedtum.html*

Wallraven, C., Schwaninger, A. and Bülthoff, H. H. (2005). Learning from humans: Computational modeling of face recognition, *Network* **16**(4): 401–418.

Yang, M.-H., Kriegman, D. J. and Ahuja, N. (2002). Detecting faces in images: a survey, *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **24**(1): 34–58.

Ye, J., Zhan, Y. and Song, S. (2004). Facial expression features extraction based on gabor wavelet transformation, *IEEE International Conference on Systems, Man and Cybernetics*, pp. 10–13.

Zhang, Y. and Ji, Q. (2005). Active and dynamic information fusion for facial expression understanding from image sequences, *Transactions on Pattern Analysis and Machine Intelligence* **27**(5): 699–714.

Zhang, Z., Lyons, M., Schuster, M. and Akamatsu, S. (1998). Comparison between geometry-based and gabor-wavelets-based facial expression recognition using multi-layer perceptron, *FG '98: Proceedings of the 3rd. International Conference on Face & Gesture Recognition*, IEEE Computer Society, Washington, DC, USA, p. 454.

Appendices

A Specification table with estimated parameters

The values of estimated parameters are presented in the next table. In the first and second columns the parameter name and its associated feature are mentioned. From column three to eleven, the associated utility for each parameter is indicated. Finally, in columns twelve to fourteen, estimated values and t-tests against zero are shown for the three models. Note that if the parameter is not present in one of the models, the corresponding cell is empty.

	BRIEF DESCRIPTION	UTILITIES									F Model		FE Model		FEC Model	
		H	SU	F	D	SA	A	N	O	DK	estimate	t test 0	estimate	t test 0	estimate	t test 0
β_1	Constant						✓				−2.22	−6.63	−1.51	−2.86	−5.91	−11.53
β_2	Constant				✓						−1.71	−6.73	0.26	0.11	2.65	1.16
β_3	Constant									✓	−2.29	−69.24	−2.29	−69.25	−2.29	−69.24
β_4	Constant			✓							−3.83	−3.53	−1.01	−0.32	−5.65	−3.85
β_5	Constant	✓									1.15	3.52	25.00	10.87	2.40	2.56
β_6	Constant								✓		−1.38	−4.54	−6.05	−3.04	−3.34	−1.67
β_7	Constant					✓					−2.69	−5.63	−14.60	−5.21	−9.61	−3.39
β_8	Constant		✓								−4.05	−21.01	1.56	2.95	−1.92	−3.83
β_9	C1				✓										5.66	7.81
β_{10}	C1			✓											−9.25	−7.83
β_{11}	C1	✓													13.60	15.84
β_{12}	C1								✓						3.07	4.47
β_{13}	C1					✓									10.90	13.61
β_{14}	C1		✓												2.75	3.52
β_{15}	C2						✓								8.87	10.13
β_{16}	C2				✓										18.60	22.87
β_{17}	C2			✓											6.56	5.15
β_{18}	C2	✓													−3.91	−3.88
β_{19}	C2								✓						12.80	17.41
β_{20}	C2					✓									10.10	11.05
β_{21}	C2		✓												−4.04	−4.06
β_{22}	C3						✓								3.05	3.35
β_{23}	C3			✓											18.00	10.40
β_{24}	C3								✓						−5.74	−7.76
β_{25}	C3					✓									−11.80	−12.64
β_{26}	C3		✓												7.29	7.30
β_{27}	C4						✓								9.24	10.36
β_{28}	C4			✓											14.50	12.46
β_{29}	C4	✓													−11.70	−11.77
β_{30}	C4					✓									7.79	9.56
β_{31}	C4		✓												13.70	14.65
β_{32}	C5						✓								−9.67	−10.66
β_{33}	C5			✓											−8.05	−6.82
β_{34}	C5	✓													1.96	2.06
β_{35}	C5								✓						−2.04	−2.59
β_{36}	C5					✓									−7.71	−8.18

	BRIEF DESCRIPTION	UTILITIES									F Model		FE Model		FEC Model	
		H	SU	F	D	SA	A	N	O	DK	estimate	t test 0	estimate	t test 0	estimate	t test 0
β_{37}	C5		✓												−12.90	−14.36
β_{38}	EDU10						✓						9.62	19.91	12.30	23.43
β_{39}	EDU10				✓								13.20	3.75	12.50	3.70
β_{40}	EDU10			✓									−8.14	−6.38	−6.02	−6.76
β_{41}	EDU10								✓				16.00	5.27	12.10	4.03
β_{42}	EDU10					✓							15.40	3.96	11.10	2.95
β_{43}	EDU10		✓										−3.17	−7.17	−2.02	−4.34
β_{44}	EDU5			✓	✓								−1.78	−11.68	−3.18	−27.36
β_{45}	EDU5	✓											2.45	15.44	2.77	15.35
β_{46}	EDU5					✓							−1.25	−8.33	−1.15	−7.39
β_{47}	EDU6				✓								−17.70	−4.15	−19.40	−4.74
β_{48}	EDU6	✓											−16.70	−6.75		
β_{49}	EDU6								✓				−25.70	−7.16	−22.10	−6.21
β_{50}	EDU6					✓							−24.30	−5.49	−21.30	−5.08
β_{51}	EDU7			✓			✓						2.31	14.29	2.21	13.30
β_{52}	EDU7				✓								1.28	5.58	2.44	11.92
β_{53}	EDU7	✓											2.46	5.76	3.13	8.21
β_{54}	EDU7								✓				2.06	10.52	2.68	14.13
β_{55}	EDU7					✓							2.03	10.60	2.05	10.26
β_{56}	EDU8			✓			✓						−2.33	−5.88		
β_{57}	EDU8				✓								−4.29	−12.49	−5.59	−16.68
β_{58}	EDU8	✓											−6.85	−14.29	−6.42	−13.74
β_{59}	EDU8								✓				0.75	2.25	1.13	3.37
β_{60}	EDU8					✓							8.39	12.02	6.15	8.89
β_{61}	EDU8		✓										−5.80	−16.54	−3.94	−11.02
β_{62}	EDU9				✓								12.20	4.29	12.00	4.36
β_{63}	EDU9			✓									−2.97	−2.57	−4.02	−5.71
β_{64}	EDU9	✓											−6.26	−10.81	−3.12	−4.94
β_{65}	EDU9								✓				12.30	5.18	8.08	3.40
β_{66}	EDU9					✓							14.80	5.24	11.50	4.16
β_{67}	RAP brow		✓				✓						−4.78	−7.94	−1.11	−2.07
β_{68}	RAP brow				✓								−10.60	−18.21	−12.20	−21.04
β_{69}	RAP brow			✓									−12.40	−12.63	−5.76	−6.39
β_{70}	RAP brow					✓							12.50	10.75	7.54	6.78
β_{71}	RAP eye						✓						−3.59	−4.87	−7.17	−11.14
β_{72}	RAP eye			✓									7.09	3.24		

	BRIEF DESCRIPTION	UTILITIES									F Model		FE Model		FEC Model	
		H	SU	F	D	SA	A	N	O	DK	estimate	t test 0	estimate	t test 0	estimate	t test 0
β_{73}	RAP eye	✓											−23.40	−10.03	−4.61	−5.54
β_{74}	RAP eye								✓				−8.79	−16.77	−10.30	−19.03
β_{75}	RAP eye					✓							−14.30	−14.53	−11.20	−11.07
β_{76}	RAP eye		✓										2.00	3.45		
β_{77}	RAP mouth						✓						−14	−15.12	−17.40	−17.85
β_{78}	RAP mouth				✓								−3.13	−2.48		
β_{79}	RAP mouth			✓									9.23	33.58	7.75	28.30
β_{80}	RAP mouth	✓											6.28	5.09	8.38	8.25
β_{81}	RAP mouth								✓				3.88	9.99	4.29	11.50
β_{82}	RAP mouth					✓							−7.30	−4.56	−9.98	−5.63
β_{83}	RAP mouth		✓										8.12	39.06	6.89	33.38
β_{84}	brow dist						✓				−9.25	−4.71	−21.70	−9.67	−24.90	−11.46
β_{85}	brow dist			✓							−32.90	−8.91	−14.40	−4.54	−9.28	−3.02
β_{86}	brow dist					✓					−23.10	−11.56	−47.60	−18.46	−43.50	−17.07
β_{87}	broweye l2						✓				−34.40	−7.33	−25.10	−4.60		
β_{88}	broweye l2								✓		24.50	5.86	40.80	10.03	50.00	11.15
β_{89}	broweye l2					✓					−4.41	−0.79	−15.30	−3.96	22.00	5.03
β_{90}	broweye l2		✓								6.48	1.59	33.60	11.30	48.80	14.48
β_{91}	broweye l3						✓				−27.50	−7.72	−28.10	−6.37	−32.80	−8.95
β_{92}	broweye l3		✓								9.99	2.92				
β_{93}	broweye r2						✓				−71.00	−16.26	−75.60	−16.21	−74.20	−19.54
β_{94}	broweye r2				✓						−55.80	−21.12	−50.10	−15.38	−31.20	−10.15
β_{95}	broweye r2			✓							−19.10	−2.02				
β_{96}	broweye r2					✓					−59.20	−9.18	−91.30	−10.28	−80.80	−8.71
β_{97}	broweye r2		✓								−4.40	−0.72				
β_{98}	browwr						✓				4.26	2.55				
β_{99}	browwr				✓						11.90	7.04	10.40	6.15	4.64	2.59
β_{100}	browwr								✓		6.31	4.48	4.28	2.97	4.33	2.95
β_{101}	browwr					✓					3.15	1.88				
β_{102}	eye angle below l			✓							−1.46	−6.07				
β_{103}	eye angle below r			✓							0.26	0.88	2.36	6.44	1.96	11.50
β_{104}	eye angle below r					✓					0.61	4.75				
β_{105}	eye angle l						✓				−0.76	−2.36	1.54	3.96	2.47	6.37
β_{106}	eye angle l			✓							5.86	12.69	5.06	10.02	5.13	12.25
β_{107}	eye angle l					✓					4.21	14.89	1.97	5.00	1.09	2.68
β_{108}	eye angle r						✓				3.37	9.89	2.03	4.84	2.72	6.15

	BRIEF DESCRIPTION	UTILITIES									F Model		FE Model		FEC Model	
		H	SU	F	D	SA	A	N	O	DK	estimate	t test 0	estimate	t test 0	estimate	t test 0
β_{109}	eye angle r			✓							0.83	2.11	−3.12	−5.34		
β_{110}	eye angle r					✓					−4.71	−17.21	−1.70	−3.93	−1.76	−4.15
β_{111}	eye brow angle l			✓							7.05	13.98	4.32	9.91	4.42	10.15
β_{112}	eye brow angle l								✓		−2.73	−7.99	−3.93	−11.11	−3.42	−9.72
β_{113}	eye brow angle l					✓					−1.13	−2.63				
β_{114}	eye brow angle r			✓							−1.46	−2.14	−2.10	−6.49	−1.54	−5.86
β_{115}	eye brow angle r								✓		−1.75	−8.49	−0.84	−4.18	−0.95	−4.24
β_{116}	eye brow angle r					✓					5.31	12.84	7.96	12.43	5.81	9.10
β_{117}	eye brow angle r		✓								−1.22	−3.69	−2.75	−12.37	−2.93	−13.69
β_{118}	eye mouth dist l2				✓						−41.10	−14.79	−16.00	−4.13		
β_{119}	eye mouth dist l2								✓		−8.29	−3.51				
β_{120}	eye mouth dist l			✓							33.30	5.07	54.00	8.57	66.30	10.12
β_{121}	eye mouth dist l	✓									−12.30	−3.23	−55.70	−10.09	−59.70	−10.66
β_{122}	eye mouth dist l								✓		−29.60	−7.74				
β_{123}	eye mouth dist l					✓					−30.70	−6.44	20.70	3.84	21.10	3.96
β_{124}	eye mouth dist r2				✓						27.70	11.86	31.60	11.19	26.70	12.72
β_{125}	eye mouth dist r2								✓		7.52	3.02	−4.50	−3.99	−4.40	−3.74
β_{126}	eye mouth dist r			✓							−30.90	−4.81	−42.40	−6.84	−46.90	−7.22
β_{127}	eye mouth dist r	✓									−79.80	−20.78	−63.40	−12.59	−58.60	−11.17
β_{128}	eye mouth dist r								✓		29.70	8.33				
β_{129}	eye mouth dist r					✓					62.20	14.47	28.80	6.39	36.50	8.12
β_{130}	eye nose dist l						✓				5.15	0.84	70.10	9.72	67.30	9.23
β_{131}	eye nose dist l				✓						90.00	15.96	96.50	13.80	49.50	8.26
β_{132}	eye nose dist l			✓							64.10	8.10	42.00	4.84	−19.70	−5.77
β_{133}	eye nose dist l								✓		90.40	16.86	78.20	15.42	54.90	10.47
β_{134}	eye nose dist l					✓					113.00	19.33	105.00	15.34	79.40	11.23
β_{135}	eye nose dist r						✓				50.20	6.72	−31.50	−3.63	−25.00	−2.87
β_{136}	eye nose dist r				✓						−94.90	−14.68	−136.00	−19.01	−96.20	−12.88
β_{137}	eye nose dist r			✓							−74.70	−7.34	−62.00	−6.09		
β_{138}	eye nose dist r								✓		−108.00	−17.09	−77.00	−12.79	−38.90	−6.05
β_{139}	eye nose dist r					✓					−135.00	−20.26	−117.00	−14.77	−95.30	−12.18
β_{140}	fore						✓				0.13	1.62				
β_{141}	fore			✓							0.87	11.21	0.67	9.09	0.74	9.39
β_{142}	fore								✓		0.29	4.82	0.16	2.67	0.20	3.21
β_{143}	fore		✓								0.56	9.29	0.54	9.03	0.47	7.56
β_{144}	leye h			✓							−81.20	−8.11	−86.70	−4.40	−32.00	−3.55

	BRIEF DESCRIPTION	UTILITIES									F Model		FE Model		FEC Model	
		H	SU	F	D	SA	A	N	O	DK	estimate	t test 0	estimate	t test 0	estimate	t test 0
β_{145}	leye h	✓									−27.60	−4.51	204.00	12.31	41.70	4.09
β_{146}	leye h		✓								−61.20	−9.21	−20.90	−2.91	−26.70	−3.63
β_{147}	mouth h						✓				−27.60	−9.90	111.00	18.17	134.00	21.23
β_{148}	mouth h				✓						−5.46	−3.54	43.00	5.40	28.20	8.66
β_{149}	mouth h			✓							42.90	29.23				
β_{150}	mouth h	✓									−4.07	−1.62				
β_{151}	mouth h					✓					−8.45	−3.91	73.20	6.00	72.50	5.38
β_{152}	mouth h		✓								55.10	43.15				
β_{153}	mouth nose dist2						✓				8.17	3.46	5.39	2.18		
β_{154}	mouth nose dist2					✓					−14.20	−7.03	−20.10	−9.69	−5.15	−2.25
β_{155}	mouth nose dist				✓						15.70	4.89	−11.80	−3.38	−19.40	−6.12
β_{156}	mouth nose dist	✓									31.20	14.28	37.90	11.87	59.10	18.56
β_{157}	mouth w						✓				23.30	11.41				
β_{158}	mouth w				✓						31.30	17.91				
β_{159}	mouth w			✓							19.80	9.43	23.10	4.07	18.60	5.09
β_{160}	mouth w	✓									103.00	41.72	34.40	4.19	105.00	37.67
β_{161}	mouth w								✓		19.30	10.54				
β_{162}	mouth w					✓					−3.07	−1.56	−44.90	−7.42	−49.90	−8.38
β_{163}	naslab				✓						0.76	14.52	0.57	11.09	0.68	12.66
β_{164}	naswr				✓						18.80	30.31	16.70	24.11	15.70	22.63
β_{165}	naswr								✓		4.73	6.68	6.35	9.16	5.94	8.22
β_{166}	reye h			✓							−33.20	−2.61				
β_{167}	reye h	✓									44.70	9.35	190.00	11.03	36.00	3.95
β_{168}	reye h		✓								30.30	7.01	38.00	5.60	44.90	9.14

Table 6: Details of models specifications