PARAMETERIZED FACIAL EXPRESSION SYNTHESIS FOR VIDEOCONFERENCING APPLICATIONS

Amaryllis Raouzaiou, Nicolas Tsapatsoulis and Stefanos Kollias

Image, Video and Multimedia Systems Lab - Dept of Electrical and Computer Engineering
National Technical University of Athens
Heroon Polytechniou 9, 157 73 Zographou, GREECE
Tel.: (301) 7722491 – Fax: (301) 7722492
email: araouz@image.ece.ntua.gr

ABSTRACT

In this paper we propose a method of creating intermediate facial expressions based on primary ones [1]. To achieve this goal we utilize both Facial Definition Parameters (FDPs) and Facial Animation Parameters (FAPs). We introduce a way for modeling the primary expressions using FAPs and we describe a rule-based technique for the synthesis of intermediate ones. Furthermore, a relation between FAPs and the *activation* parameter proposed in some classic psychological studies is established. In this way we try to get advantage of the extended work that have been done from psychologists and which covers much more expressions than the archetypal ones the computer society concentrated on. The overall scheme leads to a parameterized approach of synthesizing facial and can be used for the creation of MPEG-4 compatible synthetic video sequences.

1. INTRODUCTION

Research in facial expression analysis and synthesis has mainly concentrated on primary or archetypal emotions. In particular, sadness, anger, joy, fear, disgust and surprise are categories of emotions that attracted most of the interest in human computer interaction environments. Very few studies [2] have appeared in the computer science literature, which explore non-archetypal emotions. This trend may be due to the great influence of the works of Ekman [3], Friesen [4] and Izard [5] who proposed that the archetypal emotions correspond to distinct facial expressions which are supposed to be universally recognizable across cultures. In the contrary psychological researchers have extensively investigated [1][6] a broader variety of emotions. An extensive survey on emotion analysis can be found in [7].

Although the exploitation of the results obtained by the psychologists is far from being straightforward, computer scientists can use some hints to their research. On the other hand, the MPEG-4 indicates an alternative way of modeling facial expressions and the underlying emotions, which is strongly influenced from neurophysiological and psychological studies. For example, FAPs that are utilized in the framework of MPEG-4 for facial animation purposes, are strongly related to the Action Units (AUs) which consist the core of the Facial Action Coding System (FACS) [1].

One of the studies carried out from psychologists and which could be useful to researchers of the area of computer graphics and machine vision is the one of Whissel's [8], who suggested that emotions are points in a space with a relatively small number of dimensions, which with a first approximation, seem to occupy two dimensions: *activation* and *evaluation*.

In this work we present a methodology for creating intermediate expressions based on archetypal ones and taking into account the results of Whissel's study and in particular the *activation* parameter. The proposed methodology consists of three legs: (i) Description of the archetypal expressions through particular FAPs. In order to do that we translate the facial muscle movements - describing expressions through muscle actions stills being a very popular approach- into FAPs and create a vocabulary of FAPs for each archetypal expression. FAPs required for the description of the archetypal expressions are also experimentally verified by analyzing some prototype datasets, like the Ekman's [1]. In order to make comparisons with real expression sequences, we model the FAPs employed in the facial expression formation through the movement of particular FDPs –the selected FDPs could be automatically detected from real images or video sequences. This modeling can serve also as a bridge between the expression analysis and expression synthesis disciplines [9]. (ii) Estimation of the range of variation for the FAPs that involved in each of the archetypal expressions. This estimation is by analyzing real images and video sequences as well as by animating synthesized examples. (iii) Combination, in the framework of a rule base system, of the *activation* parameter –known from Whissel's study and being available for a variety of non-archetypal expressions- with the description of the archetypal expressions through the FAPs, for the modeling of intermediate expressions.

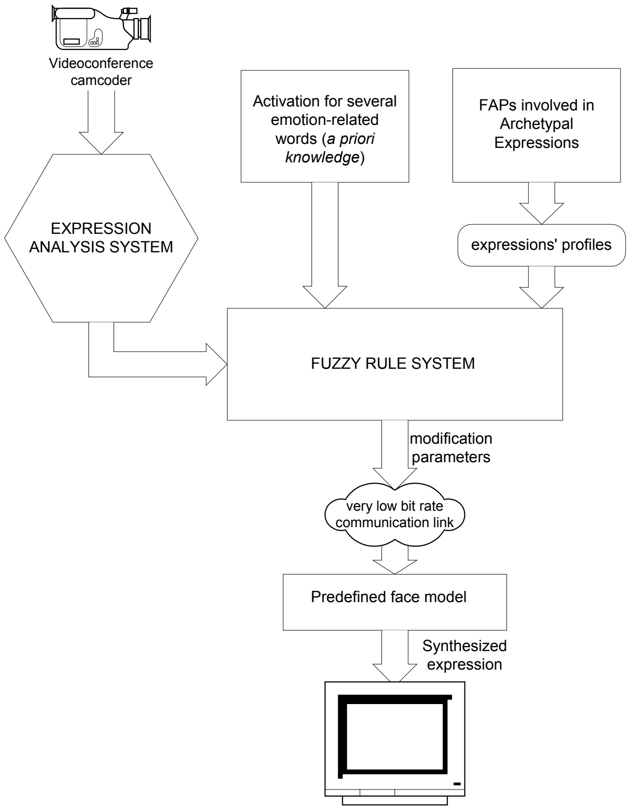


***Figure 1**: Block diagram of the proposed scheme*

The way the overall system functions is illustrated in Figure 1. The expression analysis system provides either information about the expression conveyed by the real subject, or the movement of particular points within the facial area –corresponding to specific FDPs. In the former case, the fuzzy system uses the predefined expressions' profiles as well as the activation parameter of the recognized expression, so as to provide the modification parameters - FAPs coupled with their appropriate range of variation - required by the client side application to animate the expression. In the latter case, the fuzzy rule system translates the FDPs movement to FAPs; if a predefined expression profile is recognized then the corresponding modification parameters are sent to the client side, otherwise the estimated FAP values are sent. The difference between the two cases is that a predefined profile corresponds to an expression while animating the estimated FAP values does not guarantee the creation of an expression.

Videoconferencing enables people to see, hear, discuss, teach or be taught as well as share documents and other information, utilizing low bit rates. Face isolation and FDPs detection is easier in such environments, making the proposed scheme appropriate candidate to minimize the exchanged bit rates by synthesizing the expressions at the client side, supposing that a face model is available. Even if a model is not available, the client side has the capability, using FDPs, to completely specify which face model has to animate.

2. DESCRIBING ARCHETYPAL EXPRESSIONS THROUGH FAPs

In general, facial expressions and emotions can be described as a set of measurements and transformations that can be considered atomic with respect to the MPEG-4 standard; this way, one can describe both the anatomy of a human face –basically through the FDPs, as well as any animation parameters with groups of distinct tokens, eliminating the need to specify the topology of the underlying geometry. These tokens can then be mapped to automatically detected measurements and indications of motion on a video sequence and thus help approximate a real expression conveyed by the subject with a synthetic one. Modeling facial expressions and the underlying emotions through the definition of FAPs serves several purposes: (i) Keeps the compatibility of synthetic sequences, created using the proposed methodology, with the MPEG-4 standard. (ii) Archetypal expressions occur rather infrequently, in most cases emotions are expressed by the variation of a few discrete facial features which are directly related with particular FAPs. Moreover, distinct FAPs can be utilized for communication between humans and computers in a paralinguistic form –expressed by facial signs. (iii) FAPs do not correspond to specific models or topologies; synthetic expressions can be animated by different (to the one that corresponds to the real subject) models or characters.

For the modeling of archetypal expression two basic issues should be addressed: (i) estimation of the FAPs that involved in their formation, (ii) definition of the FAPs intensities. The former is examined in the current section while the latter is explained in Section 3.

Description of archetypal expression by means of muscle movement and FACS AUs it was our starting point for setting the archetypal expression description through FAPs. Hints for the modeling were obtained from psychological studies [4][10][11] which refer to face formation during the expressions, as well as from experimental data provided from classic databases like Ekman's [1] and MediaLab's [12] –see also Section 3. Table 1 illustrates our proposition for the description of the archetypal expressions and some variations of them, using the MPEG-4 FAPs terminology. It should be noted however, that the sets shown in Table 2 consist the vocabulary of FAPs for each archetypal expression and not a particular profile for synthesizing expressions; this means that if animated they would not necessarily produce the corresponding expression. We define an *expression profile* as a subset of the FAPs vocabulary, corresponding to a particular expression, accompanied with FAP intensities –actually range of variation- which if animated creates the requested expression. Several expression profiles based on the FAPs vocabulary proposed in Table 1 are shown in the following section.

3. THE RANGE OF VARIATION OF FAPs IN REAL VIDEO SEQUENCES

An important issue, useful to both emotion analysis and synthesis systems, is the range of variation of the FAPs that involved in the facial expression formation. From the synthesis point of view, a study has been carried out by Tekalp [13] which refers to the FAPs range definition. However the suggested ranges of variation are rather loose and cannot be used for analysis purpose. In order to have a clear hint about the FAPs range of variation in real video sequences we analyzed datasets of real subjects, showing archetypal expressions and computed some statistics about the involved FAPs. However, for analysing expressions of real subjects a way of modeling FAPs through the movement of facial points is required. Analysis of FAPs range of variation in real images and video sequences is used for two purposes: (i) To verify and complete the proposed vocabulary for each archetypal expression, (ii) to define the profiles for the archetypal expressions.

3.1. Modeling FAPs through FDPs movement

Although the FAPs are practical and very useful for animation purposes, they are inadequate for analyzing facial expressions from video scenes or still images. The main reason for that is the absence of quantitative definitions for the FAPs as well as their non-additive nature. In order to be able to measure the FAPs in real images and video sequences we should define a way of describing them through the movement of some points that lie in the facial area and being able to be automatically detected. Such a modelling could get advantage of the extended research made on automatic facial points detection. Quantitative modeling of FAPs is implemented by using the features labeled as f_i ($i=1..15$) in Table 2. The feature set employs FDPs that lie in the facial area and, under some constraints, can be automatically detected and tracked. It consists of distances, noted as $s(x,y)$ where x and y correspond to FDP points shown in Figure 2(b), between these protuberant points, some of which are constant during the expressions and are used as reference points. Distances between the reference points are used for normalization (see Figure 2(a)). The units for the f_i are identical to those corresponding to FAPs even in cases where no one to one relation exists.

It should be noted that not all FAPs included in the vocabularies shown in Table 1 can be modeled by distances between facial protuberant points (e.g. *raise_b_lip_lm_o*, *lower_t_lip_lm_o*). In such cases the corresponding FAPs are retained in the vocabularies and their ranges of variation are experimentally estimated using animations. Moreover, some features serve for the estimation of range of variation of more than one FAP (e.g. features f_{12}-f_{15}).

3.2. Vocabulary verification

In order to have a clear hint about the FAPs range of variation in real video sequences, as well as for verifying the vocabulary of FAPs involved in each archetypal emotion, we analyzed two well-known datasets, showing archetypal expressions: Ekman's [1] and MediaLab's [12]. Analysis was based on the FAPs quantitative modeling described in the previous section. Computed statistics are summarized in Table 4. Mean values provide typical values that can be used for the particular expression profiles while the standard deviation is involved so as to define the range of variation (see also Section 3.3). The units of shown values are those of the corresponding FAPs [13]. Moreover, symbol (*) expresses the absence of the corresponding FAP in the vocabulary of the particular expression while symbol (–) shows that although the corresponding FAP is included in the vocabulary has not been verified by the statistical analysis. The latter case shows that not all FAPs included in the vocabularies are experimentally verified.

The detection of the facial points subset used to describe the FAPs involved in the archetypal expressions was based on the work presented in [14]. However, for accurate detection in many cases human assistance was necessary. The authors are working towards a fully automatic implementation of the point detection procedure.

3.3. Creating archetypal expression profiles

An archetypal *expression profile* is a set of FAPs accompanied by the corresponding range of variation, which, if animated, produce a visual impression of the corresponding emotion. Typically a profile of an archetypal expression consists of a subset of the corresponding FAPs vocabulary coupled with the appropriate ranges of variation. The statistical nature of the expression analysis performed in the two datasets is useful for the FAPs vocabulary completion and verification, as well as for a rough estimation of the range of variation of FAPs, but not for the profile creation. In order to define exact profiles for the archetypal expressions we combined the following three steps several times: (a) we set subsets of FAPs that probably form an archetypal expression, by translating the proposed by psychological studies [4][10][11] face formations to FAPs, (b) we used, at the first stage, the corresponding ranges of variations obtained from Table 3 and (c) we animated the corresponding profiles so as to see if we have the expected impression

The initial range of variation for the FAPs has been computed as follows: Let m_{ij} and σ_{ij} be the mean value and standard deviation of FAP F_j for the archetypal expression *i* (where i ={1⇒Anger, 2⇒Sadness, 3⇒Joy, 4⇒Disgust, 5⇒Fear, 6⇒Surprise}), as estimated in Table 3. The initial range of variation X_{ji} of FAP F_j for the archetypal expression *i* is defined as:

$$X_{ji} = [m_{ji} - \sigma_{ij}, \ m_{ji} + \sigma_{ji}] \ [1]$$

for bi-directional, and

$$X_{ji} = [0, \ m_{ji} + \sigma_{ji}] \text{ or } X_{ji} = [\ m_{ji} - \sigma_{ij}, \ 0] \ [2]$$

for unidirectional FAPs [14].

Generally speaking, for animation purposes every MPEG-4 decoder has to provide and use an MPEG-4 compliant face model, whose geometry can be defined using FDPs, or it can define the animation rules being based on Face Animation Tables (FAT). Using FATs we can specify which model vertices should be moved for each FAP and how and we can indicate the transformed nodes of the face as well the kind of the transformation. For our experiments on setting the archetypal expression profiles we used the face model developed in the context of the European Project *ACTS MoMuSys* [15], being freely available at the website http://www.iso.ch/ittf.

Figure 3 shows some examples of animated profiles. Figure 3(a) shows a particular profile for the archetypal expression "anger", while Figures 3(b) and (c) shows an alternative profile of the same expression and the difference between them is due to the FAP intensities. Difference in FAP intensities is also shown in Figures (d) and (e), both illustrating the same profile of expression "surprise". Finally Figure 3(f) shows an example of a profile of the expression "happy".

4. CREATING PROFILES FOR INTERMEDIATE EXPRESSIONS

One of the studies carried out from psychologists and which could be useful to researchers of the area of computer graphics and machine vision is the one of Whissel's [8], who suggested that emotions are points in a space with a relatively small number of dimensions, which with a first approximation, seem to occupy two dimensions: *activation* and *evaluation*. *Activation* is the degree of arousal associated with the term, with terms like patient (at 3.3) representing a midpoint, surprised (over 6) representing high activation, and bashful (around 2) representing low *activation*. *Evaluation* is the degree of pleasantness associated with the term, with guilty (at 1.1) representing the negative extreme and delighted (at 6.6) representing the positive extreme. From the practical point of view, *evaluation* seems to express internal feelings of the subject and is estimation through face formations is intractable. On the other hand, *activation* is related to the facial muscles movement and can be more easily estimated based on facial characteristics.

Grading of FAPs is strongly related with the *activation* parameter proposed by Whissel. Since this relation is expressed differently for the particular expressions, a fuzzy rule system seems appropriate for mapping FAPs to the *activation* axis. Let a_X be the *activation* parameter for the expression X . According to the modeling given in Table 1, a_X is a function of several FAPs whose initial range of variation is estimated using the statistics in Table 3 and equations (1)(2). If a_Y represents the *activation* parameter for the expression *Y,* how the difference between a_X and a_Y is disseminated in the involved FAPs, and the corresponding ranges of variations, for the two expressions? This dissemination is our basis for estimating the modification parameters that should applied to the predefined face model.

Generally speaking there are two cases: the expressions *X* and *Y* are similar (belong to the same category and involve the same FAPs e.g. *happy* and *joy*) or *X* and *Y* belong to separate categories. The first case is examined in the following paragraph while the authors are currently working in the second.

Disseminating the activation variance to the constituent FAPs is far from being straightforward. Possible cues consist the degree of freedom for the involved facial parts e.g. mouth, eyebrows, eyes etc., and available material that shows the archetypal expressions.

Expressions corresponding to high values of *activation* tend to employ more FAPs than the ones correspond to lower values. Furthermore, the FAPs values for the highly activated expressions are increased. For example surprised, which corresponds to the higher *activation,* involves almost all FAPs with extreme values while in the opposite side sadness employs very few FAPs with low values. Based on the previous observations the activation parameter can be approximated through a linear function of FAPs, i.e. $a_X = \sum_{i \in V_X} w_{Xi} \cdot F_i$

where *X* refers to the particular expression, V_x is the vocabulary of FAPs for expression *X* and w_{Xi} are constants corresponding to expression *X.*

4.1. Same Universal Emotion Category

Our approach for estimating the *activation* dissemination to the FAPs is as follows: As a general rule, one can define six general categories, each one characterized by a fundamental universal emotion; within each of these categories intermediate expressions are described by different emotional and optical intensities, as well as minor variation in expression details. From the synthetic point of view, expressions and emotions that belong in the same category can be rendered by animating the same FAPs in different intensities. For example, the emotion group "fear" also contains "worry" and "terror"; profiles of these two emotions can be synthesized using the profiles of "fear" by translating the range of variation of the employed FAPs, respectively. For example if a_w and a_f are the activation parameters for "worry" and "fear" and the range of variation for a FAP F_i belonging to a "fear" profile is X_{fi} then the corresponding range of variation for the F_i belonging to a "worry" profile is $X_{wi} = (a_w/a_f)*X_{fi}$. The same rationale can also be applied in the group of "disgust" that also contains "disdain" and "repulsion" – the fuzziness that is introduced by the range of variations provides assistance in differentiating mildly the output in similar situations. This ensures that the synthesis will not render "robot-like" animation, but drastically more realistic results.

5. CONCLUSION - DISCUSSION

In this work we proposed a complete framework for creating visual profiles, based on FAPs, for intermediate (not primary) emotions. Emotion profiles can serve either the vision part of an emotion recognition system, or a client side application that creates synthetic expressions. The main advantage of the proposed system is its flexibility: (a) No hypothesis about what the expression analysis system is –see Figure 1-, should be made; it's enough to provide either the name of the conveyed emotion, or just the movement of a predefined set of FDPs. In the former case, the proposed fuzzy system serves as an agent for synthesizing expressions, while in the latter case it functions as an autonomous emotion analysis system. (b) It is extensible w.r.t completing (or modifying) the proposed vocabulary of FAPs for the archetypal expressions. (c) The range of variation of FAPs that involved in the archetypal expression profiles can been modified. Note however that this modification affects the profiles that created for intermediate expressions. (d) It is extensible w.r.t the number of intermediate expressions that can be modeled.

6. REFERENCES

[1] P. Ekman and W. Friesen, *The Facial Action Coding System*, Consulting Psychologists Press, San Francisco, CA, 1978 (http://www.paulekman.com).

[2] *EC TMR Project PHYSTA Report*, "Development of Feature Representation from Facial Signals and Speech," January 1999.

[3] P. Ekman, "Facial expression and Emotion," *Am. Psychologist*, vol. 48 pp.384-392, 1993

[4] P. Ekman and W.V. Friesen, "Pictures of Facial Affect", Palo Alto, CA: Consulting Psychologists Press, 1978.

[5] C. Izard, L. Dougherty, and E.A. Hembree, "A System for Identifying Affect Expressions by Holistic Judgements", *Technical Report, Univ. of Delaware*, 1983.

[6] R. Cowie R. and E. Douglas-Cowie, "Speakers and hearers are people: reflections on speech deterioration as a consequence of acquired deafness," in K-E. Spens and G. Plant (Eds) *Profound deafness and speech communication*. Whurr Publications, London, 510-527, 1995.

[7] R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz and J. Taylor, "Emotion Recognition in Human-Computer Interaction", *IEEE Signal Processing Magazine*, January 2001.

[8] C. M. Whissel, *The dictionary of affect in language*, R. Plutchnik and H. Kellerman (Eds) "*Emotion: Theory, research and experience: vol 4, The measurement of emotions*", Academic Press, New York, 1989

[9] N. Tsapatsoulis, K. Karpouzis, G. Stamou, F. Piat and S. Kollias, "A Fuzzy System for Emotion Classification based on the MPEG-4 Facial Definition Parameter Set", *EUSIPCO 2000*, Tampere, Finland, September 2000.

[10] F. Parke and K. Waters, *Computer Facial Animation*, A K Peters, 1996

[11] G. Faigin, "The Artist's Complete Guide to Facial Expressions," Watson-Guptill, New York, 1990.

[12] I.A. Essa and A.P. Pentland, "Coding, Analysis, Interpretation, and Recognition of Facial Expressions", IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 19, no. 7, pp. 757-763, July 1997.

[13] M. Tekalp, "Face and 2-D Mesh Animation in MPEG-4", *Tutorial Issue On The MPEG-4 Standard, Image Communication Journal*, Elsevier, 1999.

[14] Kin-Man Lam and Hong Yan, "An Analytic-to-Holistic Approach for Face Recognition Based on a Single Frontal View," *IEEE Trans. on PAMI*, vol. 20, no. 7, July 1998. *http://www.computer.org/tpami/tp1998/i7toc.htm*

[15] ISTFACE - MPEG-4 Facial Animation System – Version 3.3.1 Gabriel Abrantes, (Developed in the context of the European Project ACTS MoMuSys © 97-98 Instituto Superior Tecnico).

Joy	*Close_t_l_eyelid, close_t_r_eyelid , close_b_l_eyelid, close_b_r_eyelid, stretch_l_cornerlip, stretch_l_cornerlip_o, stretch_r_cornerlip_o, raise_l_m_eyebrow, raise_r_m_eyebrow, lower_t_midlip, lift_r_cheek, lift_l_cheek, open_jaw, raise_b_midlip*
Sadness	*raise_l_i_eyebrow, raise_r_i_eyebrow, close_t_l_eyelid, close_t_r_eyelid, raise_l_m_eyebrow, raise_r_m_eyebrow, raise_l_o_eyebrow, raise_r_o_eyebrow, close_b_l_eyelid, close_b_r_eyelid*
Anger	*squeeze_l_eyebrow, squeeze_r_eyebrow, lower_t_midlip, raise_b_midlip, raise_l_i_eyebrow, raise_r_i_eyebrow, close_t_r_eyelid, close_t_l_eyelid, close_b_r_eyelid, close_b_l_eyelid*
Fear	*raise_l_o_eyebrow, raise_r_o_eyebrow, raise_l_m_eyebrow, raise_r_m_eyebrow, raise_l_i_eyebrow, raise_r_i_eyebrow, squeeze_l_eyebrow, squeeze_r_eyebrow, open_jaw, close_t_l_eyelid, close_t_r_eyelid, lower_t_midlip, lower_t_midlip*
Disgust	*close_t_l_eyelid, close_t_r_eyelid, close_b_l_eyelid, close_b_r_eyelid, lower_t_midlip, open_jaw, lower_t_lip_lm, raise_b_lip_lm, lower_t_lip_lm_o, raise_b_lip_lm_o, raise_l_cornerlip_o, lower_t_lip_rm, raise_b_lip_rm, lower_t_lip_rm_o, raise_b_lip_rm_o, raise_r_cornerlip_o*
Surprise	*raise_l_o_eyebrow, raise_r_o_eyebrow, raise_l_m_eyebrow, raise_r_m_eyebrow, raise_l_i_eyebrow, raise_r_i_eyebrow, open_jaw, squeeze_l_eyebrow, squeeze_r_eyebrow*

Table 1: *FAPs vocabulary for archetypal expression description*

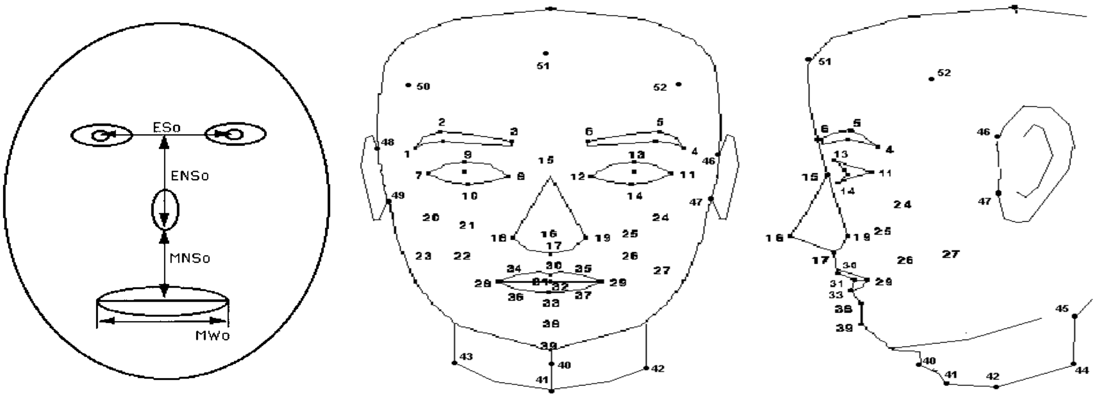
FAP name	Feature for the description	Utilized feature	Unit
Squeeze_l_eyebrow	$D_1=s(1,8)$	$f_1= D_{1\text{-}NEUTRAL} - D_1$	*ES*
Squeeze_r_eyebrow	$D_2=s(4,12)$	$f_2= D_{2\text{-}NEUTRAL} - D_2$	*ES*
Lower_t_midlip	$D_3=s(16,30)$	$f_3= D_3 - D_{3\text{-}NEUTRAL}$	*MNS*
Raise_b_midlip	$D_4=s(16,33)$	$f_4= D_{4\text{-}NEUTRAL} - D_4$	*MNS*
Raise_l_i_eyebrow	$D_5=s(3,8)$	$f_5= D_5 - D_{5\text{-}NEUTRAL}$	*ENS*
Raise_r_i_eyebrow	$D_6=s(6,12)$	$f_6= D_6 - D_{6\text{-}NEUTRAL}$	*ENS*
Raise_l_o_eyebrow	$D_7=s(1,7)$	$f_7= D_7 - D_{7\text{-}NEUTRAL}$	*ENS*
Raise_r_o_eyebrow	$D_8=s(4,11)$	$f_8= D_8 - D_{8\text{-}NEUTRAL}$	*ENS*
Raise_l_m_eyebrow	$D_9=s(2,7)$	$f_9= D_9 - D_{9\text{-}NEUTRAL}$	*ENS*
Raise_r_m_eyebrow	$D_{10}=s(5,11)$	$f_{10}= D_{10} - D_{10\text{-}NEUTRAL}$	*ENS*
Open_jaw	$D_{11}=s(30,33)$	$f_{11}= D_{11} - D_{11\text{-}NEUTRAL}$	*MNS*
close_t_l_eyelid – close_b_l_eyelid	$D_{12}=s(9,10)$	$f_{12}= D_{12} - D_{12\text{-}NEUTRAL}$	*IRISD*
close_t_r_eyelid – close_b_r_eyelid	$D_{13}=s(13,14)$	$f_{13}= D_{13} - D_{13\text{-}NEUTRAL}$	*IRISD*
stretch_l_cornerlip (stretch_l_cornerlip_o) – stretch_r_cornerlip (stretch_r_cornerlip_o)	$D_{14}=s(28,29)$	$f_{14}= D_{14} - D_{14\text{-}NEUTRAL}$	*MW*
squeeze_l_eyebrow AND squeeze_r_eyebrow	$D_{15}=s(1,4)$	$f_{15}= D_{15\text{-}NEUTRAL} - D_{15}$	*ES*

Table 2*: Quantitative FAPs modeling: (1)* ***s(x,y)*** *is the Euclidean distance between the FDPs* ***x*** *and* ***y*** *shown in Figure 2(b), (2)* $\mathbf{D_{i\text{-}NEUTRAL}}$ *refers to the distance* $\mathbf{D_i}$ *when the face is its in neutral position*

FAP name (symbol)	Stats	Anger	Sadness	Joy	Disgust	Fear	Surprise
open_jaw (F_3)	Mean	*	*	-	-	291	885
	StD	*	*	-	-	189	316
lower_t_midlip (F_4)	Mean	73	*	-271	-234	-	*
	StD	51	*	110	109	-	*
raise_b_midlip (F_5)	Mean	*	*	-	-177	218	543
	StD	*	*	-	108	135	203
stretch_l_cornerlip (F_6), stretch_l_cornerlip_o (F_{53}), stretch_r_cornerlip (F_7), stretch_r_cornerlip_o (F_{54})	Mean	*	*	234	*	*	-82
	StD	*	*	98	*	*	39
lower_t_lip_lm (F_8)	Mean	*	*	*	-	*	*
	StD	*	*	*	-	*	*
lower_t_lip_rm (F_9)	Mean	*	*	*	-	*	*
	StD	*	*	*	-	*	*
raise_b_lip_lm (F_{10})	Mean	*	*	*	-	*	*
	StD	*	*	*	-	*	*
raise_b_lip_rm (F_{11})	Mean	*	*	*	-	*	*
	StD	*	*	*	-	*	*
close_t_l_eyelid (F_{19}), close_b_l_eyelid (F_{21})	Mean	-	-153	-254	-203	244	254
	StD	-	112	133	148	126	83
close_t_r_eyelid (F_{20}), close_b_r_eyelid (F_{22})	Mean	-	-161	-242	-211	249	252
	StD		109	122	145	128	81
raise_l_i_eyebrow (F_{31})	Mean	-83	85	*	*	104	224
	StD	48	55	*	*	69	103
raise_r_i_eyebrow (F_{32})	Mean	-85	80	*	*	111	211
	StD	51	54	*	*	72	97
raise_l_m_eyebrow (F_{33})	Mean	-149	-	24	-80	72	144
	StD	40	-	22	53	58	64
raise_r_m_eyebrow (F_{34})	Mean	-144	-	25	-82	75	142
	StD	39	-	22	54	60	62
raise_l_o_eyebrow (F_{35})	Mean	-66	-	*	*	-	54
	StD	35	-	*	*	-	31
raise_r_o_eyebrow (F_{36})	Mean	-70	-	*	*	-	55
	StD	38		*	*	-	31
squeeze_l_eyebrow (F_{37})	Mean	57	*	*	*	-	-
	StD	28	*	*	*	-	-
squeeze_r_eyebrow (F_{38})	Mean	58	*	*	*	-	-
	StD	31	*	*	*	-	-
lift_l_cheek (F_{41})	Mean	*	*	-	*	*	*
	StD	*	*	-	*	*	*
lift_r_cheek (F_{42})	Mean	*	*	-	*	*	*
	StD	*	*	-	*	*	*
stretch_l_cornerlip_o (F_{53})	Mean	*	*	-	*	*	-
	StD	*	*	-	*	*	-
stretch_r_cornerlip_o (F_{54})	Mean	*	*	*	*	*	-
	StD	*	*	*	*	*	
lower_t_lip_lm_o (F_{55})	Mean	*	*	*	-	*	*
	StD	*	*	*	-	*	*

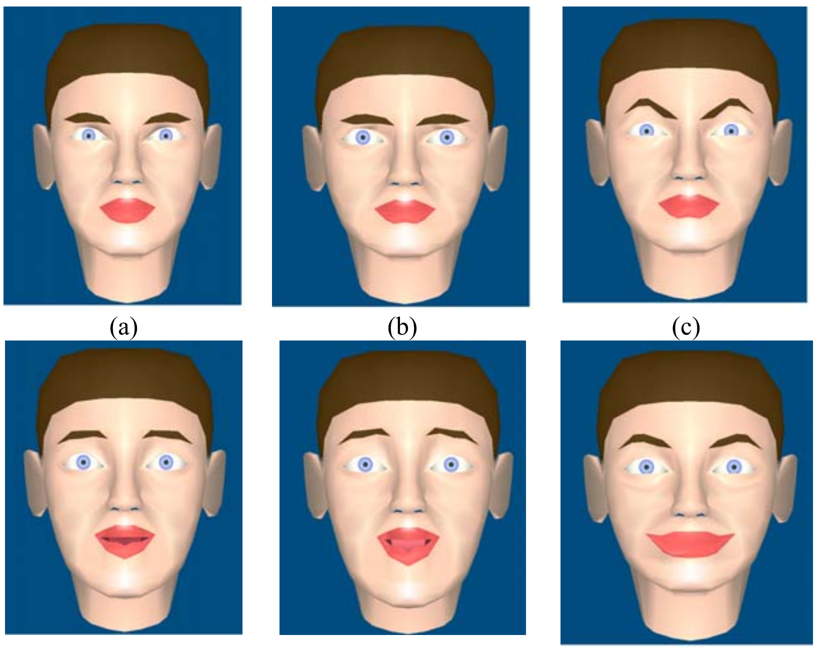
lower_t_lip_rm_o (F_{56})	Mean	*	*	*	-	*	*
	StD	*	*	*	-	*	*
raise_b_lip_lm_o (F_{57})	Mean	*	*	*	-	*	*
	StD	*	*	*	-	*	*
raise_b_lip_rm_o (F_{58})	Mean	*	*	*	-	*	*
	StD	*	*	*	-	*	*
raise_l_cornerlip_o (F_{59})	Mean	*	*	*	-	*	*
	StD	*	*	*	-	*	*
raise_r_cornerlip_o (F_{60})	Mean	*	*	*	-	*	*
	StD	*	*	*	-	*	*

Table 3: *Statistics for the vocabulary of FAPs for the archetypal expression: The symbol (*) expresses the absence of the corresponding FAP in the vocabulary of the particular expression while symbol (–) shows that although the corresponding FAP is included in the vocabulary has not been verified by the statistical analysis*



(a) (b)

Figure 2*: (a) Normalization distances (b) FDPs*



(d) (e) (f)

Figure 3: *Synthesized expressions (a)-(c) anger, (d)-(e) surprise,(f) joy*