Parameterized Facial Expression Synthesis Based on MPEG-4

Amaryllis Raouzaiou

Department of Electrical and Computer Engineering, National Technical University of Athens,
Heroon Polytechniou 9, 15773 Zographou, Greece
Email: araouz@image.ntua.gr

Nicolas Tsapatsoulis

Department of Electrical and Computer Engineering, National Technical University of Athens,
Heroon Polytechniou 9, 15773 Zographou, Greece
Email: ntsap@image.ntua.gr

Kostas Karpouzis

Department of Electrical and Computer Engineering, National Technical University of Athens,
Heroon Polytechniou 9, 15773 Zographou, Greece
Email: kkarpou@softlab.ntua.gr

Stefanos Kollias

Department of Electrical and Computer Engineering, National Technical University of Athens,
Heroon Polytechniou 9, 15773 Zographou, Greece
Email: stefanos@cs.ntua.gr

In the framework of MPEG-4, one can include applications where virtual agents, utilizing both textual and multisensory data, including facial expressions and nonverbal speech, help systems become accustomed to the actual feelings of the user. Applications of this technology are expected in educational environments, virtual collaborative workplaces, communities, and interactive entertainment. Facial animation has gained much interest within the MPEG-4 framework; with implementation details being an open research area (Tekalp, 1999). In this paper, we describe a method for enriching human computer interaction, focusing on analysis and synthesis of primary and intermediate facial expressions (Ekman and Friesen (1978)). To achieve this goal, we utilize facial animation parameters (FAPs) to model primary expressions and describe a rule-based technique for handling intermediate ones. A relation between FAPs and the *activation* parameter proposed in classical psychological studies is established, leading to parameterized facial expression analysis and synthesis notions, compatible with the MPEG-4 standard.

Keywords and phrases: facial expression, MPEG-4 facial definition parameters, activation, parameterized expression synthesis.

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 [1] have appeared in the computer science literature, which explore nonarchetypal emotions. This trend may be due to the great influence of the works of Ekman and Friesen [2, 3] and Izard et al. [4] who proposed that the archetypal emotions correspond to distinct facial expressions which are supposed to be universally recognizable across cultures. On the contrary, psychological researchers have extensively investigated [5, 6] a broader variety of emotions. An extensive survey on emotion analysis can be found in [7].

MPEG-4 indicates an alternative way of modeling facial expressions and the underlying emotions, which is strongly influenced by neurophysiological and psychological studies. The facial animation parameters (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) [8].

One of the studies carried out by psychologists and which can be useful to researchers of the area of computer graphics and machine vision is the one of Whissel [5], who suggested

that emotions are points in a space with a relatively small number of dimensions, which with a first approximation are only two: *activation* and *evaluation*. From the practical point of view, *evaluation* seems to express internal feelings of the subject and its 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.

In this work, we present a methodology for analyzing and synthesizing both primary and intermediate expressions, taking into account the results of Whissel's study and in particular the *activation* parameter. The proposed methodology consists of three steps.

Description of the archetypal expressions through particular FAPs

In order to do this, we translate facial muscle movements—describing expressions through muscle actions—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 through analysis of prototype datasets. In order to make comparisons with real expression sequences, we model FAPs employed in the facial expression formation through the movement of particular feature points (FPs)—the selected FPs can be automatically detected from real images or video sequences. The derived models can also serve as a bridge between expression analysis and expression synthesis disciplines [9].

Estimation of the range of variation of FAPs that are involved in each of the archetypal expressions

This is achieved by analyzing real images and video sequences as well as by animating synthesized examples.

Modelling of intermediate expressions

This is achieved through combination, in the framework of a rule-based system, of the *activation* parameter—known from Whissel's—with the description of the archetypal expressions by FAPs.

Figure 1 illustrates the way the proposed scheme functions. The facial expression synthesis system operates either by utilizing FAP values estimated by an image analysis subsystem, or by rendering actual expressions recognized by a fuzzy rules system. In the former case, protuberant facial points motion is analyzed and translated to FAP value variation, which in turn is rendered using the synthetic face model, so as to reproduce the expression in question. Should the results of the analysis coincide with the systems knowledge of the definition facial expression, then the expression can be rendered using predefined FAP alteration tables. These tables are computed using the known definition of archetypal emotions, fortified by video data of actual human expressions. In this case, any intermediate expressions can be rendered using interpolation rules derived by the emotion wheel.

The paper is organized as follows. In Sections 2, 3, and 4 the three legs of the proposed methodology are presented.
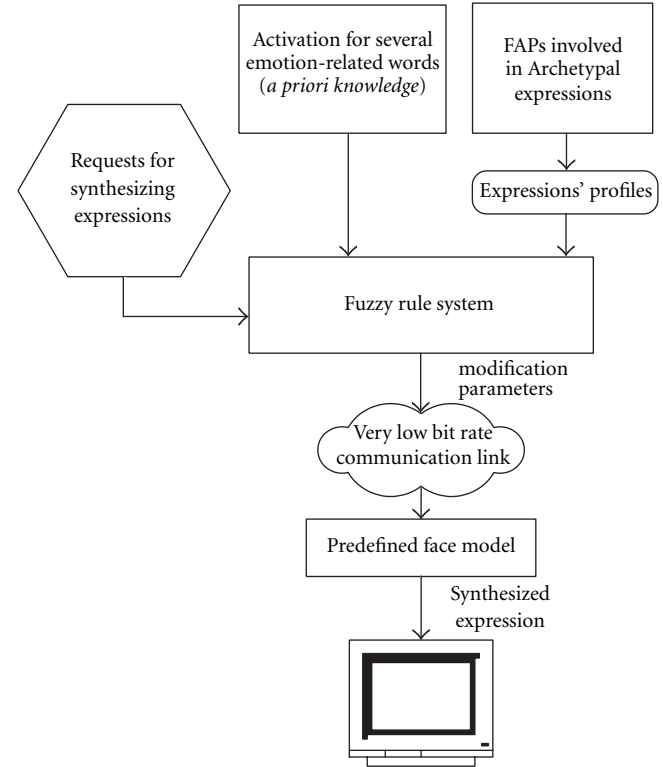

FIGURE 1: Block diagram of the proposed scheme.

In Section 5, a way of utilizing the proposed emotions synthesis scheme for emotion analysis purposes is described. In Section 6, experimental results, which illustrate the performance of the presented approach, are given. Finally, conclusions are given in Section 7.

2. DESCRIPTION OF THE ARCHETYPAL EXPRESSIONS USING FAPS

In general, facial expressions and emotions are described by a set of measurements and transformations that can be considered atomic with respect to the MPEG-4 standard. In this way, we can describe both the anatomy of a human face—basically through FDPs, as well as animation parameters, with groups of distinct tokens, eliminating the need for specifying 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 to approximate a real expression conveyed by the subject by means of a synthetic one.

Modelling facial expressions and underlying emotions through FAPs serves several purposes:

(i) provides 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 through 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 (than the one that corresponds to the real subject) models or characters.

Two basic issues should be addressed when modelling archetypal expression: (i) estimation of FAPs that are involved in their formation, (ii) definition of the FAP intensities. The former is examined in the current section, while the latter is explained in Section 5.

It is a general truth that the facial action coding system (FACS) has influenced the research on expression analysis in a high degree. FACS is a system which tries to distinguish the visually distinguishable facial movements using the knowledge of facial anatomy. FACS uses Action Units (AU) as measurement units. An AU could combine the movement of two muscles or work in the reverse way, that is, split into several muscle movement.

MPEG-4 FAPs are also strongly related to the AU; this is shown in Table 1. Description of archetypal expressions by means of muscle movements and AUs has been the starting point for setting the archetypal expression description through FAPs.

Hints for this mapping were obtained from psychological studies [2, 10, 11] which refer to face formation during expression generation, as well as from experimental data provided by classic databases as Ekman's (static) and MediaLab's (dynamic)—see also Section 3. Table 2 illustrates the description of archetypal expressions and some variations of them, using the MPEG-4 FAP's terminology. It should be noted that the sets shown in Table 2 consist of the vocabulary of FAPs to be used 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. In the following, we define an *expression profile* to be a subset of the FAPs vocabulary, corresponding to a particular expression, accompanied with FAP intensities, that is the actual ranges of variation, which if animated creates the requested expression. Several expression profiles based on the FAPs vocabulary proposed in Table 2 are shown in the experimental results 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 are involved in facial expression formation. From the synthesis point of view, a study has been carried out [5] which refers to FAP's range definition. However, the suggested ranges of variation are rather loose and cannot be used for analysis purposes. In order to have clear cues about FAP's range of variation in real video sequences, we analyzed two well-known datasets, showing archetypal expressions, Ekman's

TABLE 1: FAP to AU mapping.

Action units	FAPs
AU1	raise_l_i_eyebrow + raise_r_i_eyebrow
AU2	raise_l_o_eyebrow + raise_r_o_eyebrow
AU3	
AU4	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
AU5	close_t_l_eyelid + close_t_r_eyelid
AU6	lift_l_cheek + lift_r_cheek
AU7	close_b_l_eyelid + close_b_r_eyelid
AU8	
AU9	lower_t_midlip + raise_nose + stretch_l_nose + stretch_r_nose
AU10	raise_nose (+ stretch_l_nose + stetch_r_nose) + lower_t_midlip
AU11	
AU12	push_t_lip + Push_b_lip (+ lower_lowerlip + lower_t_midlip + raise_b_midlip)
AU13	
AU14	
AU15	lower_l_cornerlip + lower_r_cornerlip
AU16	
AU17	depress_chin
AU18	
AU19	
AU20	raise_b_midlip + lower_l_cornerlip + lower_r_cornerlip + stretch_l_cornerlip + stretch_r_cornerlip + 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

(static) [2] and MediaLab's (dynamic) [12], and computed statistics about the involved FAPs. Both sets show extreme cases of expressions, rather than every day ones. However, they can be used for setting limits to the variance of the respective FAPs [13]. To achieve this, however, a way of modeling FAPs through the movement of facial points is required. Analysis of FAP's range of variation in real images and video sequences is used next for two purposes:

(i) to verify and complete the proposed vocabulary for each archetypal expression,

(ii) to define profiles of archetypal expressions.

3.1. Modeling FAPs through FP's movement

Although FAPs are practical and very useful for animation purposes, they are inadequate for analyzing facial expressions

TABLE 2: FAPs vocabulary for archetypal expression description.

Joy	open_jaw (F_3), lower_t_midlip (F_4), raise_b_midlip (F_5), stretch_l_cornerlip (F_6), stretch_r_cornerlip (F_7), raise_l_cornerlip (F_{12}), raise_r_cornerlip (F_{13}), close_t_l_eyelid (F_{19}), close_t_r_eyelid (F_{20}), close_b_l_eyelid (F_{21}), close_b_r_eyelid (F_{22}), raise_l_m_eyebrow (F_{33}), raise_r_m_eyebrow (F_{34}), lift_l_cheek (F_{41}), lift_r_cheek (F_{42}), stretch_l_cornerlip_o (F_{53}), stretch_r_cornerlip_o (F_{54})
Sadness	close_t_l_eyelid (F_{19}), close_t_r_eyelid (F_{20}), close_b_l_eyelid (F_{21}), close_b_r_eyelid (F_{22}), raise_l_i_eyebrow (F_{31}), raise_r_i_eyebrow (F_{32}), raise_l_m_eyebrow (F_{33}), raise_r_m_eyebrow (F_{34}), raise_l_o_eyebrow (F_{35}), raise_r_o_eyebrow (F_{36})
Anger	lower_t_midlip (F_4), raise_b_midlip (F_5), push_b_lip (F_{16}), depress_chin (F_{18}), close_t_l_eyelid (F_{19}), close_t_r_eyelid (F_{20}), close_b_l_eyelid (F_{21}), close_b_r_eyelid (F_{22}), raise_l_i_eyebrow (F_{31}), raise_r_i_eyebrow (F_{32}), raise_l_m_eyebrow (F_{33}), raise_r_m_eyebrow (F_{34}), raise_l_o_eyebrow (F_{35}), raise_r_o_eyebrow (F_{36}), squeeze_l_eyebrow (F_{37}), squeeze_r_eyebrow (F_{38})
Fear	open_jaw (F_3), lower_t_midlip (F_4), raise_b_midlip (F_5), lower_t_lip_lm (F_8), lower_t_lip_rm (F_9), raise_b_lip_lm (F_{10}), raise_b_lip_rm (F_{11}), close_t_l_eyelid (F_{19}), close_t_r_eyelid (F_{20}), close_b_l_eyelid (F_{21}), close_b_r_eyelid (F_{22}), raise_l_i_eyebrow (F_{31}), raise_r_i_eyebrow (F_{32}), raise_l_m_eyebrow (F_{33}), raise_r_m_eyebrow (F_{34}), raise_l_o_eyebrow (F_{35}), raise_r_o_eyebrow (F_{36}), squeeze_l_eyebrow (F_{37}), squeeze_r_eyebrow (F_{38})
Disgust	open_jaw (F_3), lower_t_midlip (F_4), raise_b_midlip (F_5), lower_t_lip_lm (F_8), lower_t_lip_rm (F_9), raise_b_lip_lm (F_{10}), raise_b_lip_rm (F_{11}), close_t_l_eyelid (F_{19}), close_t_r_eyelid (F_{20}), close_b_l_eyelid (F_{21}), close_b_r_eyelid (F_{22}), raise_l_m_eyebrow (F_{33}), raise_r_m_eyebrow (F_{34}), lower_t_lip_lm_o (F_{55}), lower_t_lip_rm_o (F_{56}), raise_b_lip_lm_o (F_{57}), raise_b_lip_rm_o (F_{58}), raise_l_cornerlip_o (F_{59}), raise_r_cornerlip_o (F_{60})
Surprise	open_jaw (F_3), raise_b_midlip (F_5), stretch_l_cornerlip (F_6), stretch_r_cornerlip (F_7), raise_b_lip_lm (F_{10}), raise_b_lip_rm (F_{11}), close_t_l_eyelid (F_{19}), close_t_r_eyelid (F_{20}), close_b_l_eyelid (F_{21}), close_b_r_eyelid (F_{22}), raise_l_i_eyebrow (F_{31}), raise_r_i_eyebrow (F_{32}), raise_l_m_eyebrow (F_{33}), raise_r_m_eyebrow (F_{34}), raise_l_o_eyebrow (F_{35}), raise_r_o_eyebrow (F_{36}), squeeze_l_eyebrow (F_{37}), squeeze_r_eyebrow (F_{38}), stretch_l_cornerlip_o (F_{53}), stretch_r_cornerlip_o (F_{54})

from video scenes or still images. The main reason for that is the absence of a clear quantitative definition of FAPs (at least of most of them) as well as their nonadditive nature. Note here that the same problem holds for the FACS action units. This is quite normal, due to the strong relationship between particular AUs and FAPs (see Table 1). In order to be able to measure 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 are able to be automatically detected. Such a description could get advantage of the extended research made on automatic facial points detection [14, 15]. Quantitative description of FAPs based on particular FPs, that correspond to protuberant facial points' movement, provides the means of bridging the gap between expression analysis and animation/synthesis. In the expression analysis case the nonadditive property of the FAPs can be addressed by a fuzzy rule system, similar to the one described later for creating profiles for intermediate expressions.

Quantitative modeling of FAPs is implemented using the features labeled as f_i ($i = 1, \ldots, 15$) in Table 3 [16]. The feature set employs FPs 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 feature points shown in Figure 2b, between these protuberant points, some of which are constant during expressions and are used as reference points. Distances between reference points are used for normalization (see Figure 2a). The units for 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 2 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 vocabulary and their ranges of variation are experimentally defined based on facial animations. Moreover, some features serve for the estimation of range of variation of more than one FAP (e.g., features f_{12}, f_{13}, f_{14}, and f_{15}).

3.2. Vocabulary verification

To obtain clear cues about the FAPs' range of variation in real video sequences, as well as to verify the vocabulary of FAPs involved in each archetypal emotion, we analyzed two well-known datasets, showing archetypal expressions:
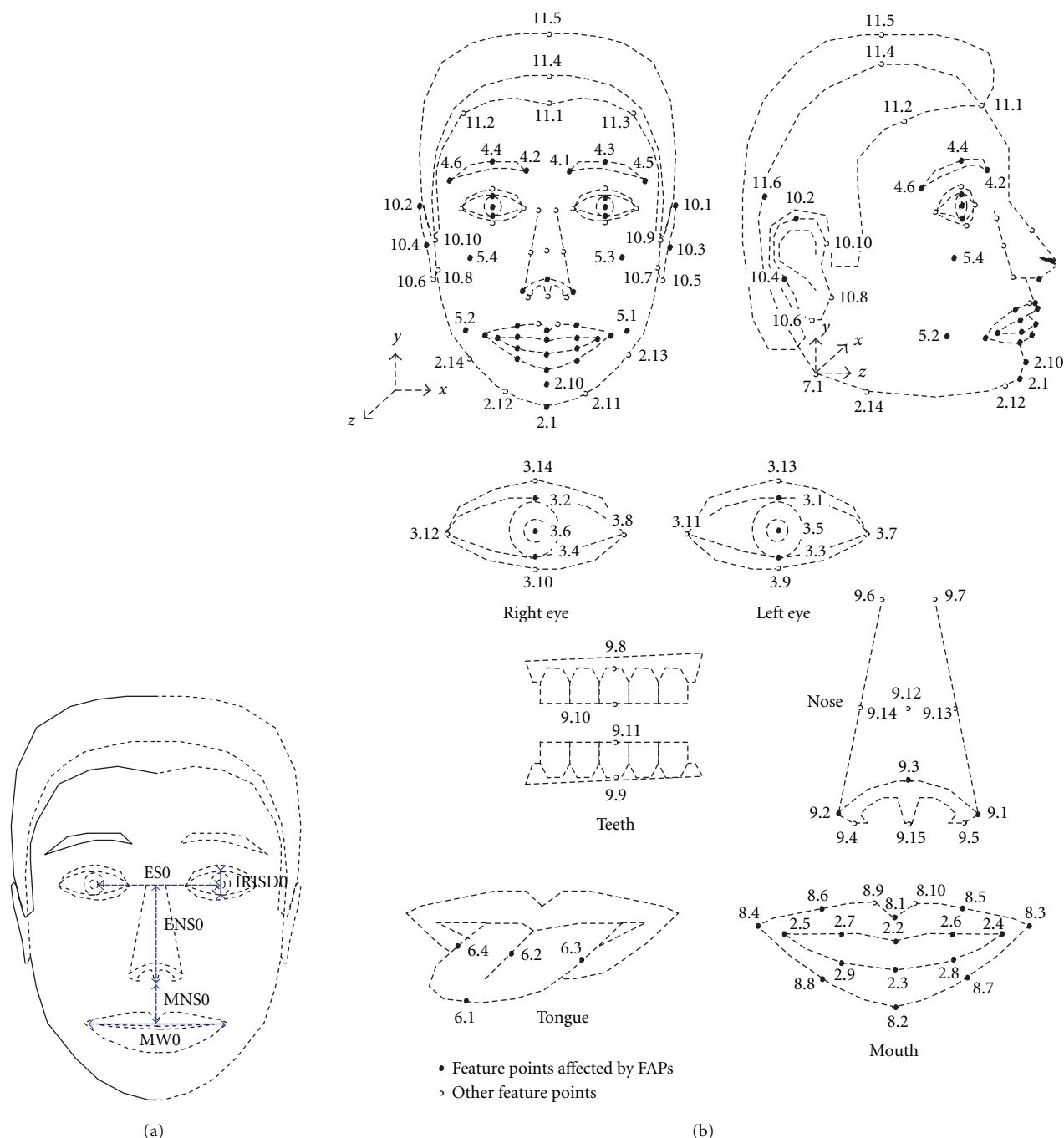
FIGURE 2: (a) A face model in its neutral state and the feature points used to define FAP units (FAPU). (b) Feature points (FPs).

Ekman's (static) [2] and MediaLab's (dynamic) [12]. The analysis was based on the FAPs' qualitative modelling described in the previous section. Computed statistics are summarized in Table 4. Mean values provide typical values that can be used for particular expression profiles, while the standard deviation can define the range of variation (see also Section 3.3). The units of shown values are those of the corresponding FAPs [17]. The symbol (∗) expresses the absence of the corresponding FAP in the vocabulary of the particular expression while the symbol (—) shows that, although the corresponding FAP is included in the vocabulary, it has not been verified by the statistical analysis. The latter case shows that not all FAPs included in the vocabulary 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 [18]. To obtain accurate detection, in many cases, human assistance was necessary. The authors are

TABLE 3: Quantitative FAPs modeling: (1) $s(\mathbf{x}, \mathbf{y})$ is the Euclidean distance between the FPs $\mathbf{x}$ and $\mathbf{y}$ shown in Figure 2b, (2) $\mathbf{D}_{i\text{-NEUTRAL}}$ refers to the distance $\mathbf{D}_i$ when the face is in its neutral position.

FAP name	Feature for the description	Utilized feature	Unit
squeeze_l_eyebrow (F_{37})	$D_1 = s(4.5, 3.11)$	$f_1 = D_{1\text{-NEUTRAL}} - D_1$	ES
squeeze_r_eyebrow (F_{38})	$D_2 = s(4.6, 3.8)$	$f_2 = D_{2\text{-NEUTRAL}} - D_2$	ES
lower_t_midlip (F_4)	$D_3 = s(9.3, 8.1)$	$f_3 = D_3 - D_{3\text{-NEUTRAL}}$	MNS
raise_b_midlip (F_5)	$D_4 = s(9.3, 8.2)$	$f_4 = D_{4\text{-NEUTRAL}} - D_4$	MNS
raise_l_i_eyebrow (F_{31})	$D_5 = s(4.1, 3.11)$	$f_5 = D_5 - D_{5\text{-NEUTRAL}}$	ENS
raise_r_i_eyebrow (F_{32})	$D_6 = s(4.2, 3.8)$	$f_6 = D_6 - D_{6\text{-NEUTRAL}}$	ENS
raise_l_o_eyebrow (F_{35})	$D_7 = s(4.5, 3.7)$	$f_7 = D_7 - D_{7\text{-NEUTRAL}}$	ENS
raise_r_o_eyebrow (F_{36})	$D_8 = s(4.6, 3.12)$	$f_8 = D_8 - D_{8\text{-NEUTRAL}}$	ENS
raise_l_m_eyebrow (F_{33})	$D_9 = s(4.3, 3.7)$	$f_9 = D_9 - D_{9\text{-NEUTRAL}}$	ENS
raise_r_m_eyebrow (F_{34})	$D_{10} = s(4.4, 3.12)$	$f_{10} = D_{10} - D_{10\text{-NEUTRAL}}$	ENS
open_jaw (F_3)	$D_{11} = s(8.1, 8.2)$	$f_{11} = D_{11} - D_{11\text{-NEUTRAL}}$	MNS
close_t_l_eyelid (F_{19})– close_b_l_eyelid (F_{21})	$D_{12} = s(3.1, 3.3)$	$f_{12} = D_{12} - D_{12\text{-NEUTRAL}}$	IRISD
close_t_r_eyelid (F_{20})– close_b_r_eyelid (F_{22})	$D_{13} = s(3.2, 3.4)$	$f_{13} = D_{13} - D_{13\text{-NEUTRAL}}$	IRISD
stretch_l_cornerlip (F_6) (stretch_l_cornerlip_o) (F_{53})– stretch_r_cornerlip (F_7) (stretch_r_cornerlip_o) (F_{54})	$D_{14} = s(8.4, 8.3)$	$f_{14} = D_{14} - D_{14\text{-NEUTRAL}}$	MW
squeeze_l_eyebrow (F_{37}) AND squeeze_r_eyebrow (F_{38})	$D_{15} = s(4.6, 4.5)$	$f_{15} = D_{15\text{-NEUTRAL}} - D15$	ES

working towards a fully automatic implementation of the FP detection procedure.

Figure 3 illustrates particular statistics, computed over the previously described datasets, for the expression *joy*. In all diagrams, horizontal axis shows the indices of the features of Table 3, while vertical axis shows the value of the corresponding feature: Figure 3a shows the minimum values of the features, Figure 3b the maximum values, and Figure 3c the mean values. From this figure, it is confirmed, for example, that *lower_t_midlip* (feature with index 3), which refers to lowering the middle of the upper lip, is employed, since even the maximum value for this FAP is below zero. In the same way, the FAPs raise_l_m_eyebrow, raise_r_m_eyebrow, close_t_l_eyelid, close_t_r_eyelid, close_b_l_eyelid, close_b_r_eyelid, stretch_l_cornerlip, stretch_r_cornerlip (indices 9, 10, 12, 13, 14) are verified. Some of the above FAPs are described using a single variable. For example the *stretch_l_cornerlip* and *stretch_r_cornerlip* are both modelled via f_{14}. The values, shown in Table 4, result by dividing the values of feature f_{14}. Similarly to Figure 3, Figure 4 illustrates particular statistics for the expression *surprise*.

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, produces a visual representation 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 expression analysis performed on the above mentioned datasets is useful for FAPs' vocabulary completion and verification, as well as for a rough estimation of the range of variation of FAPs, but not for profile creation. In order to define exact profiles for the archetypal expressions, we combined the following three steps:

(a) we defined subsets of FAPs that are candidates to form an archetypal expression, by translating the proposed by psychological studies [2, 10, 11] face formations to FAPs,

(b) we used the corresponding ranges of variations obtained from Table 4,

(c) we animated the corresponding profiles to verify appropriateness of derived representations.

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

$$X_{i,j} = [m_{i,j} - \sigma_{i,j}, m_{i,j} + \sigma_{i,j}] \tag{1}$$

TABLE 4: 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 the symbol (—) shows that although the corresponding FAP is included in the vocabulary has not been verified by the statistical analysis.

| 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	$*$	$*$	$*$	—	$*$	$*$

(a)

(b)
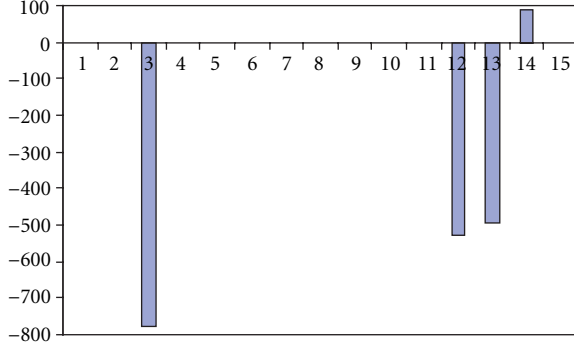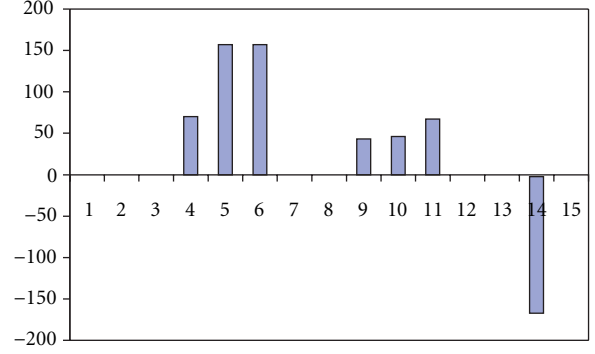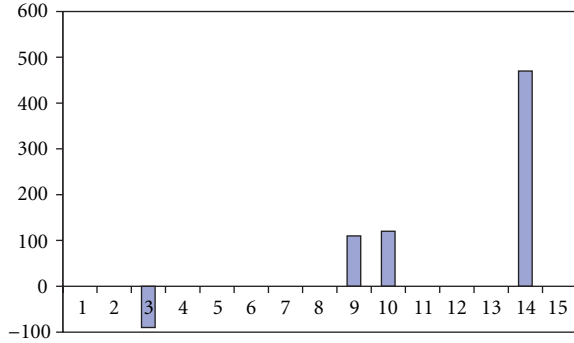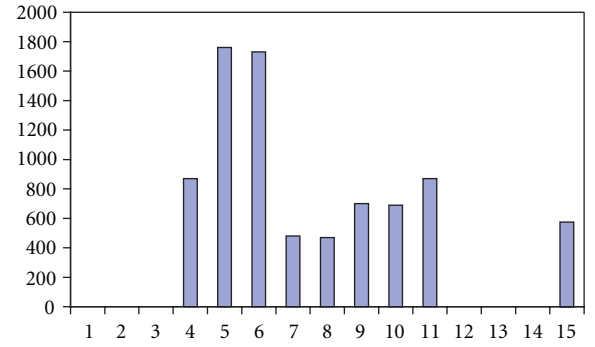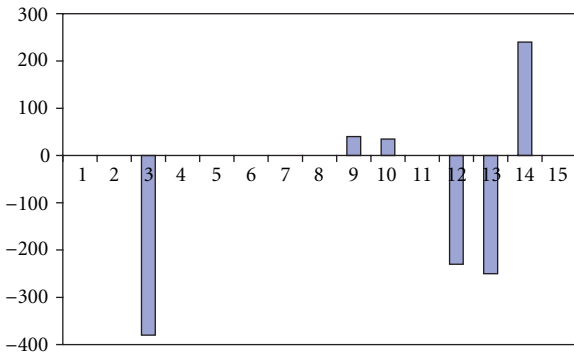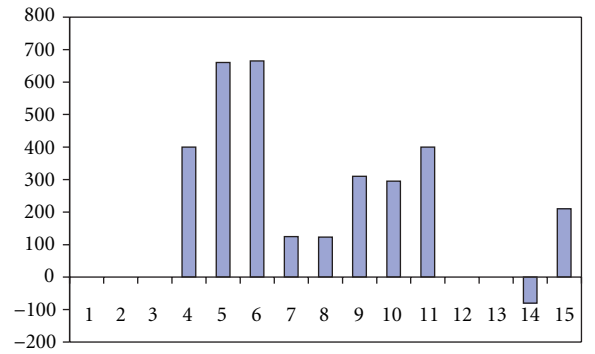
(c)

FIGURE 3: Computed statistics for the expression "joy." In all cases horizontal axis shows the indices of the features of Table 3 while vertical axis shows the value of the corresponding feature: (a) minimum values, (b) maximum values, (c) mean values.

(a)

(b)

(c)

FIGURE 4: Computed statistics for the expression "surprise." In all cases horizontal axis shows the indices of the features of Table 3 while vertical axis shows the value of the corresponding feature: (a) minimum values, (b) maximum values, (c) mean values.

for bi-directional, and

$$X_{i,j} = [\max(0, m_{i,j} - \sigma_{i,j}), m_{i,j} + \sigma_{i,j}] \quad (2)$$

or

$$X_{i,j} = [m_{i,j} - \sigma_{i,j}, \min(0, m_{i,j} + \sigma_{i,j})] \quad (3)$$

for unidirectional FAPs [17].

Generally speaking, for animation purposes, every MPEG-4 decoder has to provide and use an MPEG-4 compli-

ant face model whose geometry can be defined using FDPs, or should define the animation rules being based on face animation tables (FATs). Using FATs, we can explicitly specify the model vertices that will be spatially deformed for each FAP, as well as the magnitude of the deformation. This is in essence a mapping mechanism of each FAP, that represents a high-level semantic animation directive, to a lower-level, model specific deformation. An MPEG-4 decoder can use its own animation rules or receive a face model accompanied by the corresponding face animation tables (FATs) [19, 20]. For

our experiments on setting the archetypal expression profiles, we used the face model developed in the context of the European project ACTS MoMuSys [21], being freely available at http://www.iso.ch/ittf. Table 5 shows some examples of archetypal expression profiles, which were created based on our method.

Figure 5 shows some examples of animated profiles. Figure 5a shows a particular profile for the archetypal expression *anger*, while Figures 5b and 5c show alternative profiles of the same expression. The difference between them is due to FAP intensities. Difference in FAP intensities is also shown in Figures 5d and 5e, both illustrating the same profile of expression *surprise*. Finally Figure 5f shows an example of a profile of the expression *joy*.

4. CREATING PROFILES FOR INTERMEDIATE EXPRESSIONS

In this section we propose a way for creating profiles for intermediate expressions, used to describe the visual portion of corresponding emotions. The limited number of studies, carried out by computer scientists and engineers [7], dealing with emotions other than the archetypal ones, lead us to search in other subject/discipline bibliographies. Psychologists examined a broader set of emotions [13], but very few of the corresponding studies provide exploitable results to computer graphics and machine vision fields. One of these studies has been carried out by Whissel [5] and suggests that emotions are points in a space spanning a relatively small number of dimensions, which with a first approximation, seem to occupy two axes: *activation* and *evaluation*. *Activation* is the degree of arousal associated with the term, as shown in the "activation" column of Table 6, terms like *patient* (at 3.3 in Table 6) represent a midpoint, *surprised* (over 6) represent high activation, and *bashful* (around 2) represent low *activation*. *Evaluation* is the degree of pleasantness associated with the term, with *guilty*, as shown in the "evaluation" column of Table 6 at 1.1, representing the negative extreme and *delighted* (at 6.4) representing the positive extreme [5]. From the practical point of view, *evaluation* seems to express internal feelings of the subject and its estimation through face formations is intractable. On the other hand, *activation* is related to facial muscles' movement and can be easily estimated based on facial characteristics.

The third column in Table 6 represents Plutchik's [6] observation that emotion terms are unevenly distributed through the space defined by dimensions like Whissel's. Instead, they tend to form an approximately circular pattern called *emotion wheel*. Shown values refer to an angular measure, which runs from *Acceptance* (0) to *Disgust* (180).

For the creation of profiles for intermediate emotions we consider two cases:

(a) emotions that are similar, in nature, to an archetypal one; for example they may differ only in the intensity of muscle actions;

(b) emotions that cannot be considered as related to any of the archetypal ones.

In both cases we proceed by following the following steps:

(i) utilize either the *activation* parameter or Plutchik's *angular* measure as a priori knowledge about the intensity of facial actions for several emotions. This knowledge is combined with the profiles of archetypal expressions, through a rule based system, to create profiles for intermediate emotions;

(ii) animate the produced profiles for testing/correcting their appropriateness in terms of the visual similarity with the requested emotion.

4.1. Same universal emotion category

As a general rule, we can define six general categories, each one characterized by an archetypal 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, emotions that belong to the same category can be rendered by animating the same FAPs using different intensities. For example, the emotion group *fear* also contains *worry* and *terror* [11]; these two emotions can be synthesized by reducing or increasing the intensities of the employed FAPs, respectively. In the case of expression profiles, this affect the range of variation of the corresponding FAPs which is appropriately translated; the fuzziness, that is introduced by the varying scale of the change of FAP intensity, also 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.

Let $P_i^{(k)}$ be the kth profile of emotion i and $X_{i,j}^{(k)}$ be the range of variation of FAP F_j involved in $P_i^{(k)}$. If A, I are emotions belonging to the same universal emotion category, A being the archetypal, and I the intermediate one, then the following rules are applied.

Rule 1. $P_A^{(k)}$ and $P_I^{(k)}$ employ the same FAPs.

Rule 2. The range of variation $X_{I,j}^{(k)}$ is computed by $X_{I,j}^{(k)} = (a_I/a_A)X_{A,j}^{(k)}$.

Rule 3. a_A and a_I are the values of the *activation* parameter for emotion words A and I obtained from Whissel's study [5].

4.2. Emotions lying between archetypal ones

Creating profiles for emotions that do not clearly belong to a universal category is not straightforward. Apart from estimating the range of variations for FAPs, we should first define the vocabulary of FAPs for the particular emotion. In order to proceed, we utilize both the *emotion wheel* of Plutchik [6] and especially the *angular* measure (shown also in Table 6), and the *activation* parameter. Let I be an intermediate emotion lying between archetypal emotions A_1 and A_2 (which are supposed to be the nearest, with respect to the two sides of I emotions) according to their *angular* measure. Let also V_{A_1} and V_{A_2} be the vocabularies (sets of FAPs) corresponding to A_1 and A_2, respectively. The vocabulary V_I of

TABLE 5: Profiles for the archetypal emotions.

Profiles	FAPs and range of variation
Anger ($P_A^{(0)}$)	$F_4 \in [22, 124]$, $F_{31} \in [-131, -25]$, $F_{32} \in [-136, -34]$, $F_{33} \in [-189, -109]$, $F_{34} \in [-183, -105]$, $F_{35} \in [-101, -31]$, $F_{36} \in [-108, -32]$, $F_{37} \in [29, 85]$, $F_{38} \in [27, 89]$
$P_A^{(1)}$	$F_{19} \in [-330, -200]$, $F_{20} \in [-335, -205]$, $F_{21} \in [200, 330]$, $F_{22} \in [205, 335]$, $F_{31} \in [-200, -80]$, $F_{32} \in [-194, -74]$, $F_{33} \in [-190, -70]$, $F_{34} \in [-190, -70]$
$P_A^{(2)}$	$F_{19} \in [-330, -200]$, $F_{20} \in [-335, -205]$, $F_{21} \in [200, 330]$, $F_{22} \in [205, 335]$, $F_{31} \in [-200, -80]$, $F_{32} \in [-194, -74]$, $F_{33} \in [70, 190]$, $F_{34} \in [70, 190]$
$P_A^{(3)}$	$F_{16} \in [45, 155]$, $F_{18} \in [45, 155]$, $F_{19} \in [-330, -200]$, $F_{20} \in [-330, -200]$, $F_{31} \in [-200, -80]$, $F_{32} \in [-194, -74]$, $F_{33} \in [-190, -70]$, $F_{34} \in [-190, -70]$, $F_{37} \in [65, 135]$, $F_{38} \in [65, 135]$
$P_A^{(4)}$	$F_{16} \in [-355, -245]$, $F_{18} \in [145, 255]$, $F_{19} \in [-330, -200]$, $F_{20} \in [-330, -200]$, $F_{31} \in [-200, -80]$, $F_{32} \in [-194, -74]$, $F_{33} \in [-190, -70]$, $F_{34} \in [-190, -70]$, $F_{37} \in [65, 135]$, $F_{38} \in [65, 135]$
Sadness ($P_S^{(0)}$)	$F_{19} \in [-265, -41]$, $F_{20} \in [-270, -52]$, $F_{21} \in [-265, -41]$, $F_{22} \in [-270, -52]$, $F_{31} \in [30, 140]$, $F_{32} \in [26, 134]$
Joy ($P_J^{(0)}$)	$F_4 \in [-381, -161]$, $F_6 \in [136, 332]$, $F_7 \in [136, 332]$, $F_{19} \in [-387, -121]$, $F_{20} \in [-364, -120]$, $F_{21} \in [-387, -121]$, $F_{22} \in [-364, -120]$, $F_{33} \in [2, 46]$, $F_{34} \in [3, 47]$, $F_{53} \in [136, 332]$, $F_{54} \in [136, 332]$
$P_J^{(1)}$	$F_6 \in [160, 240]$, $F_7 \in [160, 240]$, $F_{12} \in [260, 340]$, $F_{13} \in [260, 340]$, $F_{19} \in [-449, -325]$, $F_{20} \in [-426, -302]$, $F_{21} \in [325, 449]$, $F_{22} \in [302, 426]$, $F_{33} \in [70, 130]$, $F_{34} \in [70, 130]$, $F_{41} \in [130, 170]$, $F_{42} \in [130, 170]$, $F_{53} \in [160, 240]$, $F_{54} \in [160, 240]$
$P_J^{(2)}$	$F_6 \in [160, 240]$, $F_7 \in [160, 240]$, $F_{12} \in [260, 340]$, $F_{13} \in [260, 340]$, $F_{19} \in [-449, -325]$, $F_{20} \in [-426, -302]$, $F_{21} \in [-312, -188]$, $F_{22} \in [-289, -165]$, $F_{33} \in [70, 130]$, $F_{34} \in [70, 130]$, $F_{41} \in [130, 170]$, $F_{42} \in [130, 170]$, $F_{53} \in [160, 240]$, $F_{54} \in [160, 240]$
$P_J^{(3)}$	$F_6 \in [160, 240]$, $F_7 \in [160, 240]$, $F_{12} \in [260, 340]$, $F_{13} \in [260, 340]$, $F_{19} \in [-449, -325]$, $F_{20} \in [-426, -302]$, $F_{21} \in [61, 185]$, $F_{22} \in [38, 162]$, $F_{33} \in [70, 130]$, $F_{34} \in [70, 130]$, $F_{41} \in [130, 170]$, $F_{42} \in [130, 170]$, $F_{53} \in [160, 240]$, $F_{54} \in [160, 240]$
Disgust ($P_D^{(0)}$)	$F_4 \in [-343, -125]$, $F_5 \in [-285, -69]$, $F_{19} \in [-351, -55]$, $F_{20} \in [-356, -66]$, $F_{21} \in [-351, -55]$, $F_{22} \in [-356, -66]$, $F_{33} \in [-123, -27]$, $F_{34} \in [-126, -28]$
Fear ($P_F^{(0)}$)	$F_3 \in [102, 480]$, $F_5 \in [83, 353]$, $F_{19} \in [118, 370]$, $F_{20} \in [121, 377]$, $F_{21} \in [118, 370]$, $F_{22} \in [121, 377]$, $F_{31} \in [35, 173]$, $F_{32} \in [39, 183]$, $F_{33} \in [14, 130]$, $F_{34} \in [15, 135]$
$P_F^{(1)}$	$F_3 \in [400, 560]$, $F_5 \in [333, 373]$, $F_{19} \in [-400, -340]$, $F_{20} \in [-407, -347]$, $F_{21} \in [-400, -340]$, $F_{22} \in [-407, -347]$
$P_F^{(2)}$	$F_3 \in [400, 560]$, $F_5 \in [307, 399]$, $F_{19} \in [-530, -470]$, $F_{20} \in [-523, -463]$, $F_{21} \in [-530, -470]$, $F_{22} \in [-523, -463]$, $F_{31} \in [460, 540]$, $F_{32} \in [460, 540]$, $F_{33} \in [460, 540]$, $F_{34} \in [460, 540]$, $F_{35} \in [460, 540]$, $F_{36} \in [460, 540]$
$P_F^{(3)}$	$F_3 \in [400, 560]$, $F_5 \in [-240, -160]$, $F_{19} \in [-630, -570]$, $F_{20} \in [-630, -570]$, $F_{21} \in [-630, -570]$, $F_{22} \in [-630, -570]$, $F_{31} \in [460, 540]$, $F_{32} \in [460, 540]$, $F_{37} \in [60, 140]$, $F_{38} \in [60, 140]$
$P_F^{(4)}$	$F_3 \in [400, 560]$, $F_5 \in [-240, -160]$, $F_{19} \in [-630, -570]$, $F_{20} \in [-630, -570]$, $F_{21} \in [-630, -570]$, $F_{22} \in [-630, -570]$, $F_{31} \in [460, 540]$, $F_{32} \in [460, 540]$, $F_{33} \in [360, 440]$, $F_{34} \in [360, 440]$, $F_{35} \in [260, 340]$, $F_{36} \in [260, 340]$, $F_{37} \in [60, 140]$, $F_{38} \in [60, 140]$
$P_F^{(5)}$	$F_3 \in [400, 560]$, $F_5 \in [-240, -160]$, $F_{19} \in [-630, -570]$, $F_{20} \in [-630, -570]$, $F_{21} \in [-630, -570]$, $F_{22} \in [-630, -570]$, $F_{31} \in [460, 540]$, $F_{32} \in [460, 540]$, $F_{33} \in [360, 440]$, $F_{34} \in [360, 440]$, $F_{35} \in [260, 340]$, $F_{36} \in [260, 340]$, $F_{37} \in 0$, $F_{38} \in 0$

TABLE 5: continued.

$P_F^{(6)}$	$F_3 \in [400, 560]$, $F_5 \in [-240, -160]$, $F_8 \in [-120, -80]$, $F_9 \in [-120, -80]$, $F_{10} \in [-120, -80]$, $F_{11} \in [-120, -80]$, $F_{19} \in [-630, -570]$, $F_{20} \in [-630, -570]$, $F_{21} \in [-630, -570]$, $F_{22} \in [-630, -570]$, $F_{31} \in [460, 540]$, $F_{32} \in [460, 540]$, $F_{33} \in [360, 440]$, $F_{34} \in [360, 440]$, $F_{35} \in [260, 340]$, $F_{36} \in [260, 340]$, $F_{37} \in 0$, $F_{38} \in 0$
$P_F^{(7)}$	$F_3 \in [400, 560]$, $F_5 \in [-240, -160]$, $F_{19} \in [-630, -570]$, $F_{20} \in [-630, -570]$, $F_{21} \in [-630, -570]$, $F_{22} \in [-630, -570]$, $F_{31} \in [360, 440]$, $F_{32} \in [360, 440]$, $F_{33} \in [260, 340]$, $F_{34} \in [260, 340]$, $F_{35} \in [160, 240]$, $F_{36} \in [160, 240]$
$P_F^{(8)}$	$F_3 \in [400, 560]$, $F_5 \in [-240, -160]$, $F_{19} \in [-630, -570]$, $F_{20} \in [-630, -570]$, $F_{21} \in [-630, -570]$, $F_{22} \in [-630, -570]$, $F_{31} \in [260, 340]$, $F_{32} \in [260, 340]$, $F_{33} \in [160, 240]$, $F_{34} \in [160, 240]$, $F_{35} \in [60, 140]$, $F_{36} \in [60, 140]$
$P_F^{(9)}$	$F_3 \in [400, 560]$, $F_5 \in [307, 399]$, $F_{19} \in [-630, -570]$, $F_{20} \in [-623, -563]$, $F_{21} \in [-630, -570]$, $F_{22} \in [-623, -563]$, $F_{31} \in [460, 540]$, $F_{32} \in [460, 540]$, $F_{33} \in [460, 540]$, $F_{34} \in [460, 540]$, $F_{35} \in [460, 540]$, $F_{36} \in [460, 540]$
Surprise ($P_{Su}^{(0)}$)	$F_3 \in [569, 1201]$, $F_5 \in [340, 746]$, $F_6 \in [-121, -43]$, $F_7 \in [-121, -43]$, $F_{19} \in [170, 337]$, $F_{20} \in [171, 333]$, $F_{21} \in [170, 337]$, $F_{22} \in [171, 333]$, $F_{31} \in [121, 327]$, $F_{32} \in [114, 308]$, $F_{33} \in [80, 208]$, $F_{34} \in [80, 204]$, $F_{35} \in [23, 85]$, $F_{36} \in [23, 85]$, $F_{53} \in [-121, -43]$, $F_{54} \in [-121, -43]$
$P_{Su}^{(1)}$	$F_3 \in [1150, 1252]$, $F_5 \in [-792, -700]$, $F_6 \in [-141, -101]$, $F_7 \in [-141, -101]$, $F_{10} \in [-530, -470]$, $F_{11} \in [-530, -470]$, $F_{19} \in [-350, -324]$, $F_{20} \in [-346, -320]$, $F_{21} \in [-350, -324]$, $F_{22} \in [-346, -320]$, $F_{31} \in [314, 340]$, $F_{32} \in [295, 321]$, $F_{33} \in [195, 221]$, $F_{34} \in [191, 217]$, $F_{35} \in [72, 98]$, $F_{36} \in [73, 99]$, $F_{53} \in [-141, -101]$, $F_{54} \in [-141, -101]$
$P_{Su}^{(2)}$	$F_3 \in [834, 936]$, $F_5 \in [-589, -497]$, $F_6 \in [-102, -62]$, $F_7 \in [-102, -62]$, $F_{10} \in [-380, -320]$, $F_{11} \in [-380, -320]$, $F_{19} \in [-267, -241]$, $F_{20} \in [-265, -239]$, $F_{21} \in [-267, -241]$, $F_{22} \in [-265, -239]$, $F_{31} \in [211, 237]$, $F_{32} \in [198, 224]$, $F_{33} \in [131, 157]$, $F_{34} \in [129, 155]$, $F_{35} \in [41, 67]$, $F_{36} \in [42, 68]$
$P_{Su}^{(3)}$	$F_3 \in [523, 615]$, $F_5 \in [-386, -294]$, $F_6 \in [-63, -23]$, $F_7 \in [-63, -23]$, $F_{10} \in [-230, -170]$, $F_{11} \in [-230, -170]$, $F_{19} \in [-158, -184]$, $F_{20} \in [-158, -184]$, $F_{21} \in [-158, -184]$, $F_{22} \in [-158, -184]$, $F_{31} \in [108, 134]$, $F_{32} \in [101, 127]$, $F_{33} \in [67, 93]$, $F_{34} \in [67, 93]$, $F_{35} \in [10, 36]$, $F_{36} \in [11, 37]$

TABLE 6: Selected words from Whissel's study [5].

	Activation	Evaluation	Angle		Activation	Evaluation	Angle
Accepting			0	Disgusted	5	3.2	181.3
Terrified	6.3	3.4	75.7	Joyful	5.4	6.1	323.4
Afraid	4.9	3.4	70.3	Delighted	4.2	6.4	318.6
Worried	3.9	2.9	126	Guilty	4	1.1	102.3
Angry	4.2	2.7	212	Bashful	2	2.7	74.7
Patient	3.3	3.8	39.7	Surprised	6.5	5.2	146.7
Sad	3.8	2.4	108.5	Eager	5	5.1	311

emotion I emerges as the union of vocabularies V_{A_1} and V_{A_2}, that is, $V_I = V_{A_1} \cup V_{A_2}$.

As already stated in Section 2, defining a vocabulary is not enough for modeling expressions; profiles should be created for this purpose. This poses a number of interesting issues in the case of different FAPs employed in the animation of individual profiles: in our approach, FAPs that are common in both emotions are retained during synthesis, while FAPs used in only one emotion are averaged with the respective neutral position. The same applies in the case of mutually exclusive FAPs: averaging of the intensities usually favors the most exaggerated of the emotions that are
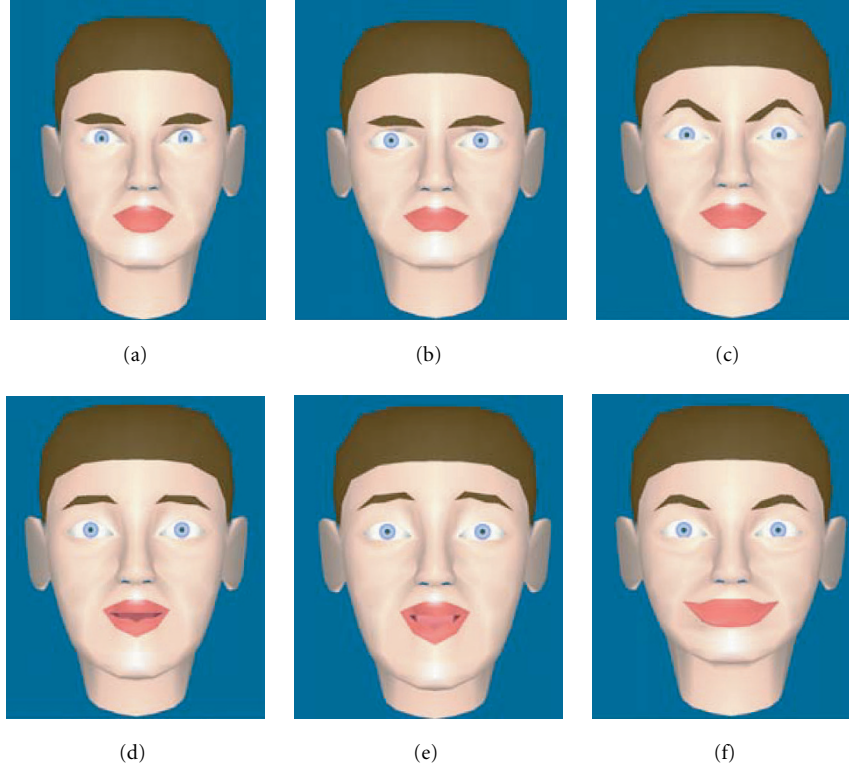
FIGURE 5: Examples of animated profiles (a), (b), and (c) Anger, (d), (e) Surprise, (f) Joy.
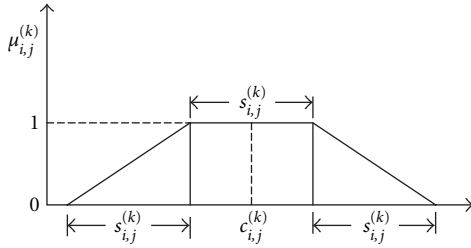

FIGURE 6: The form of membership functions.

combined, whereas FAPs with contradicting intensities are cancelled out. In practice, this approach works successfully, as shown in the actual results that follow. The combination of different, perhaps contradictory or exclusive, FAPs can be used to establish a distinct emotion categorization, similar to the semantic one, with respect to the common or neighboring FAPs that are used to synthesize and animate emotions.

Below, we describe the way to merge profiles of archetypal emotions and create profiles of intermediate ones.

Let $P_{A_1}^{(k)}$ be the kth profile of emotion A_1 and $P_{A_2}^{(l)}$ the lth profile of emotion A_2, then the following rules are applied so as to create a profile $P_I^{(m)}$ for the intermediate emotion I.

Rule 1. $P_I^{(m)}$ includes FAPs that are involved either in $P_{A_1}^{(k)}$ or $P_{A_2}^{(l)}$.

Rule 2. If the F_j is a FAP involved in both $P_{A_1}^{(k)}$ and $P_{A_2}^{(l)}$ with the same sign (direction of movement), then the range of variation $X_{I,j}^{(k)}$ is computed as a weighted translation of $X_{A_1,j}^{(k)}$ and $X_{A_2,j}^{(l)}$ (where $X_{A_1,j}^{(k)}$ and $X_{A_2,j}^{(l)}$ are the ranges of variation of FAP F_j involved in $P_{A_1}^{(k)}$ and $P_{A_2}^{(l)}$, resp.) in the following way:

(i) we compute the translated range of variations

$$t(X_{A_1,j}^{(k)}) = \frac{a_I}{a_{A_1}} X_{A_1,j}^{(k)}, \qquad t(X_{A_2,j}^{(k)}) = \frac{a_I}{a_{A_2}} X_{A_2,j}^{(k)} \qquad (4)$$

of $X_{A_1,j}^{(k)}$ and $X_{A_2,j}^{(l)}$,

(ii) we compute the center and length $c_{A_1,j}^{(k)}, s_{A_1,j}^{(k)}$ of $t(X_{A_1,j}^{(k)})$ and $c_{A_2,j}^{(k)}, s_{A_2,j}^{(k)}$ of $t(X_{A_2,j}^{(k)})$,

(iii) the length of $X_{I,j}^{(k)}$ is

$$s_{I,j}^{(m)} = \frac{\omega_I - \omega_{A_1}}{\omega_{A_2} - \omega_{A_1}} s_{A_1,j}^{(k)} + \frac{\omega_{A_2} - \omega_I}{\omega_{A_2} - \omega_{A_1}} s_{A_2,j}^{(l)} \qquad (5)$$

and its midpoint is

$$c_{I,j}^{(m)} = \frac{\omega_I - \omega_{A_1}}{\omega_{A_2} - \omega_{A_1}} c_{A_1,j}^{(k)} + \frac{\omega_{A_2} - \omega_I}{\omega_{A_2} - \omega_{A_1}} c_{A_2,j}^{(l)}. \qquad (6)$$

Rule 3. If the F_j is involved in both $P_{A_1}^{(k)}$ and $P_{A_2}^{(l)}$ but with contradictory sign (opposite direction of movement), then

the range of variation $X_{I,j}^{(k)}$ is computed by

$$X_{I,j}^{(m)} = (a_I/a_{A_1})X_{A_1,j}^{(k)} \cap (a_I/a_{A_2})X_{A_2,j}^{(l)}. \tag{7}$$

In case where $X_{I,j}^{(k)}$ is eliminated (which is the most possible situation), F_j is excluded from the profile.

Rule 4. If the F_j is involved only in one of $P_{A_1}^{(k)}$ and $P_{A_2}^{(l)}$, then the range of variation $X_{I,j}^{(k)}$ will be averaged with the corresponding of the neutral face position, that is, $X_{I,j}^{(m)} = (a_I/(2*a_{A_1}))X_{A_1,j}^{(k)}$ or $X_{I,j}^{(m)} = (a_I/(2*a_{A_2}))X_{A_2,j}^{(l)}$.

Rule 5. a_{A_1}, a_{A_2}, and a_I are the values of the *activation* parameter for emotion words A_1, A_2, and I, obtained from Whissel's study [5].

Rule 6. ω_{A_1}, ω_{A_2}, and ω_I, $\omega_{A_1} < \omega_I < \omega_{A_2}$ are the *angular* parameters for emotion words A_1, A_2, and I, obtained from Plutchik's study [6].

It should be noted that the profiles, created using the above rules, have to be animated for testing and correction purposes; the final profiles are those that present an acceptable visual similarity with the requested real emotion.

5.   THE EMOTION ANALYSIS SUBSYSTEM

In this section, we present a way of utilizing emotion modeling through profiles, for emotion understanding purposes. By doing this, we show that modeling emotions serves both synthesis as well as analysis purposes.

Consider as input to the emotion analysis sub-system a 15-element length feature vector $\bar{f}$ that corresponds to the 15 features f_i shown in Table 3. The particular values of $\bar{f}$ can be rendered to FAP values as shown in the same table (see also Section 3.1) resulting in an input vector $\bar{G}$. The elements of $\bar{G}$ express the observed values of the corresponding involved FAPs; for example G_1 refers to the value of F_{37}.

Let $X_{i,j}^{(k)}$ be the range of variation of FAP F_j involved in the kth profile $P_i^{(k)}$ of emotion i. If $c_{i,j}^{(k)}$ and $s_{i,j}^{(k)}$ are the middle point and length of interval $X_{i,j}^{(k)}$, respectively, then we describe a fuzzy class $A_{i,j}^{(k)}$ for F_j, using the membership function $\mu_{i,j}^{(k)}$ shown in Figure 6. Also let $\Delta_{i,j}^{(k)}$ be the set of classes $A_{i,j}^{(k)}$ that correspond to profile $P_i^{(k)}$; the beliefs $p_i^{(k)}$ and b_i that an observed, through the vector $\bar{G}$, facial state corresponds to profile $P_i^{(k)}$ and emotion i, respectively, are computed through the following equations:

$$p_i^{(k)} = \prod_{A_{i,j}^{(k)} \in \Delta_{i,j}^{(k)}} r_{i,j}^{(k)}, \tag{8}$$

$$b_i = \max_k (p_i^{(k)}), \tag{9}$$

where

$$r_{i,j}^{(k)} = \max \{g_i \cap A_{i,j}^{(k)}\} \tag{10}$$

expresses the *relevance* $r_{i,j}^{(k)}$ of the ith element of the input

feature vector with respect to class $A_{i,j}^{(k)}$. Actually $\bar{g} = A'(\bar{G}) = \{g_1, g_2, \ldots\}$ is the fuzzified input vector resulting from a *singleton* fuzzification procedure [22].

If a final decision about what is the observed emotion has to be made, then the following equation is used:

$$q = \arg\max_i b_i. \tag{11}$$

It is observed through (8) that the various emotion profiles correspond to the fuzzy intersection of several sets and are implemented through a *t-norm* of the form $t(a,b) = a \cdot b$. Similarly, the belief that an observed feature vector corresponds to a particular emotion results from a fuzzy union of several sets (see (9)) through an *s-norm* which is implemented as $u(a,b) = \max(a,b)$.

It should be noted that in the previously described emotion analysis system, no hypothesis has been made about the number of recognizable emotions; this number is limited only from the available modeled profiles. Thus, the system can be used for analyzing either as few as the archetypal emotions or much more, using the methodology described in Section 4 to create profiles for intermediate emotions.

6.   EXPERIMENTAL RESULTS

In this section, we show the efficiency of the proposed scheme on synthesizing archetypal and intermediate emotions according to the methodology described in the previous sections. Animated profiles were created using the face model developed in the context of the European project ACTS Mo-MuSys [21], as well as the 3D model of the software package *Poser*, edition 4 of Curious Labs Company. This model has separate parts for each moving face part. The Poser model interacts with the controls in Poser and has joints that move realistically, as in real person. Poser mirrors real face movements by adding joint parameters to each face part. This allows us to manipulate the figure based on those parameters. We can control the eyes, the eyebrows, and the mouth of the model by filling the appropriate parameters; to do this a mapping from FAPs to Poser parameters is necessary. We did this mapping mainly experimentally; the relationship between FAPs and Poser parameters is more or less straightforward.

The first set of experiments shows synthesized archetypal expressions (see Figure 7) created by using the Poser software package. The 3D nature of the face model renders the underlying emotions in a more natural way than the MPEG-4 compatible face model (compare Figures 5e and 5f for the emotions *surprise* and *joy* with Figures 7f and 7c, respectively). However in both cases the synthesized examples are rather convincing.

The second set of experiments shows particular examples in creating intermediate expressions based on our proposed method. Figures 8 and 10 were rendered with *Curious Labs Poser*, while Figures 9 and 11 are screenshots from face model developed in the context of the European project ACTS Mo-MuSys [21]. In the first case, users have control over the deformation of areas of the polygonal model and not just specific vertices. As a result, the rendered images simulate
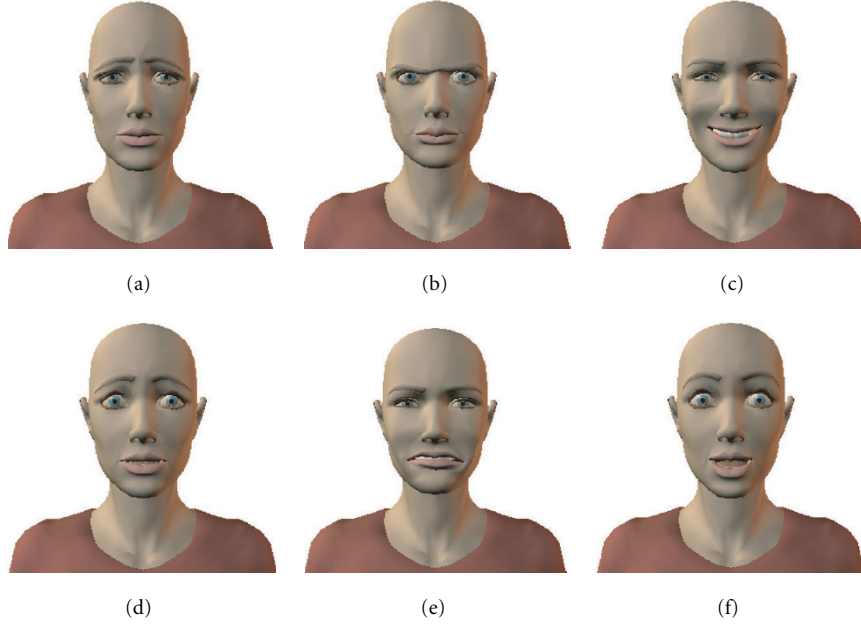
FIGURE 7: Synthesized archetypal expressions created using the 3D model of the POSER software package: (a) *sadness*, (b) *anger*, (c) *joy*, (d) *fear*, (e) *disgust*, and (f) *surprise*.

expressions more effectively, since the FAT mechanism can approximate the effect of muscle deformation, which accounts for the shape of the face during expressions. In the case of Figures 9 and 11 the decoder only utilises the FAPs supplied and thus, the final result depends on the predefined mapping between the animation parameters and the low polygon model.

6.1. Creating profiles for emotions belonging to the same universal category

In this section, we illustrate the proposed methodology for creating profiles for emotions that belong to the same universal category as an archetypal one. Emotion terms *afraid*, *terrified*, and *worried* are considered to belong to the emotion category *fear* [11] whose modeling base is the term *afraid*. In Table 7 are shown the produced profiles for the terms *terrified* and *worried* emerged by the one of the profiles of *afraid* (in particular $P_F^{(8)}$). The range of variation $X_{T,j}^{(8)}$ of FAP F_j belonging to the eighth profile of emotion term *terrified* is computed by the equation $X_{T,j}^{(8)} = (6.3/4.9)X_{F,j}^{(8)}$, where $X_{F,j}^{(8)}$ is the range of variation of FAP F_j belonging to the eighth profile of emotion term *afraid*. Similarly, $X_{W,j}^{(8)} = (3.9/4.9)X_{F,j}^{(8)}$ is the range of variation of FAP F_j belonging to the eighth profile of emotion term *worried*. Figures 8 and 9 show the animated profiles for emotion terms *afraid*, *terrified*, and *worried*, respectively. The FAP values that we used are the median ones of the corresponding ranges of variation.

6.2. Creating profiles for emotions lying between the archetypal ones

In this section, we describe the method of creating a profile for the emotion *guilt*. According to the Plutchik's *angular*

measure (see Table 6), emotion term *guilty* (angular measure 102.3 degrees) lies between the archetypal emotion terms *afraid* (angular measure 70.3 degrees) and *sad* (angular measure 108.5 degrees), being closer to the latter. According to Section 4.2 the vocabulary V_G of emotion *guilt* emerges as the union of vocabularies V_F and V_S, that is, $V_G = V_F \cup V_S$, where V_F and V_S are the vocabularies corresponding to emotions *fear* and *sad*, respectively. In Table 8 it is shown the produced profile for the term *guilty* emerged by the one of the profiles of *afraid* (in particular $P_F^{(8)}$) and *sad* ($P_S^{(0)}$). FAPs F_3, F_5, F_{33}, F_{34}, F_{35}, and F_{36} are included only in the $P_F^{(8)}$ and therefore the corresponding ranges of variation in the emerging *guilty* profile $P_G^{(m)}$ (*m*th *guilty* profile) are computed by averaging the ranges of variation of $P_F^{(8)}$ with the neutral face, according to Rule 4 in Section 4.2; for example $X_{G,3}^{(m)} = (4/2*4.9)X_{F,4}^{(8)}$. FAPs F_{19}, F_{20}, F_{21}, F_{22}, F_{31}, F_{32} are included in both $P_F^{(8)}$ and $P_S^{(0)}$, with the same direction of movement, thus Rule 2 in Section 4.2 is followed. For example the range of variation $X_{G,19}^{(m)}$ for FAP F_{29} term is computed as follows:

$$t(X_{F,19}^{(8)}) = \frac{4}{4.9}X_{F,19}^{(8)} \implies [-510, -460],$$

$$c_{F,19}^{(8)} = -485, \qquad s_{F,19}^{(8)} = 50,$$

$$t(X_{S,19}^{(0)}) = \frac{4}{3.9}X_{S,19}^{(0)} \implies [-270, -42],$$

$$c_{S,19}^{(0)} = -156, \qquad s_{S,19}^{(9)} = 228, \tag{12}$$

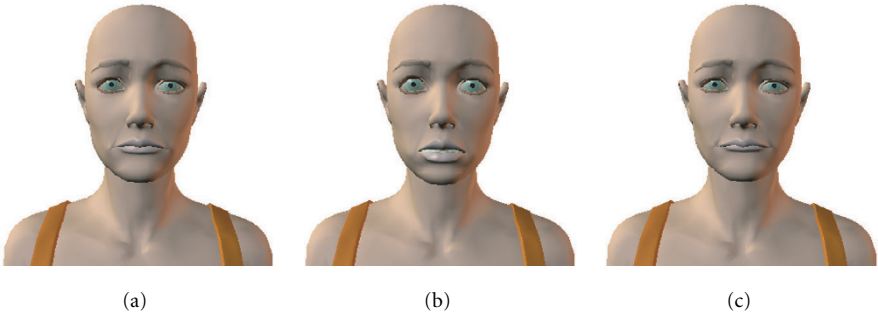(a)   (b)   (c)

FIGURE 8: Poser face model: animated profiles for emotion terms (a) *afraid*, (b) *terrified*, and (c) *worried*.

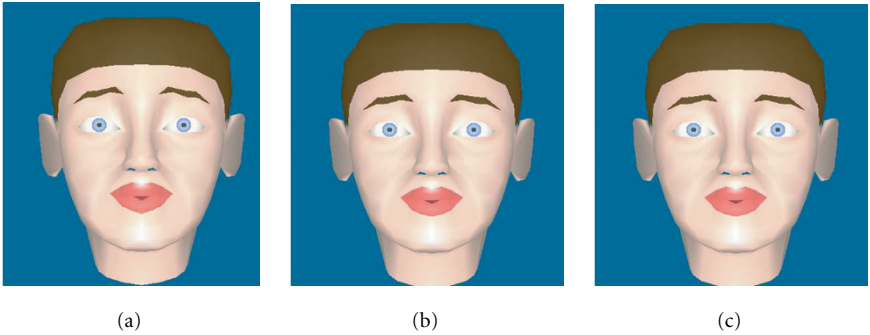(a)   (b)   (c)

FIGURE 9: MPEG-4 face model: animated profiles for emotion terms (a) *afraid*, (b) *terrified*, and (c) *worried*.
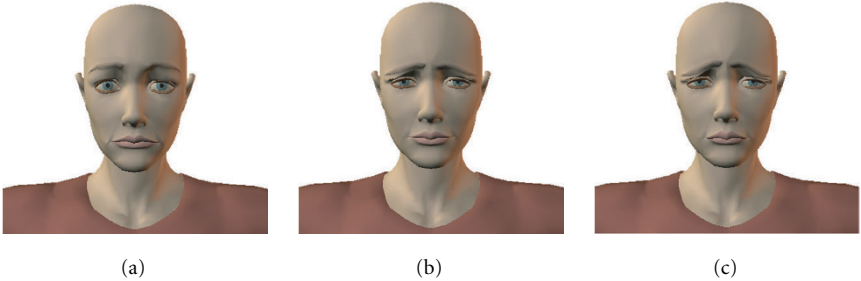

(a)   (b)   (c)

FIGURE 10: Poser face model: animated profiles for emotion terms (a) *afraid*, (b) *guilty*, and (c) *sad*.

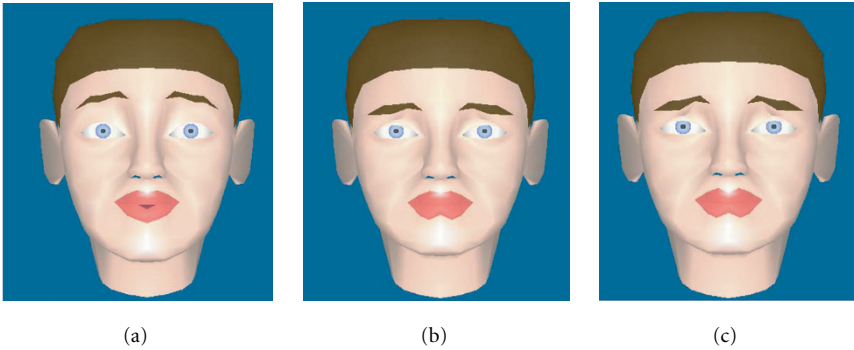(a)   (b)   (c)

FIGURE 11: MPEG-4 face model: animated profiles for emotion terms (a) *afraid*, (b) *guilty*, and (c) *sad*.

TABLE 7: Created profiles for the emotions *terror* and *worry*.

Emotion term	Activation	Profile
Afraid	4.9	$F_3 \in [400, 560]$, $F_5 \in [-240, -160]$, $F_{19} \in [-630, -570]$, $F_{20} \in [-630, -570]$, $F_{21} \in [-630, -570]$, $F_{22} \in [-630, -570]$, $F_{31} \in [260, 340]$, $F_{32} \in [260, 340]$, $F_{33} \in [160, 240]$, $F_{34} \in [160, 240]$, $F_{35} \in [60, 140]$, $F_{36} \in [60, 140]$
Terrified	6.3	$F_3 \in [520, 730]$, $F_5 \in [-310, -210]$, $F_{19} \in [-820, -740]$, $F_{20} \in [-820, -740]$, $F_{21} \in [-820, -740]$, $F_{22} \in [-820, -740]$, $F_{31} \in [340, 440]$, $F_{32} \in [340, 440]$, $F_{33} \in [210, 310]$, $F_{34} \in [210, 310]$, $F_{35} \in [80, 180]$, $F_{36} \in [80, 180]$
Worried	3.9	$F_3 \in [320, 450]$, $F_5 \in [-190, -130]$, $F_{19} \in [-500, -450]$, $F_{20} \in [-500, -450]$, $F_{21} \in [-500, -450]$, $F_{22} \in [-500, -450]$, $F_{31} \in [210, 270]$, $F_{32} \in [210, 270]$, $F_{33} \in [130, 190]$, $F_{34} \in [130, 190]$, $F_{35} \in [50, 110]$, $F_{36} \in [50, 110]$

TABLE 8: Created profile for the emotion *guilt*.

Emotion term	Activation	Angular measure	Profile
Afraid	4.9	70.3	$F_3 \in [400, 560]$, $F_5 \in [-240, -160]$, $F_{19} \in [-630, -570]$, $F_{20} \in [-630, -570]$, $F_{21}[-630, -570]$, $F_{22} \in [-630, -570]$, $F_{31} \in [260, 340]$, $F_{32} \in [260, 340]$, $F_{33} \in [160, 240]$, $F_{34} \in [160, 240]$, $F_{35} \in [60, 140]$, $F_{36} \in [60, 140]$
Guilty	4	102.3	$F_3 \in [160, 230]$, $F_5 \in [-100, -65]$, $F_{19} \in [-110, -310]$, $F_{20} \in [-120, -315]$, $F_{21} \in [-110, -310]$, $F_{22} \in [-120, -315]$, $F_{31} \in [61, 167]$, $F_{32} \in [57, 160]$, $F_{33} \in [65, 100]$, $F_{34} \in [65, 100]$, $F_{35} \in [25, 60]$, $F_{36} \in [25, 60]$
Sad	3.9	108.5	$F_{19} \in [-265, -41]$, $F_{20} \in [-270, -52]$, $F_{21} \in [-265, -41]$, $F_{22} \in [-270, -52]$, $F_{31} \in [30, 140]$, $F_{32} \in [26, 134]$

since

$$\omega_F = 70.3°, \qquad \omega_S = 108.5°, \qquad \omega_G = 102.3°,$$

$$c_{G,19}^{(m)} = \frac{102.3 - 70.3}{108.5 - 70.3} * (-156)$$
$$+ \frac{108.5 - 102.3}{108.5 - 70.3} * (-485) = -209, \tag{13}$$
$$s_{G,19}^{(m)} = \frac{102.3 - 70.3}{108.5 - 70.3} * 228$$
$$+ \frac{108.5 - 102.3}{108.5 - 70.3} * 50 = 199,$$

and $X_{G,19}^{(m)}$ corresponds to the range $[-110, -310]$.

7. CONCLUSION, DISCUSSION, AND FURTHER WORK

In this work, we have 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.

(i) No hypothesis about what the expression analysis system is (see Figure 1), should be made; it is enough to provide either the name of the conveyed emotion, or just the movement of a predefined set of FPs. 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.

(ii) It is extensible with respect to completing (or modifying) the proposed vocabulary of FAPs for the archetypal expressions

(iii) The range of variation of FAPs that involved in the archetypal expression profiles can be modified. Note however that this modification affects the profiles that created for intermediate expressions.

(iv) It is extensible with respect to the number of intermediate expressions that can be modeled.

Exploitation of the results obtained by psychological studies related with emotion recognition from computer scientists is possible although not straightforward. We have shown that terms like the *emotion wheel* and *activation* are suitable for extending the emotions that can be visually modeled.

The main focus of the paper is on synthesizing MPEG-4 compliant facial expressions; realistic generic animation is

another interesting issue which would indeed require specific FATs. This constitutes a topic for further developments.

The results presented indicate that the use of FATs, while not essential, enhances the obtained results. However, in cases of low bitrate applications where speed and responsiveness are more important than visual fidelity, the FAT functionality may be omitted, since it imposes considerable overhead on the data stream. Samples of the emotional animation, including values and models, used in this paper can be found at http://www.image.ntua.gr/mpeg4.

REFERENCES

[1] EC TMR Project PHYSTA Report, *Development of feature representation from facial signals and speech*, January 1999.

[2] P. Ekman, "Facial expression and emotion," *American Psychologist*, vol. 48, no. 4, pp. 384–392, 1993.

[3] P. Ekman and W. V. Friesen, *Pictures of Facial Affect*, Consulting Psychologists Press, Palo Alto, Calif, USA, 1978.

[4] C. Izard, L. Dougherty, and E. A. Hembree, "A system for identifying affect expressions by holistic judgements," Tech. Rep., University of Delaware, Newark, Del, USA, 1983.

[5] C. M. Whissel, "The dictionary of affect in language," in *Emotion: Theory, Research and Experience*, R. Plutchnik and H. Kellerman, Eds., vol. 4 of *The Measurement of Emotions*, pp. 113–131, Academic Press, New York, USA, 1989.

[6] R. Plutchik, *Emotion: A Psychoevolutionary Synthesis*, Harper and Row, New York, USA, 1980.

[7] R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, et al., "Emotion recognition in human-computer interaction," *IEEE Signal Processing Magazine*, vol. 18, no. 1, pp. 32–80, 2001.

[8] P. Ekman and W. Friesen, *The Facial Action Coding System*, Consulting Psychologists Press, San Francisco, Calif, USA, 1978.

[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," in *Proc. European Signal Processing Conference*, Tampere, Finland, September 2000.

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

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

[12] I. Essa and A. Pentland, "Coding, analysis, interpretation, and recognition of facial expressions," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 757–763, 1997.

[13] EC TMR Project PHYSTA Report, *Review of existing techniques for human emotion understanding and applications in human-computer interaction*, October 1998.

[14] P. Chellapa, C. Wilson, and S. Sirohey, "Human and machine recognition of faces: A survey," *Proceedings of the IEEE*, vol. 83, no. 5, pp. 705–740, 1995.

[15] H. A. Rowley, S. Baluja, and T. Kanade, "Neural network-based face detection," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 20, no. 1, pp. 23–28, 1998.

[16] K. Karpouzis, N. Tsapatsoulis, and S. Kollias, "Moving to continuous facial expression space using the MPEG-4 facial definition parameter (FDP) set," in *Human Vision and Electronic Imaging V*, vol. 3959 of *Proceedings of SPIE*, San Jose, Calif, USA, January 2000.

[17] M. Tekalp and J. Ostermann, "Face and 2-D mesh animation in MPEG-4," *Image Communication Journal*, vol. 15, no. 4-5, pp. 387–421, 2000, Tutorial Issue On The MPEG-4 Standard.

[18] K.-M. Lam and H. Yan, "An analytic-to-holistic approach for face recognition based on a single frontal view," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 20, no. 7, pp. 673–686, 1998.

[19] J. Ostermann and E. Haratsch, "An animation definition interface: Rapid design of MPEG-4 compliant animated faces and bodies," in *International Workshop on Synthetic-Natural Hybrid Coding and Three Dimensional Imaging*, pp. 216–219, Greece, September 1997.

[20] F. Lavagetto and R. Pockaj, "The facial animation engine: Toward a high-level interface for the design of MPEG-4 compliant animated faces," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 9, no. 2, pp. 277–289, 1999.

[21] G. Abrantes and F. Pereira, "MPEG-4 facial animation technology: survey, implementation and results," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 9, no. 2, pp. 290–305, 1999.

[22] G. J. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic, Theory and Applications*, Prentice-Hall, Upper Saddle River, NJ, USA, 1995.

Amaryllis Raouzaiou was born in Athens, Greece in 1977. She graduated from the Department of Electrical and Computer Engineering, the National Technical University of Athens in 2000 and she is currently pursuing the Ph.D. degree at the same university. Her current research interests lie in the areas of synthetic-natural hybrid video coding, human-computer interaction, machine vision, and neural networks. She is a member of the Technical Chamber of Greece. She is with the team of IST project ERMIS (Emotionally Rich Man-Machine Interaction Systems, IST-2000-29319).

Nicolas Tsapatsoulis was born in Limassol, Cyprus in 1969. He graduated from the Department of Electrical and Computer Engineering, the National Technical University of Athens in 1994 and received his Ph.D. degree in 2000 from the same university. His current research interests lie in the areas of human-computer interaction, machine vision, image and video processing, neural networks, and biomedical engineering. He is a member of the Technical Chambers of Greece and Cyprus and a member of IEEE Signal Processing and Computer societies. Dr. Tsapatsoulis has published nine papers in international journals and more than 20 in proceedings of international conferences. He served as Technical Program Co-Chair for the VLBV '01 workshop. He is a reviewer of the *IEEE Transactions on Neural Networks* and *IEEE Transactions on Circuits and Systems for Video Technology* journals. Since 1995 he has participated in ten research projects at Greek and European level.
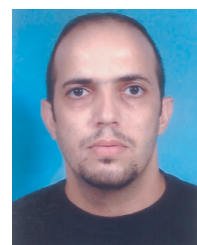
Kostas Karpouzis was born in Athens, Greece in 1972. He graduated from the Department of Electrical and Computer Engineering, the National Technical University of Athens (NTUA) in 1998 and received his Ph.D. degree in 2001 from the same university. His current research interests lie in the areas of human-computer interaction, image and video processing, 3D computer animation, and virtual reality. He is a member of the Technical Chamber of Greece and a member of ACM SIGGRAPH and SIGCHI societies. Dr. Karpouzis has published five papers in international journals and more than 15 in proceedings of international conferences. He is a member of the technical committee of the International Conference on Image Processing (ICIP). Since 1995 he has participated in seven research projects at Greek and European level.

Stefanos Kollias was born in Athens, Greece in 1956. He obtained his Diploma from National Technical University of Athens in 1979, his M.Sc. in Communication Engineering in 1980 from UMIST (University of Manchester, Institute of Science and Technology) in England and his Ph.D. in Signal Processing from the Computer Science Division of NTUA. He is with the Electrical Engineering Department of NTUA since 1986 where he serves now as a Professor. Since 1990 he is Director of the Image, Video and Multimedia Systems Laboratory of NTUA. He has published more than 120 papers in the above fields, 50 of which in international journals. He has been a member of the Technical or Advisory Committee or invited speaker in 40 international conferences. He is a reviewer of ten IEEE Transactions and of ten other journals. Ten graduate students have completed their Doctorate under his supervision, while other ten are currently performing their Ph.D. thesis. He and his team have been participating in 38 European and National projects.