An Intermediate Expressions' Generator System in the MPEG-4 Framework

Amaryllis Raouzaiou, Evaggelos Spyrou, Kostas Karpouzis, and Stefanos Kollias

Image, Video and Multimedia Systems Laboratory,
School of Electrical and Computer Engineering,
National Technical University of Athens,
Athens, Greece
{araouz, espyrou, kkarpou}@image.ntua.gr,
stefanos@cs.ntua.gr

Abstract. A lifelike human face can enhance interactive applications by providing straightforward feedback to and from the users and stimulating emotional responses from them. An expressive, realistic avatar should not "express himself" in the narrow confines of the six archetypal expressions. In this paper, we present a system which generates intermediate expression profiles (set of FAPs) combining profiles of the six archetypal expressions, by utilizing concepts included in the MPEG-4 standard.

1 Introduction

Research in facial expression analysis and synthesis has mainly concentrated on 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 non-archetypal emotions. This trend may be due to the great influence of the works of Ekman [4] and Friesen 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 a broader variety of emotions. An extensive survey on emotion analysis can be found in [9]. An expressive, realistic avatar should not "express himself" in the narrow confines of the six archetypal expressions. Intermediate expressions ought to be a part of synthesizable expressions in every possible application (online gaming, e-commerce, interactive TV etc).

Moreover, the MPEG-4 indicates an alternative way of modeling facial expressions and the underlying emotions, which is strongly influenced from neurophysiological and psychological studies (FAPs). The adoption of token-based animation in the MPEG-4 framework [6] benefits the definition of emotional states, since the extraction of simple, symbolic parameters is more appropriate to synthesize, as well as analyze facial expression and hand gestures.

In this paper we describe a system which has as output profiles of intermediate expressions, i.e. group of FAPs accompanied with FAP intensities - the actual ranges of variation, which if animated create the requested expression, taking into account

L. Atzori et al. (Eds.): VLBV 2005, LNCS 3893, pp. 129 – 136, 2006.

results of Whissel's study [9]. These results can then be applied to avatars, so as to convey the communicated messages more vividly than plain textual information or simply to make interaction more lifelike.

2 MPEG-4 and Emotion Representation

In the framework of MPEG-4 standard [8], parameters have been specified for Face and Body Animation (FBA) by defining specific Face and Body nodes in the scene graph. The goal of FBA definition is the animation of both realistic and cartoonist characters. Thus, MPEG-4 has defined a large set of parameters and the user can select subsets of these parameters according to the application, especially for the body, for which the animation is much more complex. The FBA part can be also combined with multimodal input (e.g. linguistic and paralinguistic speech analysis).

As far as facial animation is concerned, MPEG-4 specifies 84 feature points on the neutral face, which provide spatial reference for FAPs definition. The FAP set contains two high-level parameters, visemes and expressions. In particular, the Facial Definition Parameter (FDP) and the Facial Animation Parameter (FAP) set were designed in the MPEG-4 framework to allow the definition of a facial shape and texture, eliminating the need for specifying the topology of the underlying geometry, through FDPs, and the animation of faces reproducing expressions, emotions and speech pronunciation, through FAPs. By monitoring facial gestures corresponding to FDP and/or FAP movements over time, it is possible to derive cues about user's expressions and emotions. Various results have been presented regarding classification of archetypal expressions of faces, mainly based on features or points mainly extracted from the mouth and eyes areas of the faces. These results indicate that facial expressions, possibly combined with gestures and speech, when the latter is available, provide cues that can be used to perceive a person's emotional state.

The obvious goal for emotion analysis applications is to assign category labels that identify emotional states. However, labels as such are very poor descriptions, especially since humans use a daunting number of labels to describe emotion.

Psychologists have examined a broader set of emotions [1], but very few of the studies provide results which can be exploited in computer graphics and machine vision fields. One of these studies, carried out by Whissel [9], suggests that emotions are points in a space spanning a relatively small number of dimensions, which seem to occupy two axes: *activation* and *evaluation* (*Figure 1*).

- *Valence* (Evaluation level): the clearest common element of emotional states is that the person is materially influenced by feelings that are "valenced", i.e. they are centrally concerned with positive or negative evaluations of people or things or events. The link between emotion and valencing is widely agreed (horizontal axis).
- *Activation* level: research has recognized that emotional states involve dispositions to act in certain ways. A basic way of reflecting that theme turns out to be surprisingly useful. States are simply rated in terms of the associated activation level, i.e. the strength of the person's disposition to take some action rather than none (vertical axis).

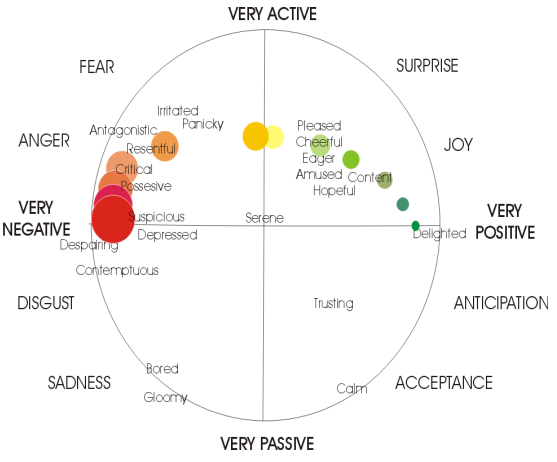


Fig. 1. The Activation – emotion space

A surprising amount of emotional discourse can be captured in terms of activation-emotion space. Perceived full-blown emotions are not evenly distributed in activation-emotion space; instead they tend to form a roughly circular pattern. In this framework, identifying the center as a natural origin has several implications. Emotional strength can be measured as the distance from the origin to a given point in activation-evaluation space. An interesting implication is that strong emotions are more sharply distinct from each other than weaker emotions with the same emotional orientation.

3 Intermediate Expressions

The limited number of studies, carried out by computer scientists and engineers [1], dealing with emotions other than the archetypal ones, lead us to search in other subject/discipline bibliographies. Psychologists examined a broader set of emotions [1], but very few of the corresponding studies provide exploitable results to computer graphics and machine vision fields, e.g. Whissel's study suggests that emotions are points in a space (*Figure 1*) spanning a relatively small number of dimensions, which in a first approximation, seem to occupy two axes: *activation* and *evaluation*, as shown in Table 1. *Activation* is the degree of arousal associated with the term, with terms like *surprised* (around 3) representing high activation, and *sad* (around -2) representing low *activation*. *Evaluation* is the degree of pleasantness associated with the term, with *angry* (at -3.0) representing the negative extreme and *delighted* (at 2.9) representing the positive extreme. 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.

Table 1. Selected Words from Whissel's Study

	Activation(a)	*Evaluation(e)*
Terrified	2.8	-2.0
Afraid	1.4	-2.0
Worried	0.4	-2.1
Angry	1.3	-3.0
Surprised	3	2.5
Sad	-2.0	-1.7
Depressed	-0.3	-2.5
Suspicious	0.2	-2.8
Delighted	0.7	2.9

The synthesis of intermediate expressions is based on the profiles of the six archetypal expressions [7].

As a general rule, one can define six general categories, each one characterized by an archetypal emotion. 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* [7]; these two emotions can be synthesized by reducing or increasing the intensities of the employed FAPs, respectively. In the case of expression profiles, this affects the range of variation of the corresponding FAPs which is appropriately translated.

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, one should first define the FAPs which are involved in the particular emotion.

One is able to synthesize intermediate emotions by combining the FAPs employed for the representation of universal ones. In our approach, FAPs that are common in both emotions are retained during synthesis, while emotions used in only one emotion are averaged with the respective neutral position. In the case of mutually exclusive FAPs, averaging of intensities usually favors the most exaggerated of the emotions that are combined, whereas FAPs with contradicting intensities are cancelled out.

Below we describe the rules used by our system to merge profiles of archetypal emotions and create profiles of intermediate ones:

Let $P_{A_1}^{(k)}$ be the *k-th* profile of emotion A_1 and $P_{A_2}^{(l)}$ the *l-th* profile of emotion A_2. Let $X_{A_1,j}^{(k)}$ and $X_{A_2,j}^{(l)}$ be the ranges of variation of FAP F_j involved in $P_{A_1}^{(k)}$ and $P_{A_2}^{(l)}$ respectively. Additionally, $\omega = \tan^{-1}\left(\frac{a}{e}\right)$, ω_{A_1}, ω_{A_2} and ω_I, $\omega_{A_1} < \omega_I < \omega_{A_2}$, a_{A_1}, a_{A_2} and a_I are the values of the *activation* parameter and e_{A_1}, e_{A_2} and e_I the values of the *evaluation* parameter for emotion words A_1, A_2 and I respectively, obtained from Whissel's study [9]. The following rules are applied in order 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 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 $X_{A_2,j}^{(l)}$ in the following way: (i) the translated ranges of variations $t(X_{A_1,j}^{(k)}) = \frac{a_I}{a_{A_1}} X_{A_1,j}^{(k)}$ and $t(X_{A_2,j}^{(k)}) = \frac{a_I}{a_{A_2}} X_{A_2,j}^{(k)}$ of $X_{A_1,j}^{(k)}$ and $X_{A_2,j}^{(l)}$ are computed, (ii) the centers $c_{A_1,j}^{(k)}$ and $c_{A_2,j}^{(k)}$ of $t(X_{A_1,j}^{(k)})$ and $t(X_{A_2,j}^{(k)})$ are the same as those of $X_{A_1,j}^{(k)}$ and $X_{A_2,j}^{(l)}$, (iii) the lengths $s_{A_1,j}^{(k)}$ and $s_{A_2,j}^{(k)}$ of $t(X_{A_1,j}^{(k)})$ and $t(X_{A_2,j}^{(k)})$ are computed using the relation $s_{Ai,j}^{(k)} = \frac{1}{3} t(X_{A_1,j}^{(k)})$, (iv) the length of $X_{I,j}^{(k)}$ is

$$s_{I,j}^{(m)} = \frac{\omega_I - \omega_{A_1}}{\omega_{A_2} - \omega_{A_1}} s_{A_1,j}^{(k)} + \frac{\omega_{A_2} - \omega_I}{\omega_{A_2} - \omega_{A_1}} s_{A_2,j}^{(l)}$$ and its midpoint is

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

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 $X_{I,j}^{(m)} = \frac{a_I}{a_{A_1}} X_{A_1,j}^{(k)} \cap \frac{a_I}{a_{A_2}} X_{A_2,j}^{(l)}$. In case where $X_{I,j}^{(k)}$ is eliminated (which is the most possible situation) then 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, i.e., $X_{I,j}^{(m)} = \frac{a_I}{2 * a_{A_1}} X_{A_1,j}^{(k)}$ or

$$X_{I,j}^{(m)} = \frac{a_I}{2 * a_{A_2}} X_{A_2,j}^{(l)}$$

4 Intermediate Expressions' Generator System

The proposed system (*Figure 2*) has as input the user request, i.e. the parameters *a* and *e* of Whissel's wheel of the desired intermediate expression or only the term of the intermediate expression (e.g. *depressed*) and the system uses the corresponding stored *a* and *e* values.

The main part of the system applies the above-mentioned rules and calculates the group of FAPs of every intermediate expression profile, accompanied by the corresponding values. This output has the exact form of the input of an MPEG-4 decoder, such as **GretaPlayer [5]**.

The profiles of each archetypal expression – every archetypal expression has more than one profile- are stored in the system and have the form of an array containing the ranges of variation for every one of the 68 FAPs of the MPEG-4 standard. However, only a subset of them can be extracted by the analysis procedure and thus can be used to create the profiles of archetypal expressions. The output of our system is also an array containing the essential information for an MPEG-4 decoder: a) a binary mask with ones representing the used FAPs and b) a value for every FAP lying near the center of the derived range of variation, picked randomly from the Gaussian distribution.

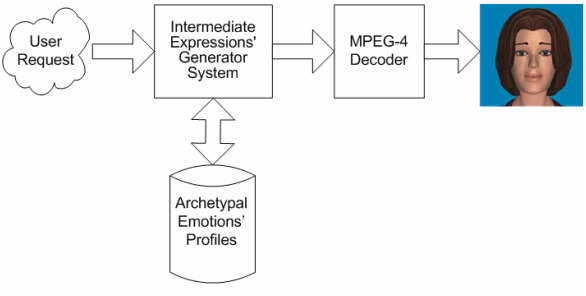


Fig. 2. Proposed System Architecture

5 Experimental Results

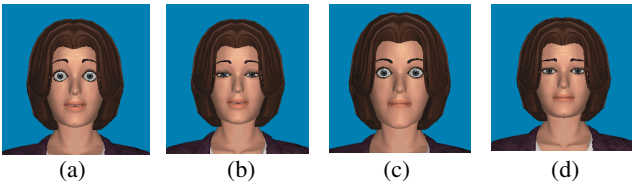
It should be noted that the profiles derived by the presented system, have to be animated for testing and correction purposes; the final profiles are those that are approved by experts, e.g. they present an acceptable visual similarity with the requested real emotion.

Table 2 presents the profiles of the basic emotions *fear* and *sadness* in the form they are stored in the system (see *Section IV*), omitting the columns that have zero values and the profile of the intermediate expression *depressed* in the same form.

Using the rules described above, *depression* (*Figures 3b, 3c*) and *guilt* (*Figure 5b*) is animated using *fear* (*Fig.3a, 5a*) and *sadness* (*Fig.3c, 5c*), *suspicious* (*Figure 4b*) using *anger* (*Fig. 4a*) and *disgust* (*Fig. 4c*). From *Figures 3b* and *3c, 3b* is approved and *3c* is rejected, after the judgment of an expert.

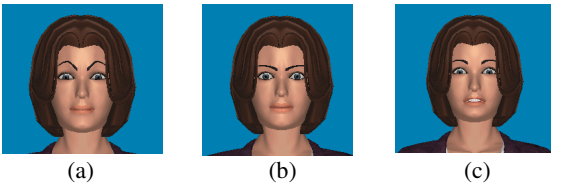
Table 2. Activation and evaluation measures used to create the profile for the emotion *depressed*

***Afraid* (1.4, -2.0):**											
F_3	**F_5**	**F_{19}**	**F_{20}**	**F_{21}**	**F_{22}**	**F_{31}**	**F_{32}**	**F_{33}**	**F_{34}**	**F_{35}**	**F_{36}**
400	-240	-630	-630	-630	-630	260	260	160	160	60	60
560	-160	-570	-570	-570	-570	340	340	240	240	140	140
***Depressed* (-0.3, -2.5):**											
F_3	**F_5**	**F_{19}**	**F_{20}**	**F_{21}**	**F_{22}**	**F_{31}**	**F_{32}**	**F_{33}**	**F_{34}**	**F_{35}**	**F_{36}**
160	-100	-110	-120	-110	-120	61	57	65	65	25	25
230	-65	-310	-315	-310	-315	167	160	100	100	60	60
***Sad* (-2.0, -1.7):**											
F_3	**F_5**	**F_{19}**	**F_{20}**	**F_{21}**	**F_{22}**	**F_{31}**	**F_{32}**	**F_{33}**	**F_{34}**	**F_{35}**	**F_{36}**
0	0	-265	-270	-265	-270	30	26	0	0	0	0
0	0	-41	-52	-41	-52	140	134	0	0	0	0



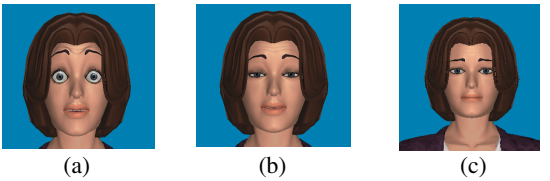
(a) (b) (c) (d)

Fig. 3. Profiles for (a) fear, (b-c) depressed (d) sadness



(a) (b) (c)

Fig. 4. Profiles for (a) anger, (b) suspicious (c) disgust



(a) (b) (c)

Fig. 5. Profiles for (a) fear, (b) guilt (c) sadness

6 Conclusions - Future Work

Expression synthesis is a great means of improving HCI applications, since it provides a powerful and universal means of expression and interaction. In this paper

we presented a system which provides realistic intermediate facial expression profiles, utilizing concepts included in established standards, such as MPEG-4, which are widely supported in modern computers and standalone devices and making human-computer interaction more lifelike.

In the future, more profiles of intermediate emotions can be created by the combination of two expressions, not necessarily archetypal ones, while the same system with slight alterations may be used to generate gesture profiles of intermediate emotions.

References

1. Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., Taylor, J.: Emotion Recognition in Human-Computer Interaction. IEEE Signal Processing Magazine (2001) 32-80
2. DeCarolis, B., Pelachaud, C., Poggi, I. and Steedman, M.: APML, A mark-up language for believable behavior generation, Life-Like Characters, Springer, 2004
3. EC TMR Project PHYSTA Report: Review of Existing Techniques for Human Emotion Understanding and Applications in Human-Computer Interaction (1998) http://www.image.ece.ntua.gr/physta/reports/emotionreview.htm
4. Ekman, P.: Facial expression and Emotion. Am. Psychologist, Vol. 48 (1993) 384-392
5. Hartmann, B., Mancini, M., Pelachaud, C.: Formational parameters and adaptive prototype instantiation for MPEG-4 compliant gesture synthesis, *Computer Animation 2002*, pp. 111. 3, 6, 7
6. Preda, M. and Prêteux, F.: Advanced animation framework for virtual characters within the MPEG-4 standard, *Proc. of the Intl. Conference on Image Processing*. Rochester, NY, 2002.
7. Raouzaiou, A., Tsapatsoulis, N., Karpouzis, K., Kollias, S.: Parameterized facial expression synthesis based on MPEG-4. EURASIP Journal on Applied Signal Processing, Vol. 2002, No. 10. Hindawi Publishing Corporation (2002) 1021-1038
8. Tekalp, M., Ostermann, J.: Face and 2-D mesh animation in MPEG-4. Image Communication Journal, Vol.15, Nos. 4-5 (2000) 387-421
9. Whissel, C.M.: The dictionary of affect in language. In: Plutchnik, R., Kellerman, H. (eds): Emotion: Theory, research and experience: Vol 4, The measurement of emotions. Academic Press, New York (1989)