Facial Expression Synthesis Using PAD Emotional Parameters for a Chinese Expressive Avatar

Shen Zhang[1,2], Zhiyong Wu[2], Helen M. Meng[2], and Lianhong Cai[1]

[1] Department of Computer Science and Technology
Tsinghua University, 100084 Beijing, China
[2] Department of Systems Engineering and Engineering Management
The Chinese University of Hong Kong, HKSAR, China
zhangshen05@mails.tsinghua.edu.cn, john.zy.wu@gmail.com
hmmeng@se.cuhk.edu.hk, clh-dcs@tsinghua.edu.cn

Abstract. Facial expression plays an important role in face to face communication in that it conveys nonverbal information and emotional intent beyond speech. In this paper, an approach for facial expression synthesis with an expressive Chinese talking avatar is proposed, where a layered parametric framework is designed to synthesize intermediate facial expressions using PAD emotional parameters [5], which describe the human emotional state with three nearly orthogonal dimensions. Partial Expression Parameter (PEP) is proposed to depict the facial expression movements in specific face regions, which act as the mid-level expression parameters between the low-level Facial Animation Parameters (FAPs) [11] and the high-level PAD emotional parameters. A pseudo facial expression database is established by cloning the real human expression to avatar and the corresponding emotion states for each expression is annotated using PAD score. An emotion-expression mapping model is trained on the database to map the emotion state (PAD) into facial expression configuration (PEP). Perceptual evaluation shows the input PAD value is consistent with that of human perception on synthetic expression, which supports the effectiveness of our approach.

Keywords: Facial expression, PAD emotional model, partial expression parameter, talking avatar.

1 Introduction

Facial expression is one of the most common and effective way for human to express emotions, and it plays a crucial role in face to face communication in that it can convey non-verbal information and intended emotion beyond speech. For expressive visual speech synthesis [1], the facial expression together with head movement is referred as visual prosody [2]. Our long-term plan is to create a Chinese expressive talking avatar for text-to-audio-visual-speech (TTAVS) synthesis [1], which leads to two questions: First, how should the emotional state be described? Second, what is the relationship between inner emotion and facial expression? Previous research on facial expression [1,3,8] are usually limited to six basic emotion categories (i.e. happy, sad, surprise, angry, disgust and fear). However, according to Mehrabian's theory [5], human emotion is not limited to

isolated categories but can be described along three nearly rthogonal dimensions: pleasure-displeasure (P), arousal-nonarousal (A) and dominance-submissiveness (D), namely the PAD emotional model which is proven to be appropriate in describing universal emotions. Ruttkay et al.[6] have proposed the *"Emotion Disc"* and *"Emotion Square"* to generate continuum of facial expression with 2-dimensional navigation. Albrecht et al. [12] have extended the work of MPEG-4 based facial expression synthesis [13], and designed a method to generate mixed expressions, where a two dimension *"Activation-Evaluation"* emotion space is adopted. Kshirsagar et al. [14] have proposed a multi-layered approach including high-level description of emotion intensity over time, mid-level configuration of static facial expressions, and low-level FAPs definitions. However, there are few quantitative results on explicit mapping between emotion dimension such as PAD, and low-level facial expression parameters such as MPEG-4 FAPs [11].

In this paper, a layered framework is proposed to synthesize facial expression using PAD emotional parameters for a Chinese expressive talking avatar. The PAD parameters are adopted as high-level description of emotion state. The MPEG-4 Facial Animation Parameters (FAPs) [11] are used for direct control of mesh deformation on avatar. Different with Kshirsagar's mid-level FAPs configuration of static expression [14], the Partial Expression Parameters (PEPs), which model the correlations among FAPs within local face region, are proposed as mid-level expression parameters to depict common expression movements, such as eyebrow-raise, mouth-open etc. The PAD-PEP mapping model and PEP-FAP translation model are then established for converting high-level PAD emotional parameters to mid-level PEP parameters and then to low-level FAP parameters.

The rest of this paper is as follows. Section 2 introduces the overall architecture of the layered framework for facial expression synthesis. Detailed information about Partial Expression Parameter (PEPs) and the PEP-FAP translation model are presented in section 3. Section 4 introduces the creation of a pseudo facial expression database on three-dimensional avatar. Section 5 presents the design of emotion-expression mapping function that models the correlation between PAD and PEP. The synthetic result and perceptual evaluation is discussed in Section 6. Finally we conclude our work and discuss the future direction.

2   Layered Framework for Facial Expression Synthesis

For geometric facial animation on 3D avatar, we adopt the Facial Animation Parameters (FAPs) provided by MPEG-4 standard [11]. FAP is designed to control the face model points to generate various facial movements. However, it is too complicated for expression synthesizer to manipulate the FAPs directly since they only define the motions (e.g. translations and rotations) of single feature points on the avatar, while ignore the high correlations between different FAPs within the same face region. This motivates us to propose a new parameter to capture the FAP correlations. Based on our previous work on FAP-driven facial expression synthesis [1,9], we propose the Partial Expression Parameters (PEPs) to depict the common expression movement within specific face regions, such as mouth-bent, eye-open and

eyebrow-raise etc. The PEPs capture the correlation among different FAPs, and thus reduces the complexity of FAP-driven expression synthesis [1].

With PAD as high-level emotion description, PEP as mid-level expression configuration and FAP as low-level animation parameter, a layered framework for PAD-driven facial expression synthesis is proposed as shown in Figure 1, where the PAD-PEP mapping model is trained on a pseudo facial expression database with PAD and PEP annotations, and the PEP-FAP translation model is defined experimentally using a home-grown expression editor [9].
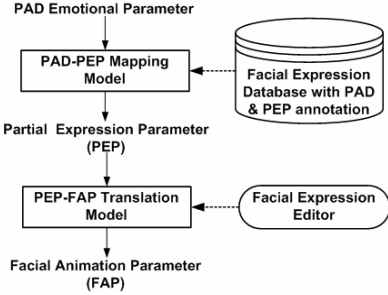

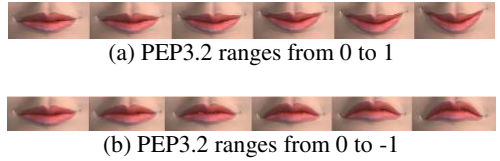**Fig. 1.** Layered framework for PAD-driven facial expression synthesis

(a) PEP3.2 ranges from 0 to 1

(b) PEP3.2 ranges from 0 to -1

Fig. 2. Partial expression movement of mouth-bent with PEP 3.2 ranging in [-1, 1]

3   PEP-FAP Translation Model

3.1   Partial Expression Parameter (PEP)

The PEPs are designed to capture the correlation among FAPs in specific face region for facial expression synthesis. The selection of each PEP features is based on our observation of Japanese Female Facial Expression (JAFFE) database [8] as well as our previous work on FAPs manipulation [1,9]. Detailed partial expression description for PEP can be found in Table 1. The value of the PEP is a scalar ranging in [-1, 1] to simulate the continuous facial expression change. For better understanding, the partial expression movement of *mouth-bent* (PEP 3.2) is illustrated in Figure 2 with its value ranging from -1 to +1.

3.2   PEP-FAP Translation Template

Previous research on FAPs correlation [7] indicates that there are high correlations among different FAPs, and this cross-correlation property has been successfully exploited to interpolate the unknown FAP value from a set of decided FAPs. This motivates us to implement the linear PEP-FAP translation templates experimentally. For each PEP parameter, we define its associated FAP group and key-FAP as shown in Table 2. The mathematical form of the linear PEP-FAP translation template is then defined in Equation 1. For the i-th PEP parameter in region R (P_i^R), the FAP that best represents the corresponding partial expression movement is selected as the key-FAP (F_k) The value of key-FAP is determined by (P_i^R) directly. The value of non-key FAP (F_j) is linear correlated with the key-FAP (F_k) with the coefficient α_k^j The F_k^{max} is the

maximum value of F_k. F_k^{max} and α_k^j are experimentally defined using a home-grown facial expression editor [9].

$$\begin{cases} F_k = P_i^R \cdot F_k^{max} & (P_i^R \in [-1,+1]) \\ F_j = \alpha_k^j \cdot F_k & (\alpha_k^j \in [-1,+1], \ k \neq j) \end{cases} \tag{1}$$

Table 1. PEP definition with partial expression description

Face Region	PEP code (Left/Right)	Partial Expression Description	
		[0,-1]	[0,1]
Eye-brow	1.1(L/R)	Eyebrow lower down	Eyebrow raise up
	1.2(L/R)	Relax Eyebrow	Squeeze Eyebrow
	1.3(L/R)	In the shape of "\ /"	In the shape of "/ \"
Eye	2.1(L/R)	Close eye-lid	Open eye-lid
	2.2(L/R)	(Eyeball) look right	(Eyeball) look left
	2.3(L/R)	(Eyeball) look up	(Eyeball) look down
Mouth	3.1	Close mouth	Open mouth
	3.2	Mouth-corner bent down	Mouth-corner bent up
	3.3	Mouth sipped	Mouth protruded (pout)
	3.4	Mouth stretched in	Mouth stretched out
Jaw	4.1	Jaw move up	Jaw lower down
	4.2	Jaw move right	Jaw move left

Table 2. Related FAP group and key FAP for each PEP parameter

Face Region	PEP Code	FAP Number	Key FAP	Related FAP Group
Eye-brow	1.1	6	F33	[F33,F31,F35,F34,F32,F36]
	1.2	2	F37	[F37,F38]
	1.3	6	F31	[F31,F35, F33, F32, F36, F34]
Eye	2.1	4	F19	[F19,F21,F20,F22]
	2.2	2	F23	[F23,F24]
	2.3	2	F25	[F25,F26]
Mouth	3.1	12	F5	[F5,F10,F11,F52,F57,F58, F4,F8,F9,F51,F55,F56]
	3.2	4	F12	[F12,F13,F59,F60]
	3.3	2	F16	[F16,F17]
	3.4	4	F6	[F6,F7,F53,F54]
Jaw	4.1	1	F3	[F3]
	4.2	1	F15	[F15]

4   Pseudo Facial Expression Database on Avatar

In order to capture the relationship between PAD emotional parameters and PEP parameters, a pseudo facial expression database is created. Here, the "*pseudo*" indicates that the database is not realistic human expression but the cartoon-like expression on the avatar as shown in Figure 3. We choose the Japanese Female Facial Expression (JAFFE) database [8] as reference to build up our pseudo database.

4.1   JAFFE Expression Database

The JAFFE expression database contains 213 expression images with 10 Japanese females posing three or four examples for each of seven basic expressions, *Neutral*, *Happy*, *Sad*, *Surprise*, *Angry*, *Disgust* and *Fear*. For each image we annotated 18 facial points manually according to MPEG-4 facial definition points (FDPs) [11] as shown in Figure 4(a) and 4(b). These facial points are then used to extract the PEP parameters for creating the pseudo facial expression database on avatar.
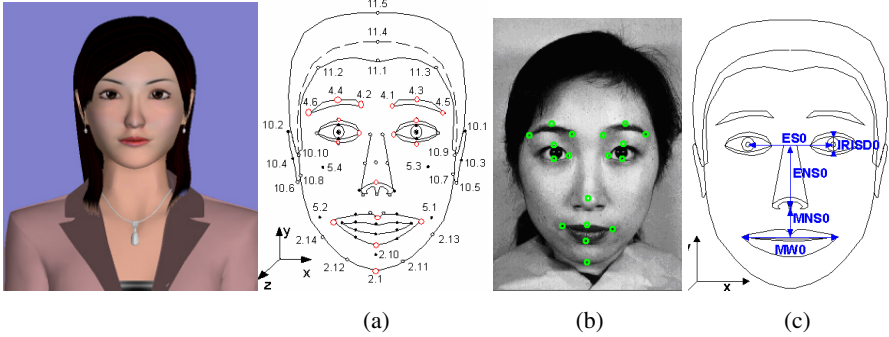

(a)   (b)   (c)

Fig. 3. Avatar front view   **Fig. 4.** Annotated facial points (b), FDPs (a) and FAPU (c)

4.2   PEP Extraction for Pseudo Facial Expression Database

Based on the above facial points annotation result, 12 PEP parameters are then extracted by measuring the movement of facial points as defined in Table 3. The other 6 PEPs are not extracted because the depth information is lacked (PEP 2.2 and 2.3 for

Table 3. PEP measurement by facial point movement

Face Region	PEP code	PEP Measurement	Units (FAPU)
Eye-brow	1.1(L/R)	4.3.y(L) and 4.4.y(R)	ENS
	1.2(L/R)	4.1.x(L) and 4.2.x(R)	ES
	1.3(L/R)	atan($\lvert$4.1.y-4.5.y$\rvert$/$\lvert$4.1.x-4.5.x$\rvert$) (L) atan($\lvert$4.2.y-4.6.y$\rvert$/$\lvert$4.2.x-4.6.x$\rvert$) (R)	AU
Eye	2.1(L/R)	$\lvert$3.9.y-3.13.y$\rvert$ (L) $\lvert$3.10.y-3.14.y$\rvert$ (R)	IRISD
	2.2(L/R)	Not-extracted	AU
	2.3(L/R)	Not-extracted	AU
Mouth	3.1	$\lvert$8.1.y-8.2.y$\rvert$	MNS
	3.2	$\lvert$(8.3.y+8.4.y)/2-(8.1.y+8.2.y)/2$\rvert$	MNS
	3.3	Not-extracted	MNS
	3.4	$\lvert$8.3.x-8.4.x$\rvert$	MW
Jaw	4.1	2.1.y	MNS
	4.2	2.1.x (Not-extracted)	MW

Table 4. Sample number of 7 expressions (labels in the first row) in each PEP clustering center (C_i in the first column)

	HAP	SUP	ANG	NEU	SAD	FEAR	DIS	Total
C1	**25**	1	2	0	1	0	0	29
C2	4	**20**	1	0	0	3	1	29
C3	2	0	**18**	0	9	4	*12*	45
C4	0	6	1	**30**	0	0	0	37
C5	0	0	0	0	**15**	9	2	26
C6	0	3	0	0	4	**11**	2	20
C7	0	0	8	0	2	5	**12**	27
Total	31	30	30	30	31	32	29	213

C1	C2	C3	C4
	C5	C6	C7

Fig. 5. Synthetic expressions for PEP value of clustering center in Table 4

eye-ball movement, PEP 3.3 for mouth movement in z-direction) or the movement is not obvious (PEP 4.2 for horizontal jaw movement) in the database. The FAPU shown in Figure 4(c) is utilized to normalize the extracted PEP to eliminate the individual difference caused by face shape of different subjects. By applying the extracted PEP on avatar, we create the pseudo expression database which has similar synthetic expressions as JAFFE.

A *k-means clustering* experiment is then conducted on the pseudo database to check the validation of the cloned facial expressions. The clustering result is presented in Table 4, which states the number of seven intended expression in each PEP clustering group. The synthetic expression of each clustering center is illustrated in Figure 5. According to the clustering result, the extracted PEP can distinguish the basic expressions to some extent in that the PEP configuration for the same expression will nearly be clustered into the same group with the exception of *Angry* and *Disgust*. This may be understandable because that some commonalities of facial movements are shared for *Angry* and *Disgust* as reported in [1], which causes the current PEP features cannot completely distinguish the two expressions.

5   PAD-PEP Mapping Model

5.1   PAD Annotation for Pseudo Expression Database

As the layered framework for PAD-driven facial expression synthesis in section 2 proposed, the PAD-PEP mapping model is learned from a facial expression database with PAD and PEP annotation. The above pseudo facial expression database already contains the PEP values, thus the PAD values for each synthetic expression in the pseudo database need to be annotated for training the PAD-PEP mapping model.

Based on the PAD model proposed by Mehrabian [5], the Chinese version of abbreviated PAD emotion scales is provided [4] so that the P, A and D values can be evaluated by a 12-item questionnaire as shown in Table 5. The abbreviated PAD scale has been proved as a versatile psychological measuring instrument which is capable of adapting to a variety of applications including emotion annotation.

**Table 5.** 12-item PAD questionnaire for expression annotation and evaluation

Emotion		-4	-3	-2	-1	0	1	2	3	4	Emotion	
Angry	(愤怒的)										Activated	(有活力的)
Wide-	(清醒的)										Sleepy	(困倦的)
Controlled	(被控的)										Controlling	(主控的)
Friendly	(友好的)										Scornful	(轻蔑的)
Calm	(平静的)										Excited	(激动的)
Dominant	(支配的)										Submissive	(顺从的)
Cruel	(残忍的)										Joyful	(高兴的)
Interested	(感兴趣的)										Relaxed	(放松的)
Guided	(被引导的)										Autonomous	(自主的)
Excited	(兴奋的)										Enraged	(激怒的)
Relaxed	(放松的)										Hopeful	(充满希望)
Influential	(有影响力)										Influenced	(被影响的)

For each synthetic expression image in the pseudo expression database, we annotate the P, A and D value using the 12-item questionnaire. During annotation, the annotator is required to describe the synthetic expression using 12 pairs of emotional words as given in Table 5. For each pair of the emotional words, which are just like two ends of a scale, the annotator should choose one of them that better describes the synthetic expression with a 9 level score varying from -4 to +4. The P, A and D values are then calculated from this questionnaire using the method described in [5]. Before annotating, the annotator is trained by expert of psychology on using the PAD questionnaire.

The annotation results are summarized in Table 6, with the corresponding distribution in PAD emotional space shown in Figure 7. We can see that the PAD annotations for each basic emotion category are distributed in nearly different areas with the exception of *Angry* and *Disgust*. This result is consistent with the clustering result of PEP configuration as described in section 4.2

Table 6. Average PAD annotation for the pseudo expression database

Intent Emotion	MEAN		
	P	A	D
ANG	-0.59	0.08	0.47
DIS	-0.59	-0.01	0.40
FEAR	-0.08	0.18	-0.39
HAP	0.63	0.40	0.29
NEU	0.03	-0.04	-0.07
SAD	-0.28	-0.12	-0.37
SUP	0.41	0.55	0.19
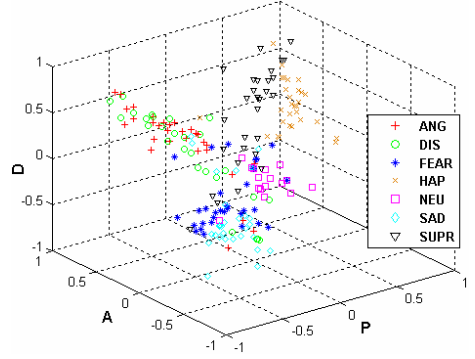
Fig. 7. Distribution of PAD annotations for the pseudo expression database

5.2 PAD-PEP Mapping Model

Different from the previous work on facial expression manipulation, our work focuses on exploring the relationship between high-level emotion description and mid-level expression configuration, in other words, our goal is to implement a mapping function to convert the PAD into PEP. The pseudo expression database containing the PAD annotation as well as PEP configuration is used to train the PAD-PEP mapping model.

In order to find a proper emotion-expression mapping function between PAD and PEP, we have tried the polynomial function with the first and second order as well as non-linear exponential function. As the experimental result reveals, it is the second order polynomial function that receive the best fitting result, and the mathematic form is shown in Equation 2.

$$PEP = \alpha \cdot E^2 + \beta \cdot E + \delta \qquad (2)$$

Where **PEP** is the PEP vector of expression configuration, and **E** is the PAD vector $[P, A, D]$, E^2 is also a vector with each element the square value of its counterpart in **E** respectively, i.e. $[P^2, A^2, D^2]$, **α** and **β** are the corresponding coefficient matrix, **δ** is the constant offset vector. It should be noticed that, each dimension of **PEP** vector is estimated respectively with the same mathematic form as equation 3 shown. The PEP_i is the i-th dimension in **PEP** vector, and all the corresponding function coefficients are indicated with the subscript i and P-A-D dimensions.

$$PEP_i = [\alpha_{Pi}, \alpha_{Ai}, \alpha_{Di}] \begin{bmatrix} P^2 \\ A^2 \\ D^2 \end{bmatrix} + [\beta_{Pi}, \beta_{Ai}, \beta_{Di}] \begin{bmatrix} P \\ A \\ D \end{bmatrix} + \delta_i \qquad (3)$$

6 Experiments

6.1 Experiments on PAD-PEP Mapping Function

The pseudo expression database consists of only 213 expression samples, which are extracted from the JAFFE database consisting of 10 subjects and 7 expressions. In order

to reduce the limitation caused by the small database, we adopt the K-fold cross-validation method to train the PAD-PEP mapping function. For the test set, we select 10*2=20 expression image samples by randomly selecting 2 different expressions for all 10 subjects. The training set, which consists of the rest 193 expression image samples, is then divided into 10 subsets with each subset covering all 10 subjects and each subject with 1-2 expressions. By such division scheme, we are able to capture the common facial expression movement shared by different people as much as possible in the PAD-PEP mapping function.

The coefficients of the PAD-PEP mapping function is estimated using the least square errors method. It should be noticed that each dimension of PEP (e.g. PEP_i) is estimated separately with the same form as shown in equation 3, which means that we finally have 12 sets of coefficients corresponding to 12 PEP parameters.

Table 7. Correlation coefficients for evaluating PAD-PEP mapping function

PEP code	Validation set (k=10)			Test set
	Min	Max	Avg	
1.1L	0.56	0.88	0.77	0.74
1.1R	0.52	0.86	0.77	0.77
1.2L	0.13	0.82	0.43	0.51
1.2R	0.13	0.64	0.44	0.62
1.3L	0.63	0.89	0.79	0.80
1.3R	0.56	0.91	0.77	0.82
2.1L	0.53	0.84	0.75	0.60
2.1R	0.51	0.85	0.76	0.62
3.1	0.36	0.89	0.65	0.52
3.2	0.24	0.93	0.71	0.75
3.4	0.06	0.77	0.52	0.78
4.1	0.36	0.83	0.57	0.66
Average	0.38	0.84	0.66	0.69

In the K-fold cross-validation training process (k=10), there are 10 iterations corresponding to 10 validating subset. For each iteration, we calculate the correlation coefficients between real and estimate data on current validating subset as criteria to evaluate the fitting performance of the trained function. The minimum, maximum and average value of correlation coefficient is summarized in Table 7. The trained function with the average fitting performance among all the 10 iterations is chosen as the final result and used to evaluate the test set.

From table 7, we can see that the performance of mapping function on test set is acceptable with correlation coefficient be around 0.70. There is still some space for performance to be improved. This may be explained as that the pseudo expression database consists only 10 subjects with 7 basic expressions, which is insufficient for training the mapping function to capture common expression patterns among different people. The possible solution is either to increase the number of the subjects or to collect more expression images for a specific subject.

6.2 Synthetic Expression and Perceptual Evaluation

A series of perceptual evaluation is conducted to evaluate the performance of the layered framework for expression synthesis. The synthetic expressions for the 6 basic emotion words (*happy*, *surprise*, *sad*, *scared*, *angry*, and *disgusting*) and some other candidate emotion words with corresponding PAD value are shown in Figure 8. 14 subjects are then invited in the evaluation to finish the 12-item PAD questionnaire as shown in Table 5 for each synthetic facial expression image. Another emotion labeling experiment is also conducted where the subjects are required to select one word that best describe the synthetic expression from 14 candidate emotion words, which are *Happy*, *Optimism*, *Relax*, *Surprise*, *Mildness*, *Dependent*, *Bored*, *Sad*, *Scared*, *Anxious*, *Scornful*, *Disgusting*, *Angry*, and *Hostile*.
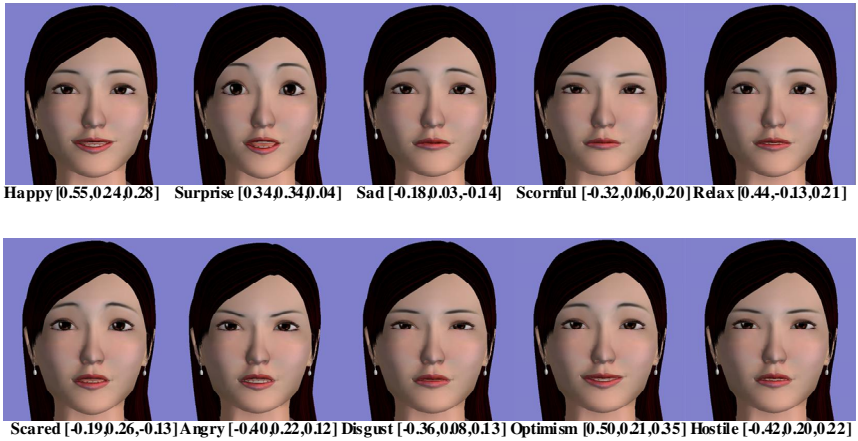

Happy [0.55,0.24,0.28]   Surprise [0.34,0.34,0.04]   Sad [-0.18,0.03,-0.14]   Scornful [-0.32,0.06,0.20] Relax [0.44,-0.13,0.21]

Scared [-0.19,0.26,-0.13] Angry [-0.40,0.22,0.12] Disgust [-0.36,0.08,0.13] Optimism [0.50,0.21,0.35] Hostile [-0.42,0.20,0.22]

Fig. 8. PAD-driven synthetic expression for selected emotions (Emotion [P,A,D])

Table 8. Result of PAD evaluation and emotion labeling. (The PAD value of *Hostile* is [-0.42, 0.20, 0.22]).

PAD-driven synthetic expression	Original PAD of emotion words			Evaluation PAD of Synthetic image			Expression Label (with voting percent)
	P	A	D	P	A	D	
Happy	0.55	0.24	0.28	0.42	0.12	0.10	Happy (67%)
Surprise	0.34	0.34	0.04	0.36	0.45	-0.05	Surprise (100%)
Sad	-0.18	0.03	-0.14	-0.01	-0.26	-0.27	Sad (42%)
Scared	-0.19	0.26	-0.13	0.01	-0.04	-0.25	Sad (50%)
Angry	-0.40	0.22	0.12	-0.17	0.02	-0.08	Hostile (58%)
Disgust	-0.36	0.08	0.13	-0.56	0.15	0.44	Disgusting (50%)

The PAD values for each basic emotion word and the PAD perception values for synthetic expression are summarized in Table 8. A voting method is applied to determine the emotion category for each synthetic facial expression image based on the result of expression labeling. The PAD correlation coefficients between emotion words and synthetic expression are 0.89(P), 0.68(A) and 0.70(D) respectively, this

result is consistent with the reliability and validity of Chinese version of abbreviated PAD emotion scales reported in [4]. Confusion between *Sad* and *Scared* as well as *Angry* and *Hostile* is found in expression labeling experiment, it is because that the input PAD value for each of the two confused words are very close as shown in Table 8, which originates from the confusion in human understanding between these emotion words and thus leads to the similarity between synthetic expressions.

The experimental results indicates that the PAD emotional parameters can be used in describing the emotion state as well as facial expression; and the proposed layered framework is effective for PAD-driven facial expression synthesis.

7   Conclusion and Further Work

This paper proposes a layered framework for facial expression synthesis on a Chinese expressive avatar. The synthetic facial expression is controlled by the high-level PAD emotional parameters, namely the Pleasure-Arousal-Dominance, by which the human emotion can be described with three independent orthogonal dimensions. The mid-level Partial Expression Parameter (PEP) is proposed to depict the facial expression movement in a specific local face region. The MPEG-4 Facial Animation Parameter (FAP) is adopted as low-level parameters for direct manipulation of model points on the avatar. The PAD-PEP mapping model and PEP-FAP translation model are then implemented to translate the PAD parameters to PEP parameters and then to FAP parameters for facial expression synthesis.

The proposed approach is effective in facial expression synthesis, which enriches the avatar with emotions. Together with our previous work on expressive head movement synthesis [10], we can generate a more engaging avatar for expressive visual speech synthesis. Further work will focus on combining these works to develop a text-driven expressive talking avatar for TTAVS synthesis.

References

1. Wu, Z.Y., Zhang, S., Cai, L.H., Meng, H.M.: Real-time Synthesis of Chinese Visual Speech and Facial Expressions using MPEG-4 FAP Features in a Three-dimensional Avatar. In: Proc. Int. Conf. on Spoken Language Processing, pp. 1802–1805 (2006)
2. Graf, H.P., Cosatto, E., Strom, V., Huang, F.J.: Visual Prosody: Facial Movements Accompanying Speech. In: Fifth IEEE Int. Conf. on Automatic Face and Gesture Recognition, pp. 381–386 (2002)

3. Du, Y.Z., Lin, X.Y.: Emotional facial expression model building. Pattern Recognition Letters 24(16), 2923–2934 (2003)
4. Li, X.M., Zhou, H.T., Song, S.Z., Ran, T., Fu, X.L.: The Reliability and Validity of the Chinese Version of Abbreviated PAD Emotion Scales. In: Int. Conf. on Affective Computing and Intelligent Interaction (ACII), pp. 513–518 (2005)
5. Mehrabian, A.: Pleasure-arousal-dominance: A General Framework for Describing and Measuring Individual Differences in Temperament. Current Psychology. Developmental, Learning, Personality, Social 14, 261–292 (1996)
6. Ruttkay, Z., Noot, H., Hagen, P.: Emotion Disc and Emotion Squares: Tools to Explore the Facial Expression Space. Computer Graphics Forum 22(1), 49–53 (2003)
7. Lavagetto, F., Pockaj, R.: An Efficient Use of MPEG-4 FAP Interpolation for Facial Animation at 70 bits/Frame. IEEE Transactions on Circuits and Systems for Video Technology 11(10), 1085–1097 (2001)
8. Lyons, M., Akamatsu, S., Kamachi, M., Gyoba, J.: Coding. facial expressions with gabor wavelets. In: Proc. of the 3rd IEEE Conf. on Face and Gesture Recognition, pp. 200–205 (1998)
9. Zhang, S., Wu, Z.Y., Cai, L.H.: Region-based Facial Expression Synthesis on a Three-dimensional Avatar. In: China Conference on Human Computer Interaction(CHCI) (2006)
10. Zhang, S., Wu, Z.Y., Meng, H.M., Cai, L.H.: Head Movement Synthesis based on Semantic and Prosodic Features for a Chinese Expressive Avatar. In: IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP) (2007)
11. Motion Pictures Expert Group, ISO/IEC 14496-2: 1999/Amd. 1: 2000(E). International Standard, Information Technology – Coding of Audio-Visual Objects. Part 2: Visual; Amendment 1: Visual Extensions
12. Albrecht, I., Schröder, M., Haber, J., Seidel, H.P.: Mixed feelings: expression of non-basic emotions in a muscle-based talking head. Virtual Reality 8(4), 201–212 (2005)
13. Tsapatsoulis, N., Raousaiou, A., Kollias, S., Cowie, R., Douglas-Cowie, E.: Emotion recognition and synthesis based on MPEG-4 FAPs MPEG-4 facial animation—the standard implementations applications, pp. 141–167. Wiley, Hillsdale (2002)
14. Kshirsagar, S., Escher, M., Sannier, G., M.-Thalmann, N.: Multimodal Animation System Based on the MPEG-4 Standard. In: Multimedia Modelling 99, Ottawa, Canada, pp. 215–232 (1999)