# Happy Companion: A System of Multimodal Human-Computer Affective Interaction

*Jia Jia*[1,2,3], *Lianhong Cai*[1,2,3], *Sirui Wang*[4], *Xiaolan Fu*[4]

1 State Key Laboratory on Intelligent Technology and Systems
2 Tsinghua National Laboratory for Information Science and Technology（TNList）
3 Department of Computer Science and Technology, Tsinghua University, Beijing, China
4 State Key Laboratory of Brain and Cognitive Science, Institute of Psychology, Chinese Academy of Science, China
E-mail: jiajia@mails.tsinghua.edu.cn, clh-dcs@tsinghua.edu.cn, wangsr@psych.ac.cn, fuxl@psych.ac.cn

## Abstract

Emotion is one of the most important aspects in the intelligent human computer inter-actions. The concept of “Affective Computing” refers to the “computing that relates to, arises from, or deliberately influences emotion or other affective phenomena”. The research on realization of emotion comprehension and expression has been a hot topic. In this paper, we proposed a novel human-computer affective interaction system, called “Happy Companion”. We adopt the PAD model with three dimensional: pleasure (i.e. valence, evaluation), arousal (i.e. activation, activity), and dominance (i.e. power, potency), as the description of human’s emotional states. By integrating relevant technologies, such as PAD quantification and annotation, emotion detection and analysis from speech and video, 2D/3D facial expression synthesis, et al., the happy companion presents a virtual guider who can detect your emotional states and respond expressively through audio-visual interface in real time. Under her guide, you can participate four basic tasks to regulate your emotions, including Breathing Relaxation, Muscle Relaxation, Facial Expression Learning, Joke and Game. The design and study of “happy companion” have made a try to create a system of multimodal human-computer affective interaction.
Key words: Affective computing, PAD model, multimodal, human-computer interaction, emotion regulation.

## 1 Introduction

Affective computing is a branch of neurological study that deals with the design of systems that can recognize, interpret, and process human emotions. It is an interdisciplinary field spanning computer sciences, psychology, and cognitive science. The concept of “Affective Computing” is proposed by Since Picard[18] in 1997, which refers to the “computing that relates to, arises from, or deliberately influences emotion or other affective phenomena”. The research on realization of emotion comprehension and expression has been a hot topic in affective human computer interactions, including emotional mechanism, acquiring of emotional signal, analysis, modeling, identification, emotional understanding and feedback, realization of human-computer interaction. Two main aspects of emotion should be addressed for affective systems that understand, manipulate, and deliver emotional information. Firstly, how to describe human’s emotional state in a computational environment? Secondly, how to detect or deliver the abundant emotional information through multimodal channels, such as emotional speech, emotional facial expression?

For the description of human’s emotional state, The most earliest academic research on human’s emotion may date back to Charles Darwin[7], Darwin proposed to use the term of “state of mind”. Ekman et al. [9] proposed the “basic emotion program” on the research of emotional human facial expression in 1970s. Due to the vagueness in human’s perception and cognition of emotion, it is hard to distinguish and recognize the subtle and complex emotion. For dimensional emotion description, ideally, arbitrary emotional state can be quantitatively measured in terms of a small number of basic dimensions. Mehrabin et.al [14] developed the PAD emotional model by which all emotions can be located in 3D emotion space where the pleasure, arousal and dominance dimensions are orthogonal and independent to each other.

For detecting and delivering emotion through multimodal channels, most research focus on modeling the emotional speech and facial expressions. The traditional facial expression synthesis is based on the emotion categories, especially the big six emotions[3][8][12][19][22]. With the development of dimensional approach to emotion description, recently, a lot of research works are devoted in dimensional emotional speech analysis[6][20], expressive speech synthesis based on emotion dimensions[20][25][26], the facial expression recognition[3], and also the facial expression synthesis[1][28], multimodal emotion recognition[5], emotional simulation[10].

In this paper, we choose the PAD model with three dimensional: pleasure (i.e. valence, evaluation), arousal (i.e. activation, activity), and dominance (i.e. power, potency). Based on PAD model, we proposed a novel human-computer affective interaction system, called “happy companion”. The happy companion pre-

sents a virtual guider who can detect your emotional states and respond expressively through audio-visual interface in real time. It is a prototype software that tries to simulate the emotion regulation in traditional psychological consultation through multimodal human-computer interaction. The virtual "happy companion" is a 3D talking avatar who can talk and behave expressively. Under her guide you can participate four basic tasks to regulate your emotions, including Breathing Relaxation, Muscle Relaxation, Facial Expression Learning, Joke and Game.

The system integrates the techniques from the research on affective computing which is supported by the fund from the National Science Foundation of China (NSFC) under grant No.60433030, including emotion appraisal, detection, analysis and expression. These techniques are jointly developed by Department of Computer Science and Technology, Tsinghua University and Institute of Psychology, Chinese Academy of Science.

The rest of this paper is organized as follows: Section 2 introduces the design and key technologies of the "happy companion" system. Section 3 describes the emotion regulation tasks based on psychology. Section 4 presents the multimodal affective interaction in the proposed system. Finally, Section 5 concludes.

## 2 Happy Companion: System Design and Key Technologies

### 2.1 System Design

The happy companion is designed to simulate the emotion regulation in traditional psychological consultation through multimodal human-computer interaction. The design of happy companion is shown as Figure 1 (a) and (b).The system consists of the following human-computer affective interaction technologies.

Getting Anxiety Level by Subjective Evaluation (sub-section 4.1).

Selecting emotion regulation tasks based on psychology (section 3).

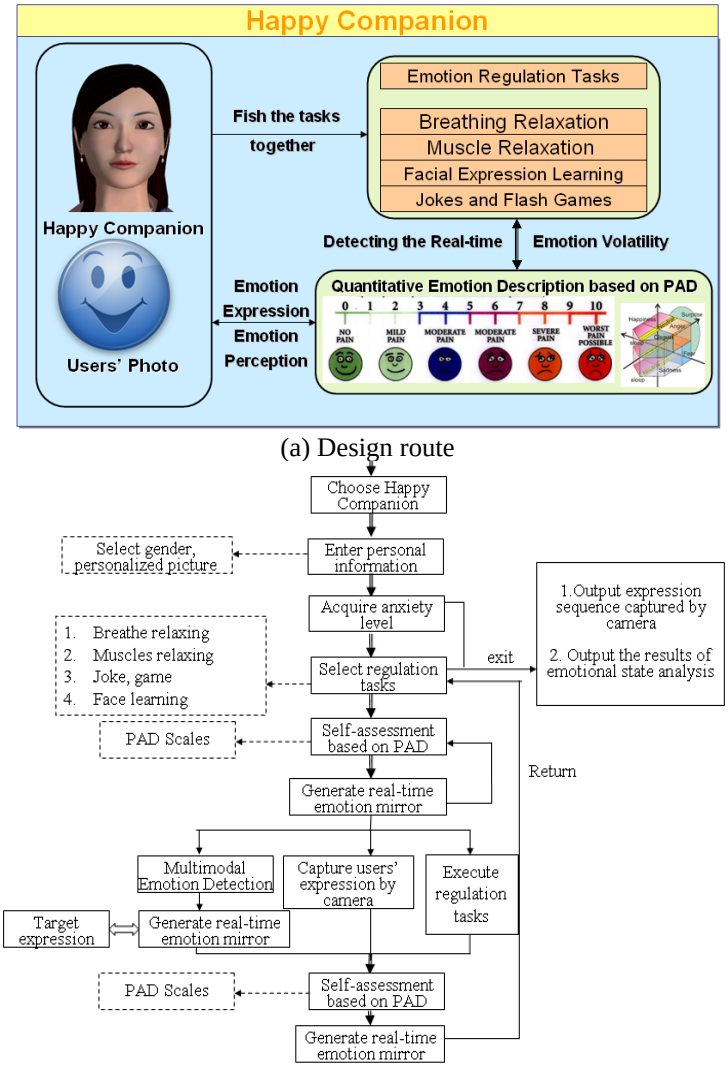
Self-assessment based on PAD Annotation (sub-section 4.2).

Synthesizing the 2D "emotion mirror" (sub-section 4.3), using the technology of facial expression synthesis based on PAD (sub-section 2.2). The "emotion mirror" means synthesize the facial expressions on a photo/2D cartoon image according to users' real-time facial expressions.

Emotional speech detection (sub-section 2.2).

Emotional state analysis based on facial image sequence (sub-section 4.4), using the technology of facial expression analysis based on PAD (sub-section 2.2).

Generating the 3D virtual "happy companion" (sub-section 4.5).




(a) Design route

(b) The flowchart of happy companion

**Figure 1** The design of happy companion system

### 2.2 Key Technologies

PAD Emotion Model: With the advanced development of affective computing and emotional research, researchers have proposed a series of emotional models that are based on the dimensional methodology. These emotional models are utilized in the description and analysis of emotional state, and further are involved in the recognition, comprehension, and expression of human emotions. Various emotional dimensions can be summarized into three basic dimensions. Among which, the PAD model proposed by Mehrabian from UCLA [14] are widely adopted and used. This model describes the emotional state from the aspects of perception and evaluation, physiological arousal and also the social communication. The PAD model contains

three basic dimensions, which are:1) Pleasure-displeasure: distinguishes the positive-negative quality of emotional state;2) Arousal-nonarousal: measures the degree of physical activity and mental alertness of emotional state;3) Dominance-submissiveness: defined in terms of control versus lack of control, mainly focus on the social aspect of emotion.The three dimensions are independent and orthogonal to each other, and their combination builds up the three dimensional space of emotion. Different emotional states are distributed in the PAD space as points or regions.

The three dimensional emotion models has been widely accepted and used in research areas in affective computing. The International Affective Picture System (IPAS) built by Bradley et al. from Florida University, has been annotated with the PAD emotion model [2]. The researcher from DFKI of Germany has adopted the PAD model to synthesize emotional speech [20], and also build the layered affect model (ALMA) [10]. The researchers from USC utilize the PAD model in their study on the analysis of distinctive perception on audio-visual emotional expression [15]. In the our affective computing research (Project finished by Tsinghua University and Chinese Academy of Science), the PAD model makes contribution to the emotion computational theory [23], the recognition and synthesis of emotional speech [6][16][26], the analysis and synthesis of facial expression [3][28].

Emotion-detecting based Model Selection for Emotional Speech Recognition: For emotion detection from the speech, most widely used approaches include robust feature extraction, speaker normalization and model tuning/retraining. We proposed a novel method, that is, adaptation technique is adopted to transform a general model into emotion-specific one with a small amount of emotion speech. Several emotion-dependent AMs by using small amount of corresponding emotional speech data are built. And the emotion of the test speech was detected using GMM-based method. Moreover, a model-selection strategy based on emotion-detection was proposed and proven to be effective, and the overall mean recognition rate increased to 80.79%with an Error Rate Reduction (ERR) of 16.55% compared to the neutral speech Acoustic Model (AM) [16].

PAD Model based Facial Expression Analysis: Considering of the complexity of expressions, it is very difficult to find a general approach to recognize not only expressions but also the grades of expressions by the discrete basic expression models. To solve this problem, we use PAD model for measuring expressions, and Multi Grades Model (MGM) is proposed [3]. This model divides facial expression into several grades. Several models classifying different grades are trained by SVM. Gabor feature is also used. Test data are classified by the models first, and then the results vote to decide the final grade. The rule of voting is as follows: if the data is classified to a certain space, put a vote to this space. After the data is classified by all models, each space's votes are counted, and the test data should be classified into the space which has the most votes [3].

PAD Model based Facial Expression Synthesis: For facial expression synthesis, PAD emotional parameters are applied to describe emotional facial state, and a set of partial expression parameters are proposed to describe the facial movement in local face region. To synthesize facial expression for emotional state in PAD space, a layered parametric framework is proposed, where the PAD is taken as high-level description of emotional state, a set of Partial Expression Parameter (PEP) is proposed to measure the dominant expressive movement in face regions, and the MPEG-4 Facial Animation Parameters (FAP) is used to animate the 2D cartoon images. More details about PEP and its relationship with PAD and FAP can be found in [28].

## 3 Emotion Regulation Tasks

### 3.1 Facial Expression Learning

Facial expression is the window of communication that can sensitively reflect a person's physical and mental states. Conversely, different expressions correspond to different emotions. When you feel depressed, you can make a conscious smile. When you feel nervous, you can consciously relax facial muscles. All these actions will produce a relaxed feeling. Smile is a simple and effective expression regulation skill. Sincere smile can express yourself and others' love. Smile from the heart could make you enjoyable.

### 3.2 Muscle Relaxing

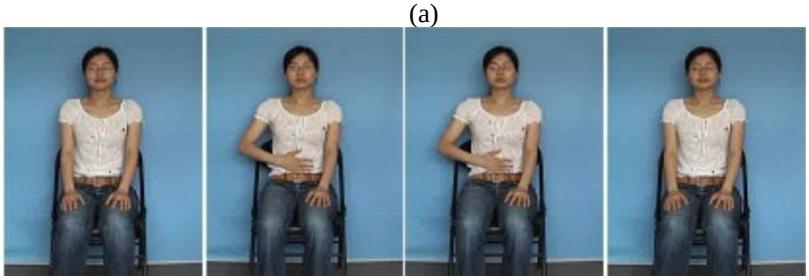
Progressive muscle relaxing is a simple skill. It has been proved that anxious mind can not exist in the body of relaxation. Progressive muscle relaxing needs to tighten and relax the sixteen muscle groups for twenty minutes. You allow yourself to put worry aside and let feeling calm down. To successful experience of relaxing feel depends on the importance of the calm mood. Do not worry whether you can use the skill well. Do not force to relax or control your body. Do not evaluate your performance. Just focus on the muscles during the practice process, and try to feel natural.

### 3.3 Breath Relaxing

People's breathing often become light and fast. And it also has severe ups and downs of the chest in a state of tension. If the frequency of light breath is too fast, it will lead to hyperventilation. Hyperventilation

can cause vertigo, dizziness and palpitations, such as anxiety. When people are in a relaxed state, breathing is often deep and full, and breathing is in the abdomen rather than the chest. Tension and abdominal breathing can not occur simultaneously. Abdominal breathing provides more oxygen and discharges the body of toxins more effectively, so it makes people more relaxed and focused. In other words, abdominal breathing can bring a range of relaxed effects.

(a)



(b)

**Figure 2** (a) The tutorial video sequence of muscle relaxing;(b)The tutorial video sequence of breath relaxing

In addition, we also offer jokes and flash games, as a means of regulating emotions, for users to choose.

## 4 Multimodal Affective Interaction

In happy companion system, we design a framework of multimodal affective interaction. For users' affective information acquisition, we have four different interfaces: 1) subjective evaluation on anxiety level, 2) self-assessment of based on PAD, 3) emotion composition detection from the speech, and 4) emotion analysis based on facial images. For users' affective information feedback, we have three interfaces: 1) the emotion regulation tasks recommended to the users according to the results of affective information acquisition. 2) the "emotion mirror" synthesis, which means synthesize the facial expressions on a photo/2D cartoon image according to users' real-time facial expressions. The input is the fusion of the multi-modal emotional information, including anxiety level, PAD values, the results of emotional speech recognition and facial expression detection. 3) the emotional state analysis based on facial image sequences. The facial image sequence is obtained during the whole course of using the software. Also, a 3D talking avatar is used as the "virtual guide", who can talk and behave expressively. Under her guide you can participate four emotion regulation tasks, including facial expression learning, breathing relaxation, muscle relaxation, Joke and Game.

### 4.1 Getting Anxiety Level by Subjective Evaluation

We use a subjective evaluation based on psychology to obtain the user's anxiety level, shown as Table 1.

**Table 1** The subjective evaluation on anxiety level

| The Question: | When you feel anxious, you will…… |
|---|---|
| 1 | Think the results of all things in the worst point of view, regardless of the reason of making you feel anxious at the beginning. |
| 2 | Start to do everything to solve the problem immediately. If you don't do like this, you will be anxious seriously. |
| 3 | Reproduce the process of solving the problem again and again in your mind, even you have solved the problem. |
| 4 | Keep thinking the situation occurred in the process of solving the problem for a few hours or days, when you have finished the problem. |
| 5 | Feel out of control. |
| 6 | Your stomach is upset; mouth is dry and heart is racing. |
| 7 | Tend to fuss. |
| 8 | Difficult to sleep at night. |
| 9 | Appear the phenomenon of expressing difficultly and hand shaking. |
| 10 | Thinking oddly. |
| Options for each question: | (a) always; (b) often; (c) seldom; (d) never |
| Score of each option: | (a)=4; (b)=3; (c)=2; (d)=1 |
| Anxiety level: | Compute the total score of the ten questions |
| High | If the total is between 25 and 40, your level of anxiety is high. |

| | |
|---|---|
| Middle | If the total is between 20 and 24, your level of anxiety is middle. |
| Low | If the total is less than 20, your level of anxiety is low. |

### 4.2 Self-assessment based on PAD Annotation

For the affective computing based on PAD emotion model, an open question is that how to obtain the PAD value for different research area (e.g. speech, text, facial expression). In another words, how to measure the emotional information from multimodal channels, and then obtain quantitatively analysis results? To this end, Mehrabian has proposed the PAD emotional scales to obtain the PAD values based on a series of psychological experiments [13]. Psychologists from Chinese Academy of Science has translated the original PAD scales into a Chinese abbreviated version[11], and conducts a lots of experiments on a large population to test and verify the Chinese version of PAD scale. The Chinese PAD scales contain 12-item semantic differential scales, as shown in Table 2. Each of the items corresponds to a couple of adjectives that describe different emotional states, and each item is quantitatively divided into nine-degree scales from -4 to +4. For better understanding and the practical use in our system, we also implemented an image version of the PAD scales as shown in Figure 3. In the image version scale, adjectives for each of the item are replaced by a cartoon images.

**Table 2** PAD Annotation Scale

| Emotion | | -4 | -3 | -2 | -1 | 0 | 1 | 2 | 3 | 4 | Emotion | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Angry | 愤怒的 | | | | | | | | | | Activated | 有活力的 |
| Wide-awake | 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 | | | | | | | | | | Autono- | 自主的 |
| Excited | 兴奋的 | | | | | | | | | | Enraged | 激怒的 |
| Relaxed | 放松的 | | | | | | | | | | Hopeful | 充满希望的 |
| Influential | 有影响力 | | | | | | | | | | Influenced | 被影响的 |


**Figure 3** Image version of the PAD scales

Based on the PAD scales, by a series of perceptual evaluations, we can annotate the emotional information embedded in human's speech, facial expression or mental activities. During the experiments, user should make judgments of their current emotional state, and make selection from each pair of the adjectives in each item. The selected adjectives are the best to describe users' current emotional state. According to the user's annotation of PAD scales, the original PAD value can be obtained by the following formula. The Q1, Q2,…, Q12 correspond to the annotated value for each item (from top to bottom) in Table 2. It can be concluded that each of the item (Qi) corresponds to one of the P, A and D dimensions measurement.

$$\begin{aligned} P &= (\ Q1 - Q4 + Q7 - Q10)/4 \\ A &= (-Q2 + Q5 - Q8 + Q11)/4 \\ D &= (\ Q3 - Q6 + Q9 - Q12)/4 \end{aligned} \qquad (1)$$

### 4.3 Synthesizing the 2D "Emotion Mirror"

According to the fusion of the multi-modal emotional information, including anxiety level, PAD values, the results of emotional speech recognition and facial expression detection, we present the real-time synthetic facial expressions on a photo or cartoon image to the users. The flowchart of facial expression synthesis based on user's 2D image is shown in Figure 4. Based on the user's 2D image, we utilize the face alignment

algorithm [27] to locate the 88 facial feature points in face image. Based on the 88 facial feature points, we transform the standard 2D face mesh into a personalized face mesh that matches the user's 2D image. The face mesh contains 218 mesh points, and 361 mesh triangles. The PAD-PEP-FAP mapping function is used to generate the target animation parameters (FAP), and the face mesh is animated by the FAP with the Raised Cosine Function [17]. Figure 5 gives an example of synthesis process and Figure 6 illustrates the synthetic facial expressions.

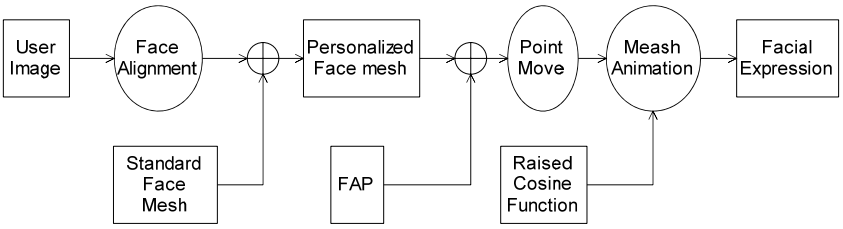



**Figure 4** The flowchart of facial expression on 2D image

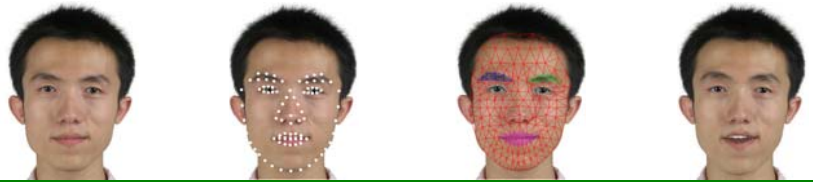


**Figure 5** Illustration of process of facial expression synthesis. (a) User image (b)Feature points (c) Personalized face mesh (d) Facial expression

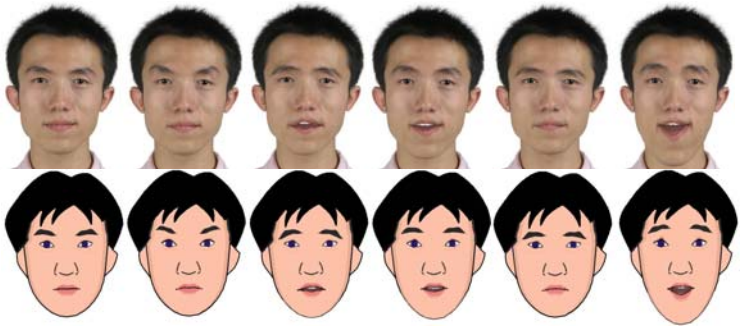


**Figure 6** Illustration of Synthetic facial expression on 2D images

### 4.4 Emotional State Analysis based on Facial Image Sequence

**Figure 7** The results of emotional state analysis in happy companion. The template of the results is "您刚来的时侯比较 ，在跟我们一起进行了游戏娱乐等活动后，现在您显得比较 ，真是不错。在与我的交流中，我感觉你的心情以比较 为主导，起伏不定。给出的建议如下："您一直笑口常开，保持就好啦"(in Chinese)

During the course of using Happy Companion, the user's facial images are obtained automatically at a certain interval. We design an emotion state analysis module to evaluate the changes of users' mood. The emotional state analysis module is using MGM model mentioned in sub-section 2.2. We mainly analyze the following four aspects: the initial emotional state, the final emotional state, emotional state volatility, and the dominant emotional state.

The initial and final emotional states are determined by the first and last facial image of the sequence respectively. Emotional volatility is determined by the amplitude and frequency of the facial image sequence. Dominant emotional state is the average of the decision by the emotional sequences. After dealing with emotion after sequence analysis module, the output shown in Figure 7.

### 4.5 Generating the 3D virtual "happy companion"

A three-dimensional talking avatar is adopted as a virtual "happy companion" in our system. Firstly, she will give a brief introduction on the purpose of the system, and also the basic flowchart on how to participate in the activities provided by the system, as shown in Figure 8. After the introduction phase, the talking avatar will stay on the left-top corner of the system interface, as shown in Figure 9. If the user encounters any problems during each stage of system, he/she can activate the talking avatar, and the talking avatar will give the proper explanation on how to use the system at current stage. The speech of talking avatar is pre-recorded, and the mouth-movement (i.e. viseme) is synthesized based the dynamic Chinese viseme model [24], and the facial expression is synthesized according to the PAD-PEP-FAP mapping model. With the talking avatar as the virtual psychologist, the intelligence and naturalness of the system is greatly enhanced.

**Figure 8** Introduction Stage

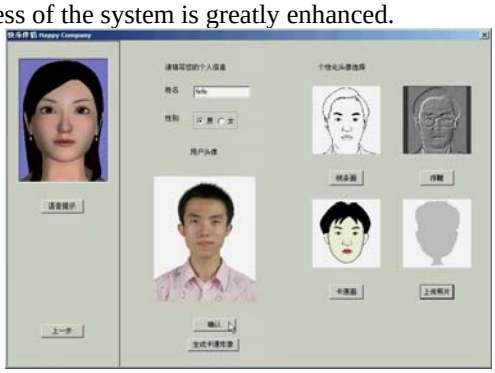


**Figure 9** Talking Avatar in system running

## 5 Conclusion

In this paper, we proposed a novel human-computer affective interaction system, "Happy Companion". We adopt the PAD model to describe and quantify the human emotional states. Based on PAD model, we design a framework of multimodal affective interaction. For users' affective information acquisition, the happy companion integrates four different interfaces: 1) subjective evaluation on anxiety level, 2) self-assessment of based on PAD, 3) emotion composition detection from the speech, and 4) emotion analysis based on facial images. For users' affective information feedback, the happy companion integrates three interfaces: 1) the emotion regulation tasks recommended to the users according to the results of affective information acquisition. 2) the "emotion mirror" synthesis, and 3) the emotional state analysis based on facial image sequences. A 3D talking avatar is used as the "virtual guide" to help the users to participate different emotion regulation tasks. The "happy companion" is still a prototype software, with our target of realize a system that tries to simulate the emotion regulation in traditional psychological consultation through multimodal human-computer interaction.

## 6 Acknowledgements

This paper is supported by the research fund from the National Science Foundation of China (NSFC) under grant No.60433030, and the National High Technology Research and Development Program("863"Program) of China(2009AA011905). The authors would also like to thank all the staff worked for the affective computing project, from the Department of Computer Science and Technology, Tsinghua University and Institute of Psychology, Chinese Academy of Science.

## References

[1] Boukricha, H., Wachsmuth, I., Hofstatter, A., et al. 2009. Pleasure-Arousal-Dominance Driven Facial Expression Simulation. Paper presented at the 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops(ACII 2009), Amsterdam, Netherlands.

[2] Bradley, M. M. and Lang, P. J. 2007. The International Affective Picture System (Iaps) in the Study of Emotion and Attention. In Coan, J. A. & Allen, J. J. B. (Eds.), Handbook of Emotion Elicitation and Assessment (pp. 29-46). New York: Cambridge University Press.

[3] Cao, J., Wang, H., Hu, P., et al. 2008. Pad Model Based Facial Expression Analysis. Paper presented at the 4th International Symposium Advances in Visual Computing(ISVC 2008), Las Vegas, NV, United states.

[4] Cao, Y., Tien, W. C., Faloutsos, P., et al. 2005. Expressive Speech-Driven Facial Animation. ACM Transaction on Graph, 24(4), 1283-1302.
[5] Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., et al. 2001. Emotion Recognition in Human-Computer Interaction. IEEE Signal Processing Magazine, 18(1), 32-80.
[6] Cui, D., Meng, F., Cai, L., et al. 2007. Affect Related Acoustic Features of Speech and Their Modification. Paper presented at the Affective Computing and Intelligent Interaction (ACII2007).
[7] Darwin, C. 1965. The Expression of the Emotions in Man and Animals: Chicago: University of Chicago Press.
[8] Deng, Z., Neumann, U., Lewis, J. P., et al. 2006. Expressive Facial Animation Synthesis by Learning Speech Coarticulation and Expression Spaces. IEEE Transaction on Visualization and Computer Graphics, 12(6), 1523-1534.
[9] Ekman, P. and Oster, H. 1979. Facial Expressions of Emotion. Annual Review of Psychology, 30, 527-554.
[10] Gebhard, P. 2005. Alma: A Layered Model of Affect. Paper presented at the fourth international joint conference on Autonomous agents and multiagent systems (AAMAS '05), New York, NY, USA.
[11] Li, X., Zhou, H., Song, S., et al. 2005. The Reliability and Validity of the Chinese Version of Abbreviated Pad Emotion Scales. Paper presented at the Affective Computing and Intelligent Interaction (ACII2005).
[12] Lyons, M., Akamatsu, S., Kamachi, M., et al. 1998. Coding Facial Expressions with Gabor Wavelets. Paper presented at the Third IEEE International Conference on Automatic Face and Gesture Recognition (AFGR98).
[13] Mehrabian, A. 1995. Framework for a Comprehensive Description and Measurement of Emotional States. Genetic Social and General Psychology, 121(3), 339-361.
[14] Mehrabian, A. 1996. Pleasure-Arousal-Dominance: A General Framework for Describing and Measuring Individual Differences in Temperament. Current Psychology: Developmental, Learning, Personality, Social, 14(4), 261-292.
[15] Mower, E., Mataric, M. J. and Narayanan, S. 2009. Human Perception of Audio-Visual Synthetic Character Emotion Expression in the Presence of Ambiguous and Conflicting Information. IEEE Transaction on Multimedia, 11(5), 843-855.
[16] Pan, Y. C., Xu, M. X., et al. 2006. Emotion-detecting Based Model Selection for Emotional Speech Recognition. IMACS Multiconference on Computational Engineering in Systems Applications.
[17] Pelachaud, C., Magno-Caldognetto, E., Zmarich, C., et al. 2001. An Approach to an Italian Talking Head. Paper presented at the Eurospeech' 01, Aalborg, Danemark.
[18] Picard, R. W. 1997. Affective Computing: MIT Press.
[19] Raouzaiou, A., Tsapatsoulis, N., Karpouzis, K., et al. 2002. Parameterized Facial Expression Synthesis Based on Mpeg-4. EURASIP Journal on Applied Signal Processing, 10, 1021-1038.
[20] Schröder, M. 2004. Dimensional Emotion Representation as a Basis for Speech Synthesis with Non-Extreme Emotions. Paper presented at the Affective Dialogue Systems, Kloster Irsee, Germany.
[21] Schröder, M. 2006. Expressing Degree of Activation in Synthetic Speech. IEEE Transactions on Audio, Speech, and Language Processing, 14(4), 1128-1136.
[22] Tang, H. and Huang, T. S. 2008. Mpeg4 Performance-Driven Avatar Via Robust Facial Motion Tracking. Paper presented at the IEEE International Conference on Image Processing (ICIP2008), San Diego, CA, United states.
[23] Tao, L. M., Liu, Y., Fu, X. L., et al. 2008. A Computational Study on Pad Based Emotional State Model. Paper presented at the 26th Computer-Human Interaction Conference.
[24] Wang, Z. M. 2002. Research on Chinese Viseme Modeling and Visual Speech. Doctorial Dissertation of Tsinghua University.
[25] Wu, Z., Meng, H. M., Yang, H., et al. 2009. Modeling the Expressivity of Input Text Semantics for Chinese Text-to-Speech Synthesis in a Spoken Dialog System. IEEE Transactions on Audio, Speech, and Language Processing, 17(8), 1567-1576.
[26] Yang, H., Meng, H. M. and Cai, L. 2006. Modeling the Acoustic Correlates of Expressive Elements in Text Genres for Expressive Text-to-Speech Synthesis. Paper presented at the INTERSPEECH 2006 and 9th International Conference on Spoken Language Processing(ICSLP), Pittsburgh, PA, United states.
[27] ZHANG, L., AI, H., XIN, S., et al. 2005. Robust Face Alignment Based on Local Texture Classifiers. Paper presented at the IEEE International Conference on Image Processing (ICIP 2005).
[28] Zhang, S., Wu, Z., Meng, H. M., et al. 2007. Facial Expression Synthesis Using Pad Emotional Parameters for a Chinese Expressive Avatar. Paper presented at the Lecture Notes in Computer Science, Lisbon, Portugal.