

國立交通大學

電控工程研究所

博士論文

機器人情感模型及情感辨識設計
Design of Robotic Emotion Model and
Human Emotion Recognition



研究生：韓孟儒

指導教授：宋開泰 博士

中華民國一百零二年一月

機器人情感模型及情感辨識設計
Design of Robotic Emotion Model and
Human Emotion Recognition

研究生：韓孟儒

Student: Meng-Ju Han

指導教授：宋開泰 博士

Advisor: Dr. Kai-Tai Song



A Dissertation

Submitted to Institute of Electrical Control Engineering
College of Electrical and Computer Engineering
National Chiao Tung University
in Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy
in
Electrical Control Engineering
January 2013
Hsinchu, Taiwan, Republic of China

中華民國一百零二年一月

機器人情感模型及情感辨識設計

學生:韓孟儒

指導教授:宋開泰 博士

國立交通大學電控工程研究所

摘要

本論文之主旨在研究機器人情感模型(emotion model)及其互動設計，文中提出一種擬人化之心情變遷(mood transition)設計方法，提高機器人與人類作自主情感互動之能力。為使機器人能產生富有類人情感表達之互動行為，本論文提出一個二維的情感模型，同時考慮機器人之情感(emotion)、心情(mood)與人格特性(personality)等狀態，使產生擬人化情感反應。在本設計中，機器人之人格特性模型建立係參考心理學家所提出之五大性格特質(Big Five factor)來達成，而機器人之心情變遷所造成之影響，則可藉由此五大人格特質參數來決定。

為能經由連續的互動行為來呈現機器人自主情感狀態，本論文亦提出一種可融合基本情緒行為之方法，來建立不同心情狀態下之行為表達方式。根據上述之心理學研究成果，本研究以模糊 Kohonen 群集網路(fuzzy Kohonen clustering networks)之方法，將人格特性、心情與情緒行為三者整合成一情感模型，使之能具體實現於機器人上。與其他研究相比，具有客觀之學理依據，而非憑藉研究人員本身主觀經驗來做假設。

在情感辨識方面，本論文提出結合影像與聲音之雙模情緒辨識以及語音情緒辨識等二種方法，使機器人可辨識使用者之情緒狀態。在雙模情緒辨識之設計中，論文中提出基於支持向量機(support vector machine)之分類特性與機率策略，用以決定二種特徵資料

之融合權重(fusing weights)。融合權重係根據待測資料與切割平面之距離，以及學習樣本之標準差所決定。而在分類階段，融合權重較高之特徵所辨識之結果，將成為最後系統辨識之結果。此外，在語音情緒辨識之設計中，本論文提出採用聲音訊號進行處理與分類。首先，在預處理時先將語音訊號進行端點偵測(end-point detection)以取得音框所在位置，而後再以統計方式將能量計算成特徵之型態，並以費雪線性辨別分析法(Fisher's linear discriminant analysis)來增強辨識率。

本論文寫作基於 DSP 之影像、語音處理系統驗證所發展的辨識方法，並整合至機器人上展示與人情感互動的功能。為了評估所開發之情感模型，文中並建立一人臉模擬器展示情緒表情之變化。為了解所提方法對於使用者之感受，本研究透過觀察人臉模擬器對使用者之情感表達狀況，以問卷調查方式來作評估。評估結果顯示，受訪者之感受與原設計目標相符。



Design of Robotic Emotion Model and Human Emotion Recognition

Student: Meng-Ju Han

Advisor: Dr. Kai-Tai Song

Institute of Electrical Control Engineering
National Chiao Tung University

ABSTRACT

This thesis aims to develop a robotic emotion model and mood transition method for autonomous emotional interaction with human. A two-dimensional (2-D) emotional model is proposed to combine robotic emotion, mood and personality in order to generate emotional behaviors. In this design, the robot personality is programmed by adjusting the *big five factors* referred from psychology. Using *Big Five* personality traits, the influence factors of robot mood transition are analyzed.

A method to fuse basic robotic emotional behaviors is proposed in this work in order to manifest robotic emotional states via continuous facial expressions. Through reference psychological results, we developed the relationships of personality vs. mood transition for robotic emotion generation. Based on these relationships, personality, mood transition and emotional behaviors have been integrated into the robotic emotion model. Comparing with existing models, the proposed method has the merit of having a theoretical basis to support the human-robot interaction design.

In order to recognize the user's emotional state, both bimodal emotion recognition and speech-signal-based emotion recognition methods are studied. In the design of the bimodal

emotion recognition system, a novel probabilistic strategy has been proposed for a classification design to determine statistically suitable fusing weights for two feature modalities. The fusion weights are selected by the distance between test data and the classification hyperplane and the standard deviation of training samples. In the latter bimodal SVM classification, the recognition result with higher weight is selected.

In the design of the proposed speech-signal-based emotion recognition method, the proposed method uses voice signal processing and classification. Firstly, end-point detection and frame setting are accomplished in the pre-processing stage. Then, the statistical features of the energy contour are computed. Fisher's linear discriminant analysis (FLDA) is used to enhance the recognition rate.

In this thesis, the proposed emotion recognition methods have been implemented on a DSP-based system in order to demonstrate the functionality of human-robot interaction. We have realized an artificial face simulator to show the effectiveness of the proposed methods. Questionnaire surveys have been carried out to evaluate the effectiveness of the proposed emotional model by observing robotic responses to user's emotional expressions. Evaluation results show that the feelings of the testers coincide with the original design.

誌謝

一路走來，由衷感謝我的指導教授宋開泰博士，感謝他多年來在我多次感到氣餒之時不斷給予鼓勵，使我終能順利到站，展開人生新的旅途；另一方面，在專業及論文寫作上的指導，不厭其煩的給我建議及修正方向，使我受益良多，也讓本論文得以順利完成。亦感謝論文口試委員一傅立成教授、王文俊教授、蔡清池教授、胡竹生教授、莊仁輝教授，對於本論文的建議與指引，強化本論文的嚴整性與可讀性。

感謝實驗室的夥伴嘉豪、晉懷、濬尉及仕傑在實務驗證時所提供的協助，也感謝學長戴任詔博士、蔡奇謚博士和博士班學弟妹巧敏、信毅對本論文的建議與討論，同時亦感謝過往相互鼓勵的碩士班學弟妹崇民、松峙、俊璋、富聖、振暘、煥坤、舒涵、科棟、奕廷、哲豪、仕晟、建宏、上峻、章宏等在生活上所帶來的樂趣。

另外，特別感謝我的父母，由於他們辛苦栽培，在生活上給予我細心地關愛與照料，使我才得以順利完成此論文；也感謝我妻子在我身邊的全力支持，在我最無助及失意的時候給予背後安定的力量。願能貢獻一己棉薄之力，成就有益家庭、社會之能事。

Contents

摘要	i
Abstract	iii
誌謝	v
Contents	vi
List of Figures	ix
List of Tables	xii
Chapter 1 Introduction	1
1.1 Motivation	1
1.2 Literature Survey	2
1.3 Research Objectives and Contributions	7
1.4 Organization of the Thesis	8
Chapter 2 Robotic Emotion Model and Emotional State Generation	10
2.1 Robotic Mood Model and Mood Transition	12
2.1.1 Robotic Mood Model	13
2.1.2 Robot Personality	14
2.1.3 Facial Expressions in Two-Dimensional Mood Space	16
2.1.4 Robotic Mood State Generation	16
2.2 Emotional Behavior Generation	18
2.2.1 Rule Table for Behavior Fusion	20
2.2.2 Evaluation of Fusion Weight Generation Scheme	22
2.2.3 Animation of Artificial Face Simulator	26
2.3 Summary	29
Chapter 3 Human Emotion Recognition	30
3.1 Bimodal Information Fusion Algorithm	30

3.1.1	Facial Image Processing	32
3.1.2	Speech Signal Processing	36
3.1.3	Bimodal Information Fusion Algorithm	37
3.1.4	Hierarchical SVM Classifiers	41
3.2	Speech-signal-based Emotion Recognition	42
3.2.1	Speech Signal Pre-processing.....	43
3.2.2	Feature Extraction.....	48
3.2.3	Emotional State Classification.....	50
3.2.4	Implementation of the Emotion Recognition Embedded System	53
3.3	Summary.....	55
Chapter 4	Experimental Results	56
4.1	Experimental Results of Robotic Emotion Generation	56
4.1.1	Experiments on an Anthropomorphic Robotic Head.....	56
4.1.2	Experimental Setup for the Artificial Face Simulator	58
4.1.3	Evaluation of Robotic Mood Transition Due to Individual Personality.....	60
4.1.4	Evaluation of Emotional Interaction Scheme.....	65
4.2	Experiments on Bimodal Information Fusion Algorithm.....	69
4.2.1	Off-line Experimental Results	70
4.2.2	On-line Experimental Results.....	71
4.3	Experiments on Speech-signal-based Emotion Recognition.....	73
4.3.1	Experiments Using the Self-built Database.....	73
4.3.2	Experiments with the Entertainment Robot.....	75
4.4	Experiments on Image-based Emotional State Recognition	77
4.5	Summary.....	80
Chapter 5	Conclusions and Future Work	81
5.1	Dissertation Summary	81

5.2 Future Directions	82
Appendix A Evaluation Questionary of Emotional Interaction	83
Bibliography.....	86
Vita	94
Publication List.....	95



List of Figures

Fig. 1-1: Structure of the thesis.	9
Fig. 2-1: Block diagram of the autonomous emotional interaction system (AEIS) for an artificial face.	10
Fig. 2-2: Two-dimensional scaling for facial expressions based on pleasure-displeasure and arousal-sleepiness ratings.	17
Fig. 2-3: The fuzzy-neuro network for fusion weight generation.	19
Fig. 2-4: Fusion weights distribution for seven facial expressions.	25
Fig. 2-5: Illustration of the proposed robotic emotion model.	29
Fig. 3-1: The experimental setup.	31
Fig. 3-2: Block diagram of the robotic audio-visual emotion recognition system.	31
Fig. 3-3: The functional block diagram of facial image processing.	33
Fig. 3-4: Face detection procedure. (a) Original image, (b) Color segmentation and closing operation, (c) Candidate face areas, (d) Final result obtained by attentional cascade. ...	33
Fig. 3-5: Definition of the facial feature points and feature values.	34
Fig. 3-6: Test results of feature extraction of eyes and eyebrows. (a) Binary operation using IOD, (b) Edge detection, (c) AND operation. (d) Extracted feature points.	35
Fig. 3-7: Feature extraction of lips.	36
Fig. 3-8: The functional block diagram of facial image processing.	37
Fig. 3-9: Representing recognition reliability using the distance between test sample and hyperplane.	38
Fig. 3-10: Representing recognition reliability using the standard deviation of training samples. (a) Smaller standard deviation, (b) Larger standard deviation.	39
Fig. 3-11: SVM bimodal recognition procedure.	41

Fig. 3-12: Block diagram of the proposed speech-signal-based emotion recognition system.	42
Fig. 3-13: Energy of a speech signal.	44
Fig. 3-14: Zero-crossing rate of a speech signal.	45
Fig. 3-15: Example of real human speech detection.	46
Fig. 3-16: An example of end-point detection.	46
Fig. 3-17: Frame-signal separation using a Hamming window.	47
Fig. 3-18: Procedure for speech signal extraction in each frame. (a) A frame of original speech signal. (b) Hamming window. (c) Result of original speech signal multiplied by Hamming window.	48
Fig. 3-19: The original time response of the speech signal and the results for feature extraction of the fundamental frequency.	49
Fig. 3-20: Structure of the hierarchical SVM classifier.	52
Fig. 3-21: The TMS320C6416 DSK codec interface.	53
Fig. 3-22: Interaction scenario for a user and the entertainment robot.	54
Fig. 3-23: Control architecture of the entertainment robot.	55
Fig. 4-1: Architecture of the self-built anthropomorphic robotic head.	57
Fig. 4-2: Examples of facial expressions of the robotic head.	57
Fig. 4-3: Interaction scenario of a user and robotic head.	58
Fig. 4-4: Experiment setup: interaction scenario with an artificial face.	59
Fig. 4-5: Robotic mood transition of RobotA.	63
Fig. 4-6: Robotic mood transition of RobotB.	63
Fig. 4-7: Weights variation for RobotA (active trait).	64
Fig. 4-8: Weights variation for RobotB (passive trait).	65
Fig. 4-9: Questionary result of psychological impact.	66
Fig. 4-10: Questionary result of Natural vs. Artificial.	67
Fig. 4-11: Representation of robot personality parameters.	69

Fig. 4-12: Examples of database..... 70

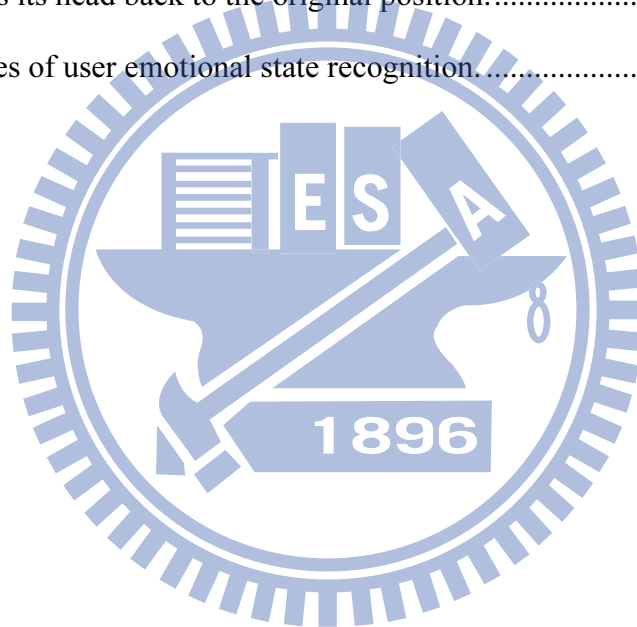
Fig. 4-13: Experimental results of recognition rate for any two emotional categories. 74

Fig. 4-14: Block diagram of the emotional interaction system. 76

Fig. 4-15: Interactive response of the robot as the user says, “I am angry!” (a) The robot puts down its hands to portray fear. (b) The robot continues to put down its hands to the lowest position. (c) The robot raises its hands back to the original position..... 76

Fig. 4-16: Interactive response of the robot, when the user speaks in a surprised tone. (a) The robot shakes its head to the right. (b) The robot shakes its head to the left. (c) The robot puts its head back to the original position..... 77

Fig. 4-17: Examples of user emotional state recognition..... 79



List of Tables

Table 2-1: Big five model of personality.	15
Table 2-2: Rule table for interactive emotional behavior generation.	20
Table 2-3: Basic facial expressions with various weights executed in the simulator.	27
Table 2-4: Linear combined facial expressions with various weights on the simulator.	28
Table 3-1: The description of facial feature values.	35
Table 3-2: The description of speech feature values.	38
Table 3-3: The description of speech feature values.	49
Table 4-1: List of the conversation dialogue and corresponding subject's facial expressions.	59
Table 4-2: Regulated user emotion intensity of conversion sentence 1 and 2.	60
Table 4-3: Definition of personality scales using Big Five factors.	60
Table 4-4: Facial expressions for the RobotA.	62
Table 4-5: Facial expressions for the RobotB.	62
Table 4-6: Estimation of personality parameters by questionnaire survey.	68
Table 4-7: Standard deviation of questionnaire results.	68
Table 4-8: Experimental results using speech features.	70
Table 4-9: Experimental results using image features.	71
Table 4-10: Experimental results using information fusion.	71
Table 4-11: On-line experimental results using only image features.	72
Table 4-12: On-line experimental results using information fusion.	72
Table 4-13: Meaning of sentence content for five emotional categories.	73
Table 4-14: Experimental results of recognizing five emotional categories.	75
Table 4-15: Test result of emotion state recognition.	79

Chapter 1

Introduction

1.1 Motivation

In recent years, many useful domestic and service robots, including museum guide robots, personal companion robots and entertainment robots have been developed for various applications [1]. It has been forecasted that edutainment and personal robots will be very attractive products in the near future [2-3]. One of the most interesting features of intelligent service robots is their human-centered functions. Actually, intelligent interaction with a user is a key feature for service robots in healthcare, companion and entertainment applications. For a robot to engage in friendly interaction with human, the function of emotional expression will play an important role in many real-life application scenarios. However, it is known that to make a robot behave human-like emotional expressions is still a challenge in robot design.

On the other hand, the ability to recognize a user's emotion is also important in human-robot interaction applications. The emotion communicator, Kotohana, developed by NEC [4] is a successful example of vocal emotion recognition. Kotohana is a flower-shaped terminal equipped with Light Emitting Diodes (LEDs). It can recognize a visitor's emotional speech and respond with a color display to convey the interaction. The terminal responds in a lively manner to the detected emotional state, via color variation in the flower. For human beings, facial expression and voice reveal a person's emotion most. They also provide important communicative cues during social interaction. A robotic emotion recognition system will enhance the interaction between human and robot in a natural manner. Base on the above discussion, it is observed that a proper emotion model is desirable in robotic

emotional behavior generation. This also motivates us to investigate mood transition algorithms based on physiological findings for humans.

1.2 Literature Survey

Methodologies for developing emotional robotic behaviors have drawn much attention in robotic research community [5]. Breazeal *et al.* [6] presented the sociable robot Leonardo, which has an expressive face capable of near human-level expression and possesses a binocular vision system to recognize human facial features. The humanoid robot Nexi [7] demonstrated a wide range of facial expressions to communicate with people. Wu *et al.* [8] explored the process of self-guided learning of realistic facial expression by a robotic head. Mavridis *et al.* [9-10] developed the Arabic-language conversational android robot; it can become an exciting educational or persuasive robot in practical use. Hashimoto *et al.* [11-12] developed a reception robot SAYA to realize realistic speaking and natural interactive behaviors with six typical facial expressions. In [13], a singer robot EveR-2 is able to acquire visual and speech information, while expressing facial emotion during performing robotic singing. For some application scenarios such as persuasive robotics [14] or longer-term human-robot interaction [15], interactive facial expression has been demonstrated to be very useful.

There have been increasing interests in the study of robotic emotion generation schemes in order to give a robot more human-like behaviors. Reported approaches to emotional robot design often adopted results from psychology in order to design robot behaviors to mimic human beings. Miwa *et al.* proposed a mental model to build the robotic emotional state from external sensory inputs [16-17]. Duhaut [18] presented a computational model which includes emotion and personality in the robotic behaviors. The TAME (Traits, Attributes, Moods, and Emotions) framework proposed by Moshkina *et al.* gives a model of time-varying affective response for humanoid robots [19]. Itoh *et al.* [20] proposed an emotion generation model

which can assess the robot's individuality and internal state through mood transitions. Their experiments showed that the robot could provide more human-like communications to users based on the emotional model. Banik *et al.* [21] demonstrated an emotion-based task sharing approach to a cooperative multi-agent robotic system. Their approach can give a robot a kind of personality through accumulation of past emotional experience. Park *et al.* [22] developed a hybrid emotion generation architecture. They proposed a robot personality model based on human personality factors to generate robotic interactions. Kim *et al.* [23] utilized the probability-based computational algorithm to develop the cognitive appraisal theory for designing artificial emotion generation systems. Their method was applied to a sample of interactive tasks and led to a more positive human-robot interaction experience. In order to allow a robot to express complex emotion, Lee *et al.* [24] proposed a general behavior generation procedure for emotional robots. It features behavior combination functions to express complex and gradational emotions. In [25], a design of autonomous robotic facial expression generation is presented.

Previous related works show abundant tools for designing emotional robots. It is observed, however, that a proper mood state transition plays a key role in robotic emotional behavior generation. Robotic mood transition from current to next mood state directly influences the interaction behavior of robot and also a user's feeling to the robot. Most existing models treat mood transition by simple and intuitive representations. These representations lack a theoretical basis to support the assumptions for their mood state transition design. This motivated us to investigate a human-like mood transition model for a robot by adopting well-studied mood state conversion criteria from psychological findings. The transition among mood states would become smoother and thus might enable a robot to respond with more natural emotional expressions. We further combine personality into the robotic mood model to represent the trait of individual robot.

In order to manifest emotional intelligence of a robot, responsive interaction behaviors

need to be designed. The relationship between mood states and responding behavior of a robot should not be a fixed, one-to-one relation. A continuous robotic facial expression would be more interesting and natural to manifest the mood state transition. Instead of being arbitrary defined, the relationships between robot emotional behaviors (e.g. in a form of facial expression) and mood state can be modeled from psychological analysis and utilized to build the interaction patterns in the design of expressive behaviors.

To respond to a user sensationally, a robot needs first to understand the user's emotion state. There are many approaches to building-up a robotic emotion recognition system. The majority of studies focus on image-based facial expression recognition [26-27]. Approaches using speech signal processing have also been investigated for sociable robotics [28-29]. Recently, there has been an increasing interest in audio-visual biometrics [30]. The combination of audio and visual information provides more reliable estimate of emotional states. The complementary relationship of these two modalities makes a recognition decision more accurate than using only a single modality. De Silva *et al.* [31] proposed to process audio and visual data separately. They have shown that some emotional states are visual dominant and some are audio dominant. They exploited this observation to recognize emotion efficiently by assigning a weight matrix to each emotion state. In [32], De Silva combined the audio and visual features using a rule-based technique to obtain improved recognition results. Rather simple rules were used in his design. For example, a rule is such that if a sample has been classified as certain emotion by both audio and visual processing methods, then the final result is that emotion. If samples have been classified differently by audio and visual analyses, the dominant mode is used as the emotion decision. Negative emotional expressions, such as anger and sadness, were assigned to be audio dominant, while joy and surprise were assigned to be visual dominant. Go *et al.* [33] combined audio and visual features directly to recognize different emotions using a neural network classifier. However, they did not give comparative experimental results between using bimodal and single modality. Wang *et al.* [34] proposed to

use cascade audio and visual feature data to classify variant emotions. They built one-against-all (OAA) linear discriminant analysis (LDA) classifiers for each emotion state and computed the probability of each emotion type. They set two rules in the decision module with several multi-class classifiers to determine the most possible emotion.

It is clear that audio and visual information are related to each other. In many situations, they offer complementary effect for recognizing emotion states. However, current related works do not deal with the robustness of emotion classification of such bimodal systems. Existing approaches to combining audio-visual information employ some straightforward and simple rules. The reliability of individual modality is not taken into consideration in the decision stage. One solution to this problem is that the classifier output is a calibrated posterior probability $P(\text{class}|\text{input})$ to perform post-processing. Platt [35] proposed a probabilistic support vector machine (SVM) to produce a calibrated posterior probability. The method trained parameters of a sigmoid function to map SVM outputs into probabilities. Although this method is valid to estimate the posterior probability, a sigmoid function cannot represent all the modals of SVM outputs. In this study, we develop a new method for reliable emotion recognition utilizing audio-visual information. We emphasize the decision mechanism of the recognition procedure when fusing visual and audio information. By setting proper weights to each modality based on their recognition reliability, a more accurate recognition decision can be obtained.

In the design of emotion recognition systems that use speech signals, most methods employ vocal features, including the statistics of fundamental frequency, energy contour, duration of silence and voice quality [36]. In order to improve the recognition rate when more than two emotional categories are to be classified, Nwe *et al.* [37] used short time log frequency power coefficients (LFPC) to represent speech signals and a discrete hidden Markov model (HMM) as the classifier. Based on the assumption that the pitch contour has a Gaussian distribution, Hyun *et al.* [38] proposed a Bayesian classifier for emotion recognition

in speech information. They reported that the zero value of a pitch contour causes errors in the Gaussian distribution and proposed a non-zero-pitch method for speech feature extraction. Pao and Chen [39] used 16-bit linear predictive coding (LPC) and twenty Mel-frequency cepstral coefficients (MFCC) to identify the emotional state of a speaker. Five emotional categories were classified using the minimum-distance method and the nearest mean classifier. Neiberg *et al.* [40] modeled the pitch feature by using standard MFCC and MFCC-low, which is calculated between 20 and 300 Hz. Their experiments showed that MFCC-low outperformed the pitch features.

You *et al.* [41] indicated that the effectiveness of principal component analysis (PCA) and linear discriminant analysis (LDA) is limited by their underlying assumption that the data is in a linear subspace. For nonlinear structures, these methods fail to detect the real number of degrees of freedom of the data, so they proposed the method of Lipschitz embedding [42]. The method is not limited by an underlying assumption that the data belong to a linear subspace, so it can analyze the speech signal in more practical situations. Schuller *et al.* [29] considered an initially large set of more than 200 features; they ranked the statistical features according to LDA results and selected important features by ranking statistical features. Chuang and Wu [43] showed that the contours of the fundamental frequency and energy are not smooth. In order to remove discontinuities in the contour, they used the Legendre polynomial technique to smooth the contours of these features. Their feature extraction procedures firstly estimated the fundamental frequency, energy, formant 1 (F1) and zero-crossing rate. From these four features, the feature values are transformed to 33 statistical features. PCA was then used to select 14 principal components from these 33 statistical features, for the analysis of emotional speech. Busso *et al.* [44] indicated that gross fundamental frequency contour statistics, such as mean, maximum, minimum and range, are more emotionally prominent than features that describe the shape of the fundamental frequency. Using psychoacoustic harmony perception from music theory, Yang *et al.* [45]

proposed a new set of harmony features for speech emotion recognition. They reported improved recognition by the use of harmony parameters and state of the art features.

For robotics applications, Li *et al.* [46] developed a prototype chatting robot, which can communicate with a user in a speech dialogue. The recognition of the speech emotion of a specific person was successful for two emotional categories. Kim *et al.* [47] focused on speech emotion recognition for a thinking robot. They proposed a speaker-independent feature, namely the ratio of a spectral flatness measure to a spectral center, to solve the problem of diverse interactive users. Similarly, Park *et al.* [48] also studied the issue of service robots interacting with diverse users who are in various emotional states. Acoustically similar characteristics between emotions and variable speaker characteristics, caused by different users' style of speech, may degrade the accuracy of speech emotion recognition. They proposed feature vector classification for speech emotion recognition, to improve performance in service robots.

For practical application, several important problems exist. Firstly, a robust speech signal acquisition system must be built on the front end of the design. It is also required that the robot is equipped with a stand-alone system for realistic human-robot interaction. One of the greatest challenges in emotion recognition for robotic applications is the performance required for nature and daily life environments.

1.3 Research Objectives and Contributions

The objective of this thesis is to develop a robot emotion model in order to interact with people emotionally. A two-dimensional (2-D) emotional model is proposed to represent robot emotion, mood transition and personality in order to generate human-like emotional expressions. In this design, the robot personality is programmed by adjusting the factors of the Five Factors model proposed by psychologists. From Big Five personality traits, the influence factors of robot mood transition are determined.

A method to fuse on basic robotic emotional behaviors is proposed in order to manifest robotic emotional states via continuous facial expressions. An artificial face on a screen is an effective way to evaluate a robot with a human-like appearance. An artificial face simulator has been implemented to show the effectiveness of the proposed methods. Questionnaire surveys have been carried out to evaluate the effectiveness of the proposed method by observing robotic responses to user's emotional expressions. Preliminary experimental results on a robotic head show that the proposed mood state transition scheme appropriately responds to a user's emotional changes in a continuous manner.

The second part of this thesis aims to develop suitable emotion recognition methods for human-robot interaction. A bimodal emotion recognition method was proposed in this thesis. In the design of the bimodal emotion recognition system, a probabilistic strategy has been studied for a support vector machine (SVM)-based classification design to assign statistically selected fusing weights to two feature modalities. The fusion weights are determined by the distance between test data and the classification hyperplane and the standard deviation of training samples. In the latter bimodal SVM classification, the recognition result with higher weight is selected.

In the design of the speech-signal-based emotion recognition method, speech signals are used to recognize several basic human emotional states. The proposed method uses voice signal processing and classification. In order to determine the effectiveness of emotional human-robot interaction, an embedded system was constructed and integrated with a self-built entertainment robot.

1.4 Organization of the Thesis

Figure 1-1 shows the organization of this thesis. In Chapter 2, a novel robotic emotion generation system is developed based-on mood transition model. A robotic mood state

generation algorithm is proposed using a two-dimensional emotional model. An interactive emotional behaviors generation is then proposed to generate an unlimited number of emotional expressions by fusing seven basic facial expressions. In Chapter 3, several human emotion recognition methods are developed to provide user's emotional state. Here bimodal information fusion algorithm and speech-signal-based emotion recognition method are proposed for human-robot interaction. Simulation and experimental results of the proposed robotic emotion generation system and the proposed human emotion recognition methods are reported and discussed in Chapter 4. Chapter 5 concludes the contributions of this work and provides the recommendations for future research.

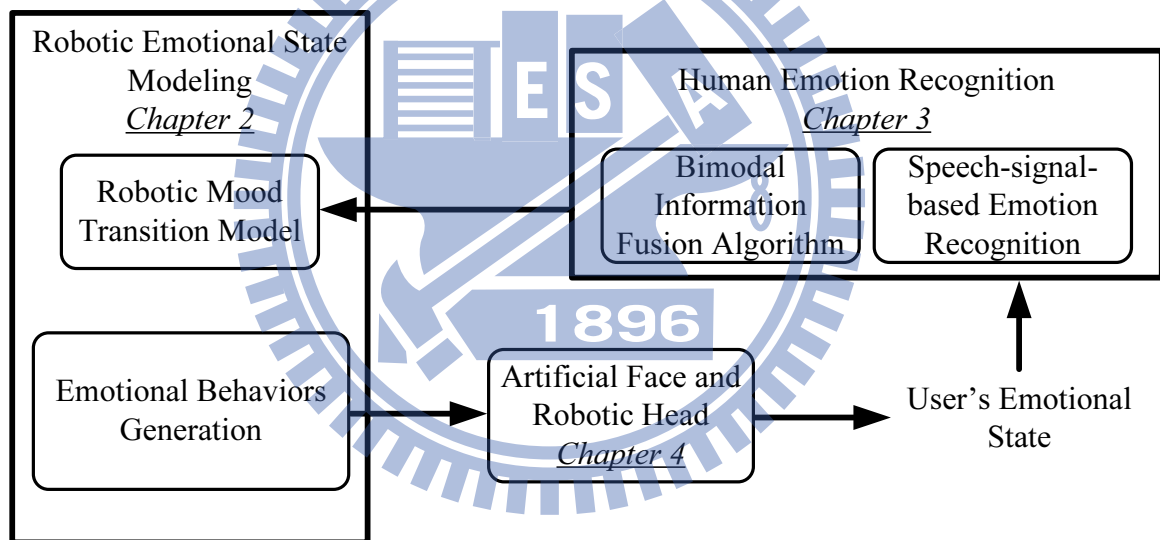


Fig. 1-1: Structure of the thesis.

Chapter 2

Robotic Emotion Model and Emotional State Generation

Figure 2-1 shows the block diagram of the proposed autonomous emotional interaction system (AEIS). Taking a robotic facial expression as the emotion behavior, the robotic interaction is expected not only to react to user's emotional state, but also to reflect the mood state of the robot itself. We attempt to integrate three modules to construct the AEIS, namely, user emotional state recognizer, robotic mood state generator and emotional behavior decision maker. An artificial face is employed to demonstrate the effectiveness of the design. A camera is provided to capture the user's face in front of the robot. The acquired images are sent to the image processing stage for emotional state recognition [49]. The user emotional state recognizer is responsible for obtaining user's emotional state and its intensity. In this design, user's emotional state at instant k (UE_k^n) is recognized and represented as a vector of four

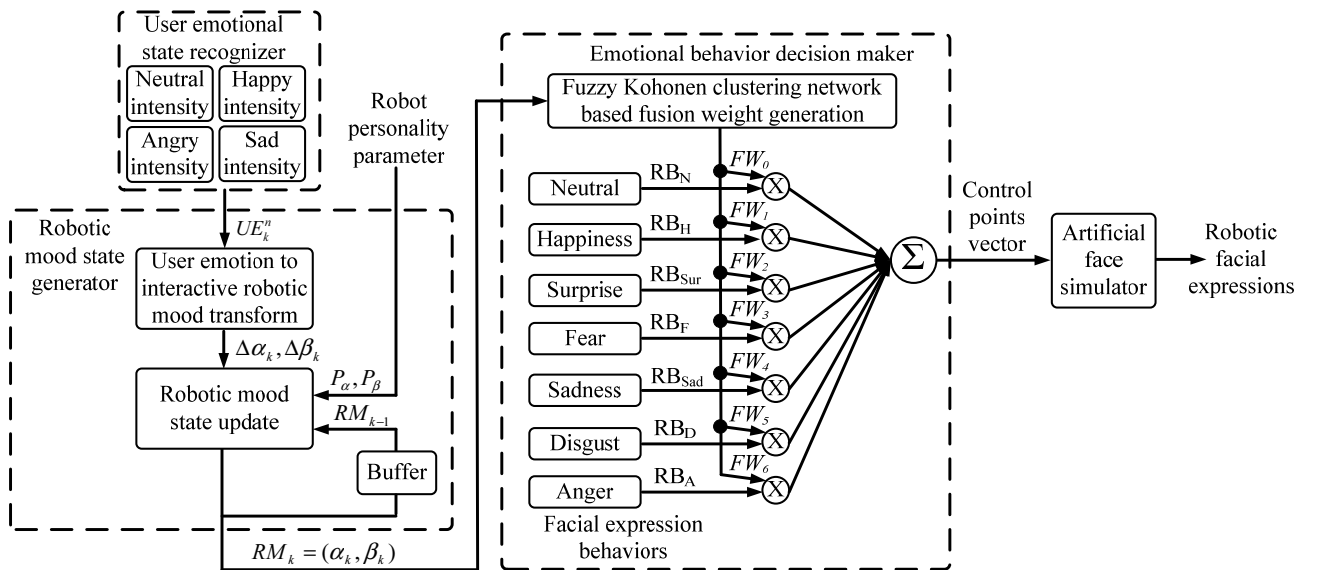


Fig. 2-1: Block diagram of the autonomous emotional interaction system (AEIS).

emotional intensities: *neutral* ($ue_{N,k}^n$), *happy* ($ue_{H,k}^n$), *angry* ($ue_{A,k}^n$) and *sad* ($ue_{S,k}^n$). Several existent emotional intensity estimation methods [50-53] provide effective tools to recognize the intensity of human's emotional state. Their results can be applied and combined into the AEIS.

In this work, an image-based emotional intensity recognition module (see 4.5) has been designed and implemented for current design of AEIS. The recognized emotional intensity consists of basic emotional categories at each sampling instant and is represented by a value between 0 and 1. These intensities are sent to the robotic mood state generator. Moreover, other emotion recognition modalities and methods (e.g. emotional speech recognition) can also be input to AEIS, only the recognized emotional states contain intensity values between 0 and 1.

In the robotic mood state generator, the recognized user's emotional intensities are transformed into interactive robotic mood variables represented by $(\Delta\alpha_k, \Delta\beta_k)$ (see 2.1.1 for detailed description). These two variables represent the way that user's emotional state influences the robotic mood state transition. Furthermore, the robotic emotional behavior depends not only on user's emotional state, but also on robot personality and previous mood state. Therefore the proposed method takes into account the interactive robotic mood variables $(\Delta\alpha_k, \Delta\beta_k)$, previous robotic mood state (RM_{k-1}) and robot personality parameters (P_α, P_β) to compute current robotic mood state (RM_k) (see 2.1.4). Note that the previous robotic mood state (RM_{k-1}) is temporary stored in a buffer. In this work, the current robotic mood state is represented as a point in the two-dimensional (2D) emotional plane. Furthermore, robotic personality parameters are created to describe the distinct human-like personality of a robot. Based on the current robotic mood state, the emotional behavior decision unit autonomously generates suitable robot behavior in response to the user's emotion state.

For robotic emotional behavior generation, in response to recognized user's emotional

intensities, a set of fusion weights ($FW_i, i=0\sim6$) corresponding to each basic emotional behavior are generated by using a fuzzy Kohonen clustering network (FKCN) [54] (see 2.2). Similar to human beings, the facial expression of a robotic face is very complex and is difficult to be classified by limited number of categories. In order to demonstrate interaction behaviors similar to that of humans, FKCN is adopted to generate an unlimited number of emotional expressions by fusing seven basic facial expressions. Outputs of FKCN are sent to the artificial face simulator to generate the interactive behaviors (facial expressions in this work). An artificial face has been designed exploiting the method in [55] to demonstrate the facial expressions generated in human-robot interaction. Seven basic facial expressions are simulated, including *neutral, happiness, surprise, fear, sadness, disgust and anger*. The facial expressions are depicted by moving control points determined from Ekman's model [56]. In the practical interaction scenario, each expression can be generated with different proportions of seven basic facial expressions. The actual facial expression of the robot is generated by summation of each behavior output multiplied by its corresponding fusion weight. Therefore, more subtle emotional expressions can be generated as desired. Detailed design of the proposed robotic mood transition model, emotional behavior generation and image-based emotional state recognition will be described in the following sections.

2.1 Robotic Mood Model and Mood Transition

Emotion is a complex psychological experience of an individual's state of mind as interacting with people or environmental influences. For humans, emotion involves "physiological arousal, expressive behaviors, and conscious experience" [57]. Emotional interaction behavior is associated with mood, temperament, personality, disposition, and motivation. In this study, the emotion for robotic behavior is simplified to association with mood and personality. We apply the concept that emotional behavior is controlled by current emotional state and mood, while the mood is influenced by personality. In this thesis, a novel

robotic mood state transition method is proposed for a given human-like personality. Furthermore, the corresponding interaction behavior will be generated autonomously for a determined mood state.

2.1.1 Robotic Mood Model

A simple way to develop robotic emotional behaviors that can interact with people is to allow a robot to respond emotional behaviors by mimicking humans. In human-robot emotional interaction, users' emotional expressions can be treated as trigger inputs to drive the robotic mood transition. Furthermore, transition of robotic mood depends not only on user's emotional states, but also on the robot mood and personality of itself. For a robot to interact with several individuals or a group of people, users' current (at instant k) emotional intensities (UE_k^n) are sampled and transformed into interactive mood variables $\Delta\alpha_k$ and $\Delta\beta_k$ to represent how user's emotional state influences the variation of robotic mood state transition.

From the experience of emotional interaction of human beings, a user's neutral intensity, for instance, usually affects the arousal and sleepiness mood variation directly. Thus, the robotic mood state tends to arousal while the user's neutral intensity is low. Similarly, the user's happiness, anger and sadness intensities affect the pleasure-displeasure axes. Thus, user's happy intensity will lead robotic mood into pleasure. On the other hand, the robotic mood state behaves more displeasure while user's angry and sad intensities are high. Based on the above observations, a straightforward case is designed for the interactive robotic mood variables ($\Delta\alpha_k, \Delta\beta_k$), which represent the reaction from current users' emotional intensities on the pleasure-arousal plane, such that:

$$\Delta\alpha_k = \frac{1}{N_s} \sum_{n=1}^{N_s} [ue_{H,k}^n - (ue_{A,k}^n + ue_{S,k}^n) / 2] \quad (2.1)$$

$$\Delta\beta_k = \frac{1}{N_s} \sum_{n=1}^{N_s} 2 \cdot (0.5 - ue_{N,k}^n) \quad (2.2)$$

$$UE_k^n = \begin{bmatrix} ue_{N,k}^n \\ ue_{H,k}^n \\ ue_{A,k}^n \\ ue_{S,k}^n \end{bmatrix} = \begin{bmatrix} k^{\text{th}} \text{ neutral intensity for user } n \\ k^{\text{th}} \text{ happiness intensity for user } n \\ k^{\text{th}} \text{ anger intensity for user } n \\ k^{\text{th}} \text{ sadness intensity for user } n \end{bmatrix}, \quad (2.3)$$

where N_s denotes the number of users and UE_k^n represents four kinds of the n^{th} user's emotional intensities. By using (2.1)-(2.3), the effect on robotic mood from multiple users' emotional inputs is represented. However, in this work, only one user is considered for better concentrating on the illustration of the proposed model, i.e. $N_s=1$ in the following discussion. It is worth to extend the number of users in the next stage of this study, such that a scenario like the Massachusetts Institute of Technology mood meter [58] can be investigated. Furthermore, the mapping between facial expressions of interacting human and robotic internal state may be modeled in a more sophisticated way. For example, $\Delta\alpha_k$ can be designed as $(ue_{A,k}^i + ue_{S,k}^i)/2 - ue_{H,k}^i$ such that alternative (opposite) responses to a user can be obtained.

2.1.2 Robot Personality

McCrae *et al.* [59] proposed *Big Five* factors (*Five Factor model*) to describe the traits of human personality. *Big Five model* is an empirically based result, not a theory of personality. The *Big Five* factors were created through a statistical procedure, which was used to analyze how ratings of various personality traits are correlated for general humans. Table 2-1 lists the *Big Five* factors and their descriptions [60]. Besides, Mehrabian [61] utilized the *Big Five* factors to represent the pleasure-arousability-dominance (PAD) temperament model. Through linear regression analysis, the scale of each PAD value is estimated by using the *Big Five* factors [62]. These results are summarized as three equations of temperament, which includes pleasure, arousability and dominance.

In this work, we adopted Big Five model to represent the robot personality and determine the mood state transition on a two-dimensional pleasure-arousal plane. Hence only two equations are utilized to represent the relationship between robot personality and

Table 2-1: Big five model of personality.

Factor	Description
Openness	Open mindedness, interest in culture.
Conscientiousness	Organized, persistent in achieving goals.
Extraversion	Preference for and behavior in social situations.
Agreeableness	Interactions with others.
Neuroticism	Tendency to experience negative thoughts.

pleasure-arousal plane. The reason that we utilize this two-dimensional pleasure-arousal plane rather than the three-dimensional PAD model is based on the Russell's study. Russell and Pratt [63] indicated that pleasure and arousal each account for large proportions of variance in the meaning of affect terms, each dimension beyond these two accounted for only a tiny proportion. More importantly, these secondary dimensions became more and more clearly interpretable as cognitive rather than emotional in nature. The secondary dimensions thus appear to be aspects of the cognitive appraisal system that has been suggested for emotions. Here elements of the *Big Five* factors are assigned based on a reasonable realization of Table 2-1. Referring to [61], the robot personality parameters (P_α, P_β) are adopted such that:

$$P_\alpha = 0.21E + 0.59A + 0.19N \quad (2.4)$$

$$P_\beta = 0.15O + 0.3A - 0.57N, \quad (2.5)$$

where O, E, A and N represent the *Big Five* factors of openness, extraversion, agreeableness and neuroticism respectively. Therefore the robot personality parameters (P_α, P_β) are given as the robot personality is known, i.e. O, E, A and N are determined constants. Later we will show that (P_α, P_β) works as the mood transition weightings on pleasure (α axis) and arousal (β axis) plane.

Note that the conscientiousness of *Big Five* factors was not used in this design, because this factor only influences the dominance axis of three-dimensional PAD model. In this study, the pleasure-arousal plane of two-dimensional emotional model was applied, so only four out

of five parameters are used to translate the mood transition weighting from the *Big Five* factors.

2.1.3 Facial Expressions in Two-Dimensional Mood Space

The relationship between mood states and emotional behaviors has been studied by psychologists. Russell and Bullock [64] proposed a two-dimensional scaling on the pleasure-displeasure and arousal-sleepiness axes to model the relationships between the facial expressions and mood state. In this work, the results from [64] are employed to model the relationship between mood state and output emotional behavior. Figure 2-2 illustrates a two-dimensional scaling result for general adult's facial expressions based on pleasure-displeasure and arousal-sleepiness ratings. The scaling result was analyzed by the Guttman-Lingoes smallest space analysis procedure [65]. This two-dimensional scaling procedure provides a geometric representation (stress and orientation) of the relations among the facial expressions by placing them in a space (Euclidean space is used here) of specified dimensionality. Greater similarity between two facial expressions is represented by their closeness in the space. Hence the coordinate in this space can be used to represent the characteristic of each facial expression. As shown in Fig. 2-2, axis α and β represent the amount of pleasure and arousal respectively. Eleven facial expressions are analyzed and located on the plane. The location of each facial expression is represented by a square along with its coordinates. The coordinates of each facial expression is obtained by measuring the location in the figure (interested readers are referred to [64]). The relationship between robotic mood and output behavior, facial expression in this case, is determined.

2.1.4 Robotic Mood State Generation

As mentioned in 2.1.1, both user's current emotional intensity and robot personality affect the robotic mood transition. The way that robot personality affects the mood transition

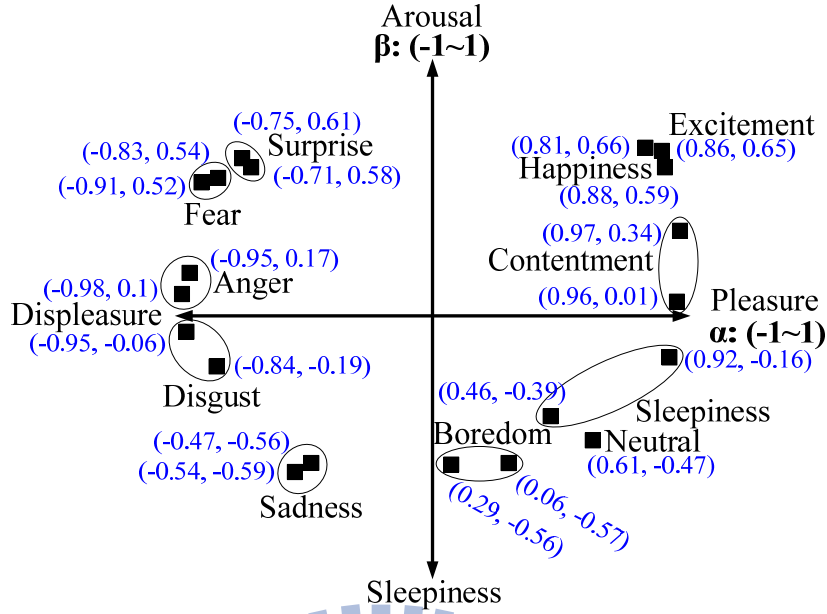


Fig. 2-2: Two-dimensional scaling for facial expressions based on pleasure-displeasure and arousal-sleepiness ratings.

is described by robot personality parameters (P_α, P_β) . As given in 2.1.2, these two parameters act as weighting factors on α and β axis respectively. When P_α and P_β vary, the speed of mood transition in the direction of α and β axes is affected. On the other hand, the interactive mood variables $(\Delta\alpha_k, \Delta\beta_k)$ give the influence of user's emotional intensity on the variation of robotic mood state transition. To reveal the relationship between robot personality and mood transition, we suggest to multiply robot personality parameters (P_α, P_β) with interactive mood variables $(\Delta\alpha_k, \Delta\beta_k)$. This indicates the influence of robotic mood transition from current user's emotional intensity as well as robot personality.

Furthermore, the manifested emotional state is determined not only by current robotic emotional variable but also by previous robotic emotional states. The manifested robotic mood state at sample instant k (RM_k) is calculated such that:

$$RM_k \equiv (\alpha_k, \beta_k) = RM_{k-1} + (P_\alpha \cdot \Delta\alpha_k, P_\beta \cdot \Delta\beta_k), \quad (2.6)$$

where $(\alpha_k, \beta_k) \in [-1, 1]$ represents the coordinates of robotic mood state at sample instant k on pleasure-arousal plane. By using (2.6), the current robotic mood state is determined and

located on emotional plane. Moreover, the mood transition is influenced by personality, which is reflected by the *Big Five* factors. After obtaining the manifested robotic mood state (RM_k), the coordinate of (α_k, β_k) will be mapped onto pleasure-arousal plane, and a suitable corresponding facial expression can be determined, as shown in Fig. 2-2.

2.2 Emotional Behavior Generation

After the robotic mood state is determined by using (2.6), a suitable emotional behavior is expected to respond to the user. In this work, we propose a design based on fuzzy Kohonen clustering network (FKCN) to generate smooth variation of interaction behaviors (facial expressions) as mood state transits gradually.

In this approach, pattern recognition techniques were adopted to generate interactive robotic behaviors [25, 54]. By adopting FKCN, robotic mood state, obtained from (2.6), is mapped to fusion weights of basic robotic emotional behaviors. The output will be a linear combination of weighted basic behaviors. In the current design, the basic facial expression behaviors are *neutral, happiness, surprise, fear, sadness, disgust and anger*, as shown in Fig. 2-1. FKCN is employed to determine the fusion weight of each basic emotional behavior based on the current robotic mood. Figure 2-3 illustrates the structure of the fuzzy-neuro network for fusion weight generation. In the input layer of the network, the robotic mood state (α_k, β_k) is regarded as inputs of FKCN. In the distance layer, the distance between input pattern and each prototype pattern is calculated such that:

$$d_{ij} = \|X_i - P_j\|^2 = (X_i - P_j)^T (X_i - P_j), \quad (2.7)$$

where X_i denotes the input pattern and P_j denotes the j^{th} prototype pattern (see 2.3.2). In this layer, the degree of difference between the current robotic mood state and the prototype pattern is calculated. If the robotic mood state is not similar to the built-in prototype patterns, then the distance will reflect the dissimilarity. The membership layer is provided to map the distance d_{ij} to membership values u_{ij} , and it calculates the similarity degree between the input

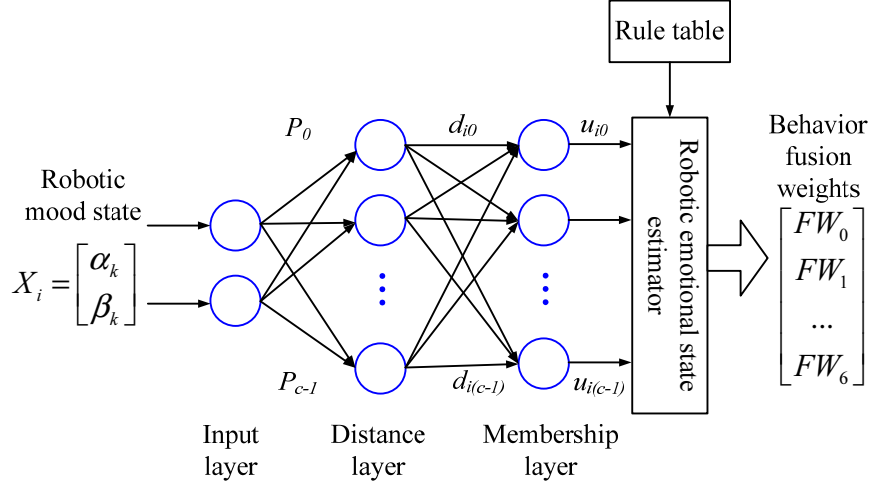


Fig. 2-3: The fuzzy-neuro network for fusion weight generation.

pattern and the prototype patterns. If an input pattern does not match any prototype pattern, then the similarity between the input pattern and each individual prototype pattern is represented by a membership value from 0 to 1. The determination of the membership value is given such that:

$$u_{ij} = \begin{cases} 1 & \text{if } d_{ij} = 0 \\ 0 & \text{if } d_{ik} = 0 \quad (k > 0, j \leq c-1) \end{cases}, \quad (2.8)$$

where c denotes the number of prototype patterns, otherwise,

$$u_{ij} = \left[\sum_{l=0}^{c-1} d_{il} \right]^{-1}. \quad (2.9)$$

Note that the sum of the outputs of the membership layer equals 1. Using the rule table (see later) and the obtained membership values, the current fusion weights ($FW_i, i=0\sim 6$) are determined such that:

$$FW_i = \sum_{j=0}^{c-1} w_{ji} u_{ij}, \quad (2.10)$$

where w_{ji} represents the prototype-pattern weight of i^{th} output behavior. The prototype-pattern weights are designed in a rule table to define basic primitive emotional behaviors corresponding to carefully chosen input states.

2.2.1 Rule Table for Behavior Fusion

In the current design, several representative input emotional states were selected from the two-dimensional model in Fig. 2-2, which gives the relationship between facial expressions and mood states. Each location of facial expression on the mood plane in Fig. 2-2 is used as a prototype pattern for FKCN. Thus, a rule table is constructed accordingly following the structure of FKCN. As shown in Table 2-2, seven basic facial expressions were selected to build the rule table. The IF-part of the rule table is the emotional state of α_k and β_k of the pleasure-arousal space and the THEN-part is the prototype-pattern weight (w_{ji}) of seven basic expressions. For example, the neutral expression in Fig. 2-2 occurs at (0.61, -0.47), which forms the IF-part of the first rule and the prototype pattern for neutral behavior. The THEN part of this rule is the neutral behavior expressed by a vector of prototype-pattern weights (1, 0, 0, 0, 0, 0, 0). The other rules and prototype patterns are set up similarly following the values in Fig. 2-2. Some facial expressions are located at two distinct points on the mood space, both locations are employed, and two rules are set up following the analysis results from psychologist. There are all together 13 rules as shown in Table 2-2. Note that Table 2-2 gives us suitable rules to mimic the behavior of human, since the content of Fig. 2-2 is referenced from psychology results. However, other alternatives and more general rules can

Table 2-2: Rule table for interactive emotional behavior generation.

IF-part prototype patterns			THEN-part weighting						
#j	α_k	β_k	Neutral	Happiness	Surprise	Fear	Sadness	Disgust	Anger
1	0.61	-0.47	1						
2	0.81	0.66		1					
3	0.88	0.59		1					
4	-0.75	0.61			1				
5	-0.71	0.58			1				
6	-0.83	0.54				1			
7	-0.91	0.52				1			
8	-0.47	-0.56					1		
9	-0.54	-0.59					1		
10	-0.95	-0.06						1	
11	-0.84	-0.19						1	
12	-0.95	0.17							1
13	-0.98	0.1							1

also be employed. FKCN works to generalize from these prototype patterns all possible situations (robotic mood state in this case) that may happen to the robot. In the FKCN generalization process, proper fusion weights for the corresponding pattern are calculated. After obtaining the fusion weights of output behaviors from FKCN, the robot's behavior is determined from seven basic facial expressions weighted by their corresponding fusion weights such that:

$$\begin{aligned} \text{Facial Expression} = & RB_N \times FW_0 + RB_H \times FW_1 + RB_{Sur} \times FW_2 + RB_F \times FW_3, \\ & + RB_{Sad} \times FW_4 + RB_D \times FW_5 + RB_A \times FW_6 \end{aligned} \quad (2.11)$$

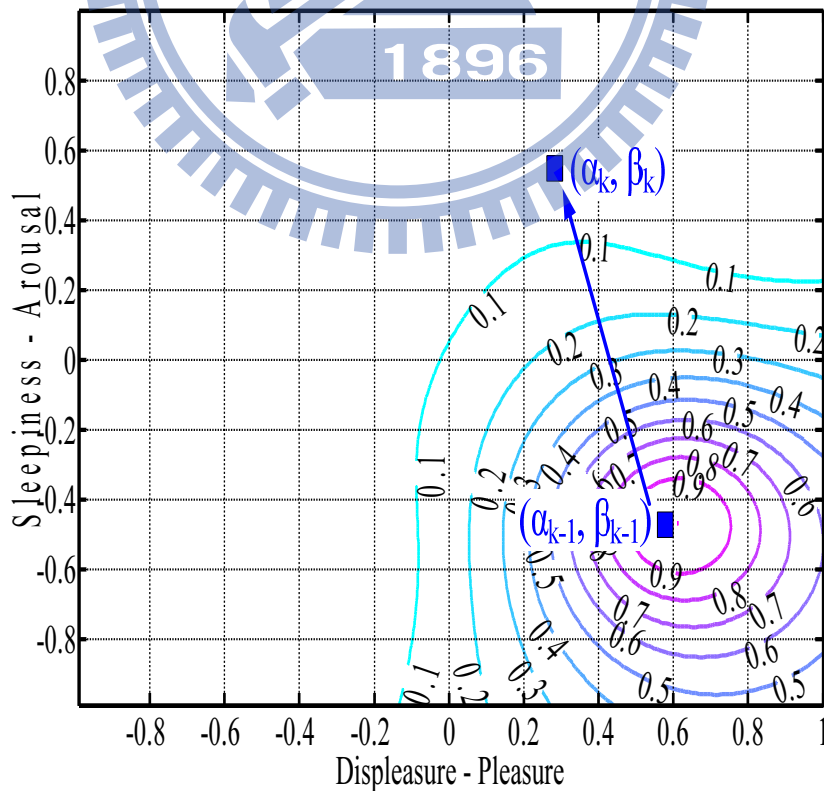
where RB_N , RB_H , RB_{Sur} , RB_F , RB_{Sad} , RB_D , RB_A , represent the seven basic facial expressions of neutral, happiness, surprise, fear, sadness, disgust and anger respectively. It is seen that (2.11) gives us a method to generate facial expressions by combining and weighting the seven basic expressions.

The linear combination of basic facial expressions gives a straightforward yet effective way to express various emotional behaviors. In order to make the combined facial expression to be more consistent with human experience, an evaluation and adjusting procedure was carried out by a panel of students in the lab. The features of seven basic facial expressions were adjusted as distinguished as possible to approach human perception experience. Some results of linear combination are demonstrated using a face expression simulator, please refer to 2.2.3.

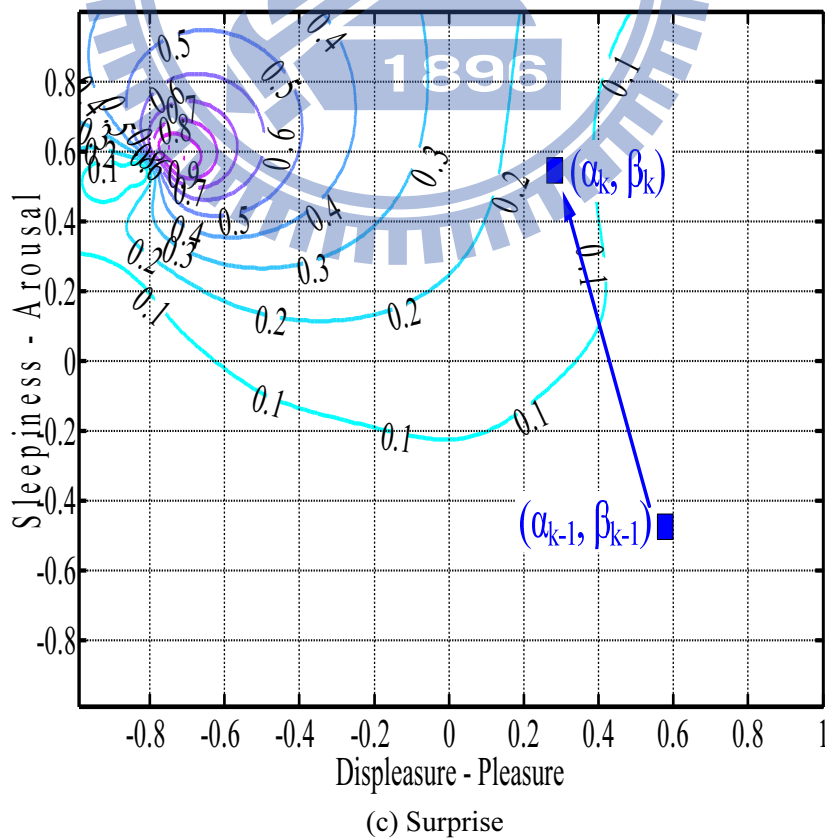
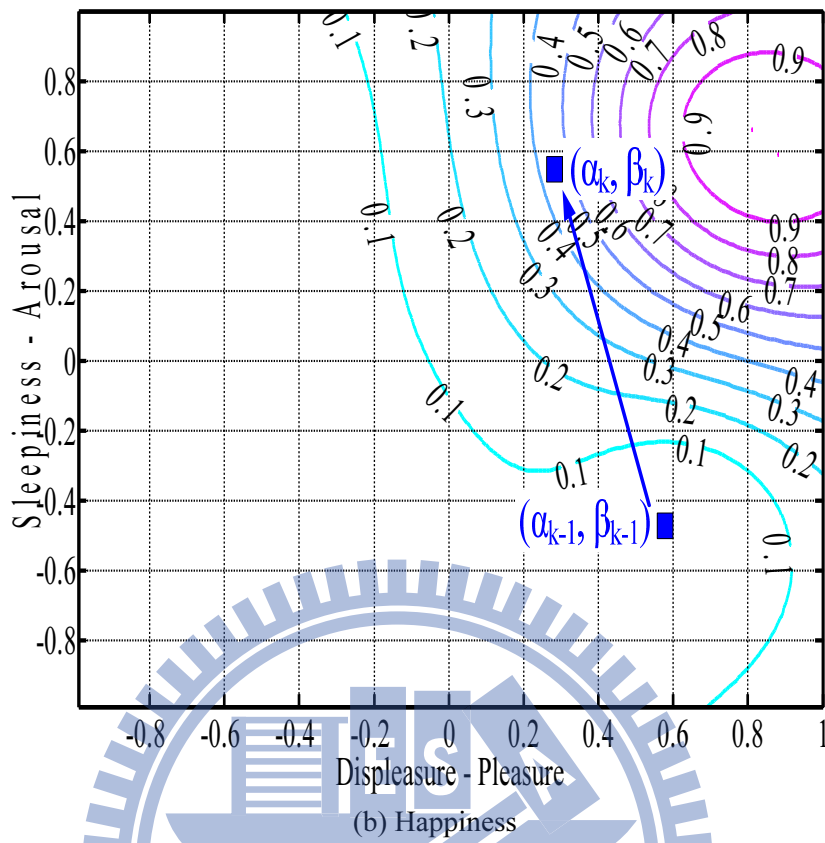
In fact, human emotional expressions are difficult to be represented by a mathematical model or several typical rules. On the other hand, FKCN is very suitable for building up the emotional expressions. The merit of FKCN is its capacity to generalize the results using limited assigned rules (prototypes). Furthermore, dissimilar emotional types can be designed by adjusting the rules. For the artificial face, facial expressions are defined as the variation of control points, which are positions of eyebrow, eye, lips and wrinkles of the artificial face.

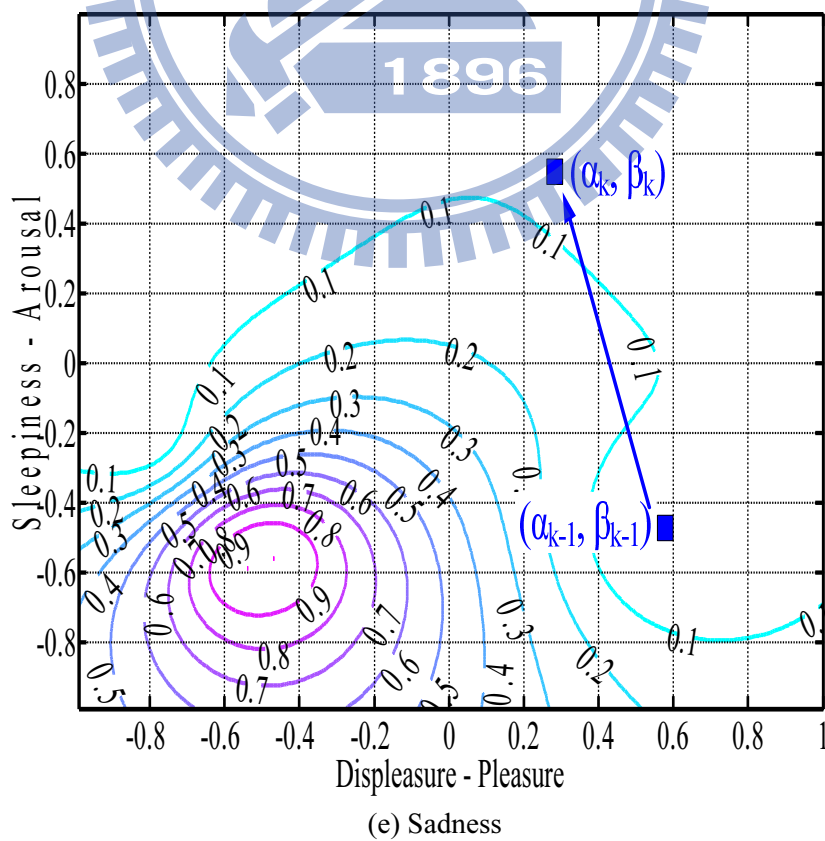
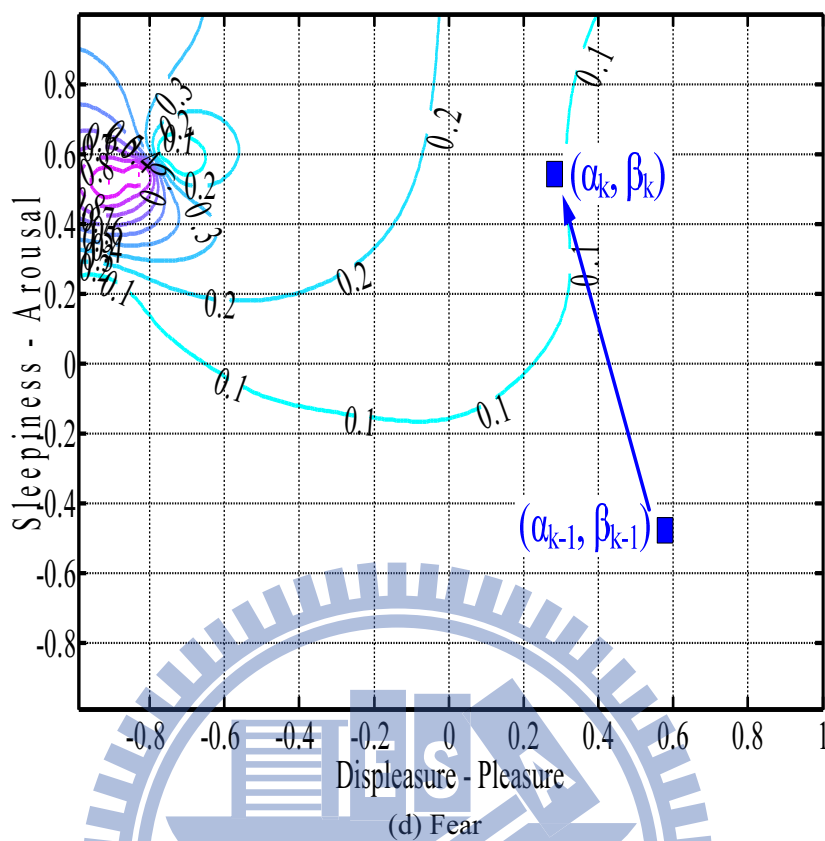
2.2.2 Evaluation of Fusion Weight Generation Scheme

In order to verify the result of fusion-weight generation using FKCN, we applied the rules in Table 2-2 and simulated the weight distribution for various emotional states. The purpose is to evaluate how the proposed FKCN does work to generalize any input emotional state (α_k, β_k) and give a set of output fusion weights corresponding to the input. Figure 2-4 shows the simulation results of weight distribution vs. robotic mood variation on pleasure-arousal plane. In order to check seven fusion weights corresponding to seven basic emotional expressions for a given mood transition from $(\alpha_{k-1}, \beta_{k-1})$ to (α_k, β_k) , the simulation outputs for seven basic emotional expressions are illustrated respectively. The blue squares in Fig. 2-4 indicate the robotic mood transition from $(\alpha_{k-1}, \beta_{k-1})$ to (α_k, β_k) . Every position or point in this two-dimensional mood space has corresponding fusion weights. Figure 2-4(a) shows the weight distribution of neutral expression for the whole robotic mood space. The



(a) Neutral





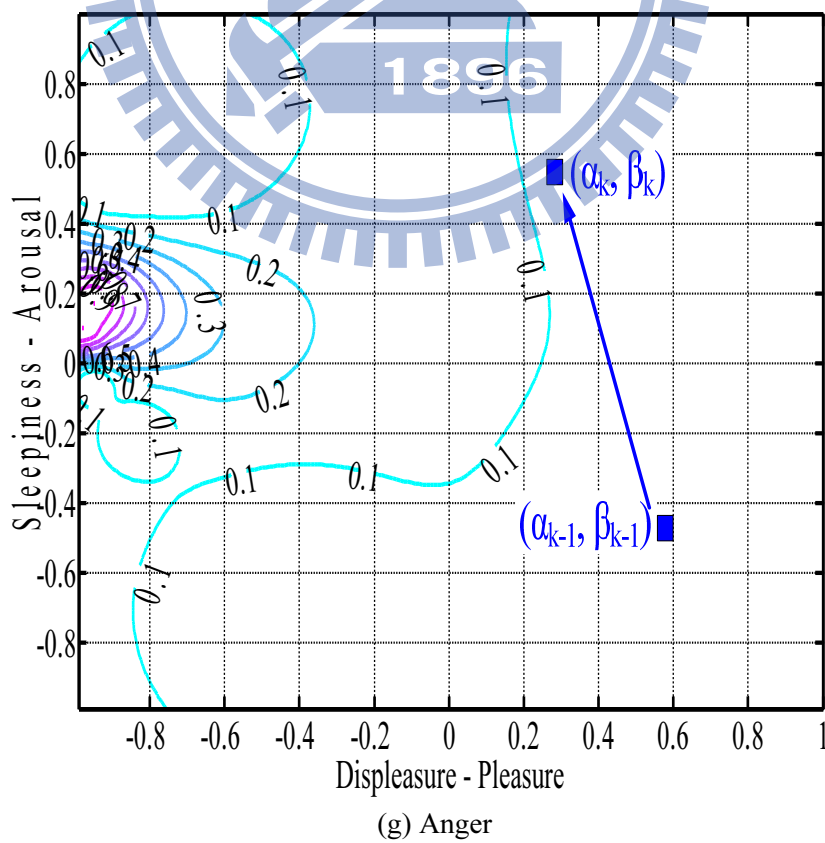
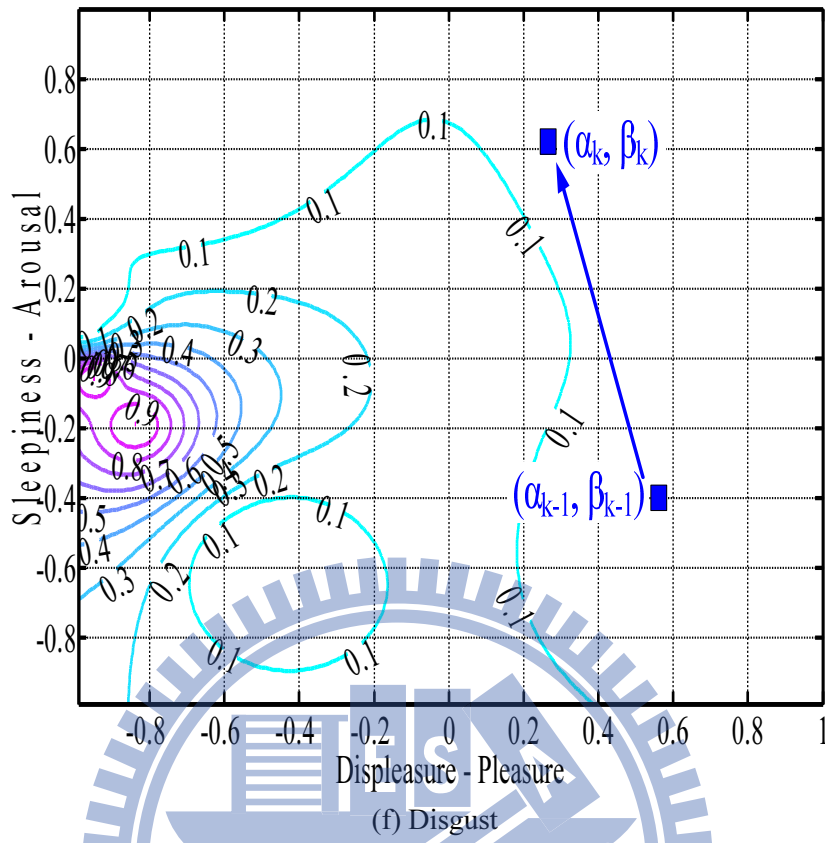


Fig. 2-4: Fusion weights distribution for seven facial expressions.





























same contour curve in the figure has the identical neutral weight. The maximum weight (1) occurs at (0.61, -0.47) in the pleasure-arousal plane. It is seen that the neutral weight decreases while the robotic mood state moves away from (0.61, -0.47). Figure 2-4(b) shows the weight distribution of happiness expression for the whole robotic mood space. The maximum weight (1) occurs at (0.81, 0.66) and (0.88, 0.59) in the pleasure-arousal plane. It is seen that the happiness weight increases while the robotic mood state moves to the upper right quadrant. Figures 2-4(c)-(g) show similar results that the maximum weight positions are located in corresponding coordinates in Fig. 2-2. These results coincide with the two-dimensional emotional state of facial expressions in Fig. 2-2. Furthermore, the correlation among seven basic emotional behaviors is also checked in the simulation. It is seen that a point on the mood plane will map to a corresponding fusion weight for each of seven basic emotional expressions.

2.2.3 Animation of Artificial Face Simulator

To evaluate the effectiveness of the FKCN-based behavior fusion on actual emotional expressions, we developed an artificial face simulator exploiting the method in [55] to examine robotic facial expressions. The method follows a muscle-based approach and thus mimics the way biological faces operate. The artificial face illustrates the expression based on the contraction of facial muscles. It can also dynamically generate features such as wrinkles [55]. Emotions are the high-level concept which is aimed to display via facial expressions. Each emotion influences a different set of muscles. For each emotion and each intensity level, muscles were adjusted to match the reference drawing.

In this simulation, seven basic facial expressions: *neutral*, *happiness*, *surprise*, *fear*, *sadness*, *disgust* and *anger* are first designed by specifying muscles tensions of each expression composed of 7 different fusion weights. Table 2-3 shows some examples of the 7 basic facial expressions generated by the simulator with different weights. One observes that










Table 2-3: Basic facial expressions with various weights executed in the simulator.

	30%	60%	80%	100%
Happiness (RB _H)				
Surprise (RB _{Sur})				
Fear (RB _F)				
Sadness (RB _{Sad})				
Disgust (RB _D)				
Anger (RB _A)				
Neutral (RB _N)				

the facial expression changes from smiling to laughing as the weight of happiness increases and from staring to screaming as the weight of surprise increases. Similarly, the facial expression changes from dreading to panic as the weight of fear increases and from gloomy to crying as the weight of sadness increases. Note that the facial expression of neutral is invariable because it is set as a normal facial expression.

Finally, fused emotional expressions are depicted by linear combination of weighted basic facial expressions. Table 2-4 shows some examples of facial expressions generated by linear combination. The facial expressions with different fusion weights of sadness, anger, surprise and fear are fused to show the complex variation of emotion transition. It provides a quantitative and vivid way to express the feeling of human emotion.

Table 2-4: Linear combined facial expressions with various weights on the simulator.

		
Sadness:70% Anger:30%	Sadness:50% Anger:50%	Sadness:30% Anger:70%
		
Suprise:70% Anger:30%	Suprise:50% Anger:50%	Suprise:30% Anger:70%
		
Fear:70% Sadness:30%	Fear:50% Sadness:50%	Fear:30% Sadness:70%

2.3 Summary

A method of robotic mood transition for autonomous emotional interaction has been developed. An emotional model is proposed for mood state transition exploiting a robotic personality approach. We apply the concept that emotional behavior is controlled by current emotional state and mood, while the mood is influenced by personality. Here the psychological Big Five factors are utilized to represent the personality. By referring Eq. (2.4) and (2.5), the relationship between personality and mood is described. Furthermore, a two-dimensional scaling result (see Fig. 2-2) is adopted to represent general adult's facial expressions based on pleasure-displeasure and arousal-sleepiness ratings. Based on above mention, an illustration of the proposed robotic emotion model is illustrated in Figure 2-5. Finally, via adopting psychological Big Five factors in the 2-D emotional model, the proposed method generates facial expressions in a more natural manner. The FKCN architecture together with rule tables from psychological findings sufficiently provides behavior fusion capability for a robot to generate emotional interactions.

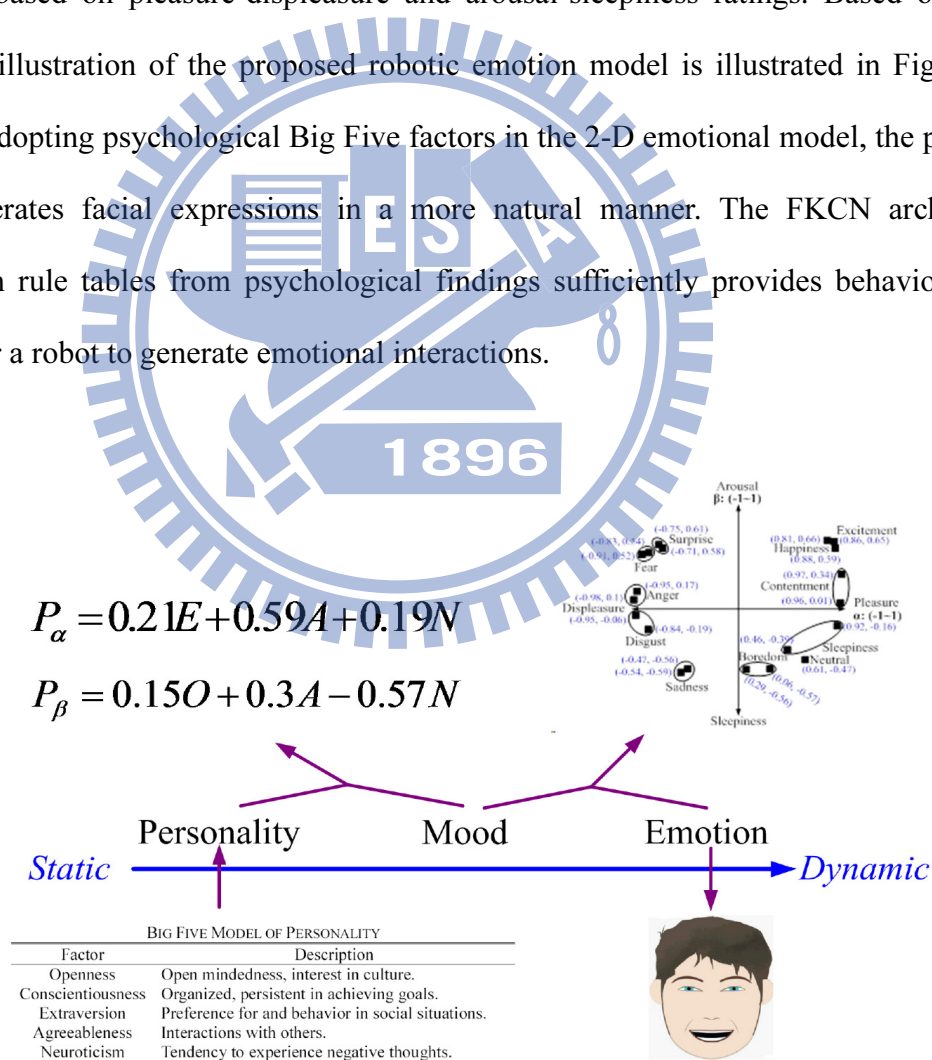


Fig. 2-5: Illustration of the proposed robotic emotion model.

Chapter 3

Human Emotion Recognition

The capability of recognizing human emotion is an important factor in human-robot interaction. For human beings, facial expression and voice reveal a person's emotion most. They also provide important communicative cues during social interaction. A robotic emotion recognition system will enhance the interaction between human and robot in a natural manner. In this chapter, several emotion recognition methods are proposed in the following sections. In 3.1, a bimodal information fusion algorithm is proposed to recognize human emotion by using both facial image and speech signal. In 3.2, a speech-signal-based emotion recognition method is presented.

3.1 Bimodal Information Fusion Algorithm

An embedded speech and image processing system has been designed and realized for real-time audio-video data acquisition and processing. Figure 3-1 illustrates the experimental setup of the emotion recognition system. The stand-alone vision system uses a CMOS image sensor to acquire facial images. The image data from the CMOS sensor are first stored in a frame buffer. Then the image data are passed to a DSP board for further processing. The audio signals are acquired through the analogue I/O port of the DSP board. The recognition results are transmitted via RS-232 serial link to a host computer (PC) to generate the interaction responses of a pet robot.

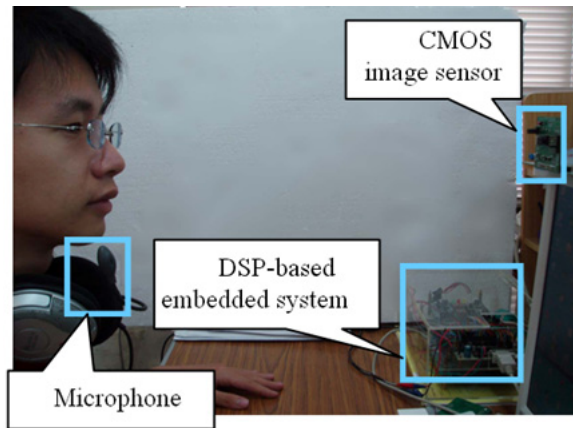


Fig. 3-1: The experimental setup.

Figure 3-2 shows the block diagram of robotic audio-visual emotion recognition (RAVER) system. After a face is detected in the image frame, facial feature points are extracted. Twelve feature values are then computed for facial expression recognition. Meanwhile, the speech signal is acquired from a microphone. Through a pre-processing procedure, the statistical feature values are calculated for each voice frame [66]. After the feature extraction procedures of both sensors are completed, the two feature modalities are sent to an SVM-based classifier [67] with the proposed bimodal decision scheme. Detailed design of facial image processing, speech signal processing and bimodal information fusion will be described in the following sections.

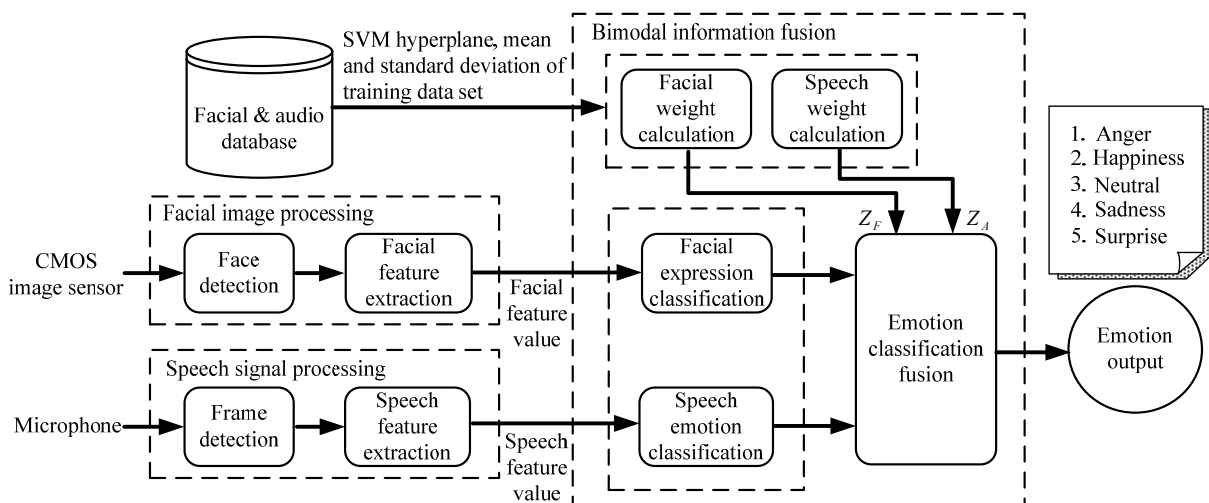


Fig. 3-2: Block diagram of the robotic audio-visual emotion recognition system.

We propose in this section a probabilistic bimodal SVM algorithm. As shown in Fig. 3-2, the extracted features using visual and audio sensors are sent to a facial expression classifier and an audio emotion classifier respectively. In the current design, five emotional categories are determined, namely, anger, happiness, sadness, surprise and neutral. Cascade SVM classifiers are developed for each modality to determine the current emotion state.

3.1.1 Facial Image Processing

The facial image processing part consists of face detection module and feature extraction module. The functional block diagram of the proposed facial image processing is illustrated in Fig. 3-3. After an image frame is captured from the CMOS image sensor, color segmentation and attentional cascade procedure [68] are performed to detect human faces. As a face is detected and segmented, the feature extraction stage is performed to locate the eyes, eyebrows and lips region in the human face area. The system employs edge detection and adaptive threshold to find these feature points. According to the distance between the two selected feature points, several feature vectors are obtained for later emotion recognition. The processing steps will be described in more detail in the following paragraphs.

A. Face Detection

The first step of the proposed emotion recognition system is to detect the human face in the image frame. As shown in Fig. 3-4(a), the skin color is utilized to segment possible human face area in a test image. The morphology closing procedure is then performed to reduce the noise in the image frame, as shown in Fig. 3-4(b). The color region mapping is applied to obtain the human face candidates, as depicted by two white squares in Fig. 3-4(c). Finally, the attentional cascade method is used to determine which candidate is indeed a human face. In Fig. 3-4(d) the black square region indicates a detected human face region.

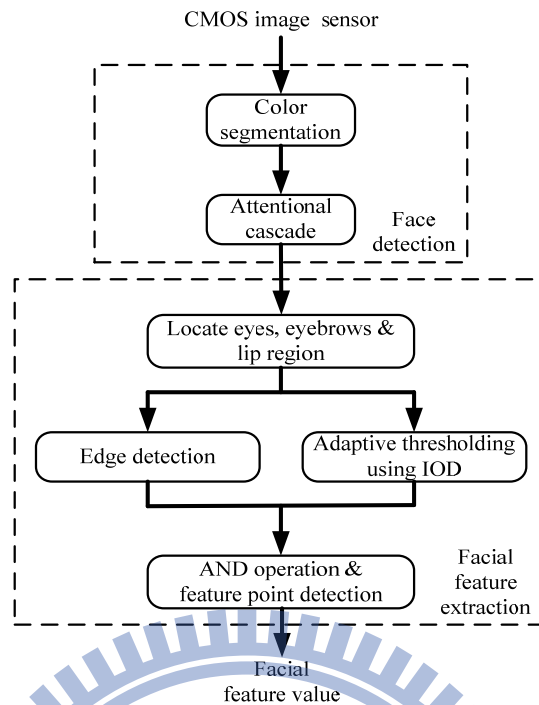


Fig. 3-3: The functional block diagram of facial image processing.

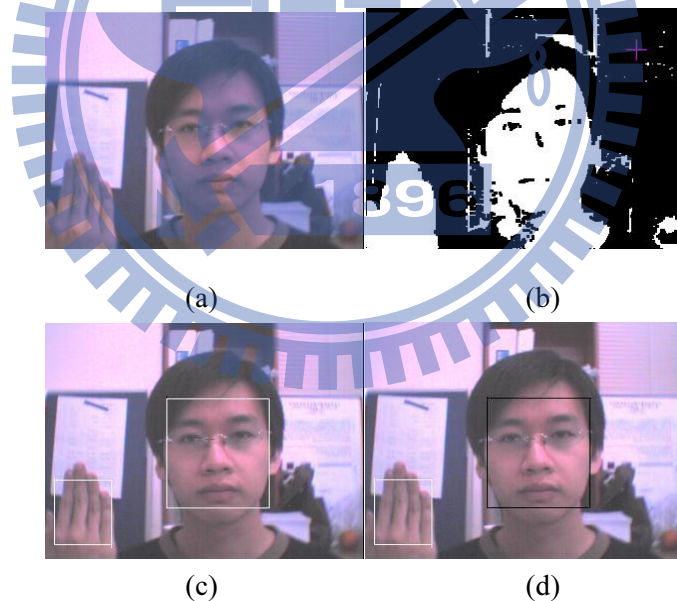


Fig. 3-4: Face detection procedure. (a) Original image, (b) Color segmentation and closing operation, (c) Candidate face areas, (d) Final result obtained by attentional cascade.

B. Facial Feature Extraction

The feature extraction module finds feature points from a frontal face image. The feature points are represented by a vector of numerical data, which represent the position of the facial

features such as eyes, eyebrows, and lips. To search positions of eyes and eyebrows on the upper part of the face image, the characteristics that eyeballs are the darkest areas on the upper face is utilized. Further, the system employs integral optical density (IOD) [69] to find the area of eyes and eyebrows. IOD works on binary images and gives reliable position information of both eyes.

In order to increase the robustness of feature point extraction, our method combines IOD and edge detection. Passing through an AND operation of two successive binary images, the outlines of eyes and eyebrows can be extracted. Figure 3-5 illustrates the definition of all facial feature values and Table 3-1 lists the corresponding detailed descriptions. We defined three feature points for each eye and two feature points for each eyebrow. We locate the upper, lower and inner points of eyes as feature points, and set the central, inner points of eyebrows as feature points. Further, there are four feature points for lips, as shown in Fig. 3-5. Figure 3-6 shows the image processing results of extracting eyes and eyebrows feature points. In Fig. 3-6(a), the detected facial image is processed using IOD while edge detection is performed in

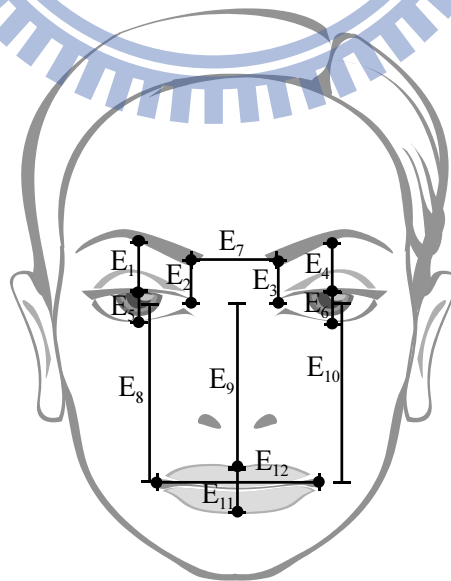


Fig. 3-5: Definition of the facial feature points and feature values.

Table 3-1: The description of facial feature values.

Features	Description
E_1	Distance between the central of right eyebrow and eye
E_2	Distance between the right eyebrow and eye
E_3	Distance between the left eyebrow and eye
E_4	Distance between the central of left eyebrow and eye
E_5	Distance between upper and lower right eye contour
E_6	Distance between upper and lower left eye contour
E_7	Distance between right and left eyebrows
E_8	Distance between right side lip and right eye
E_9	Distance between upper lip and eyes
E_{10}	Distance between left side lip and left eye
E_{11}	Distance between upper and lower lip
E_{12}	Distance between right and left side lip

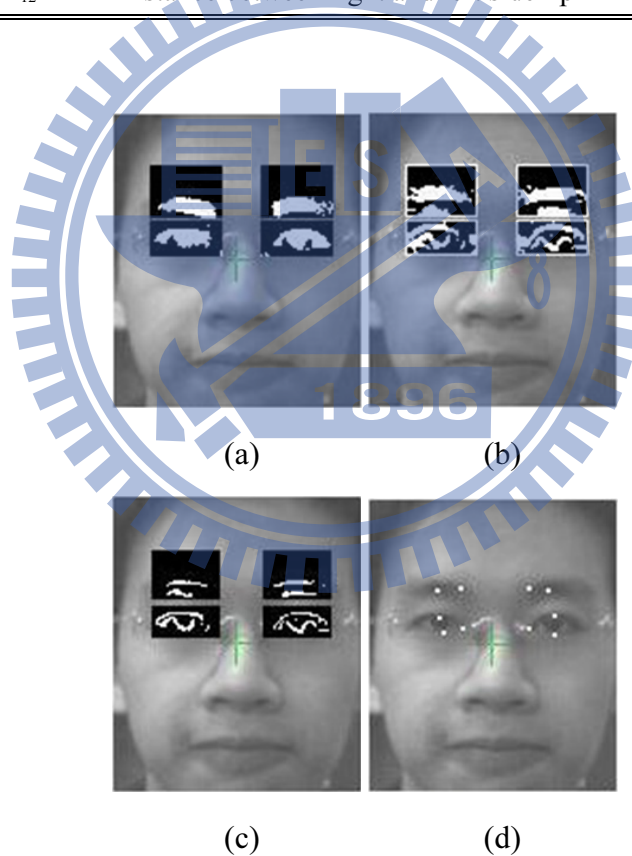


Fig. 3-6: Test results of feature extraction of eyes and eyebrows. (a) Binary operation using IOD, (b) Edge detection, (c) AND operation. (d) Extracted feature points.

Fig. 3-6(b). In Fig. 3-6(c), the AND operation of IOD and edge detection are performed. The feature extraction result is shown in Fig. 3-6(d). Similarly, Fig. 3-7 depicts the result of

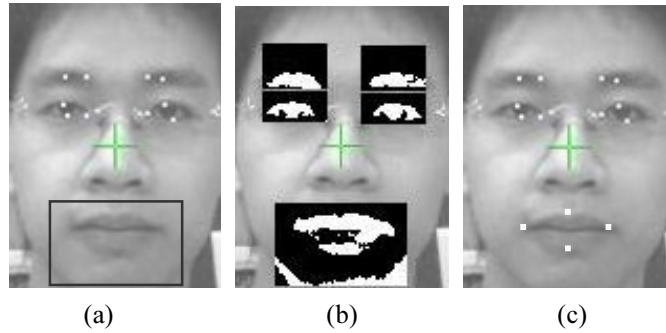


Fig. 3-7: Feature extraction of lips.

feature points extraction of lips. The candidate area in Fig. 3-7(a) is processed by using IOD. The binary detection result is shown in Fig. 3-7(b). Finally, the feature extraction result is obtained as shown in Fig. 3-7(c).

After obtaining the position of facial feature points, we calculate twelve significant feature values, which are distances between two selected feature points as shown in Table 3-1. In order to reduce the influence of distance between a user and the CMOS image sensor, these feature values are normalized for emotion recognition.

3.1.2 Speech Signal Processing

The functional block diagram of the proposed speech signal processing method is shown in Fig. 3-8. The procedure of speech signal processing is divided into two parts. The first part is the pre-processing of speech signal, including endpoint detection and frame setting. The second part is responsible for extracting speech features. The processing steps will be described in more detail in the following paragraphs.

A. Frame Detection

The endpoint detection determines the location of real speech signals by short time energy detection and zero-crossing rate detection. We use the first 128 samples to determine the threshold value in energy detection and then divide a frame into 32ms periods for further

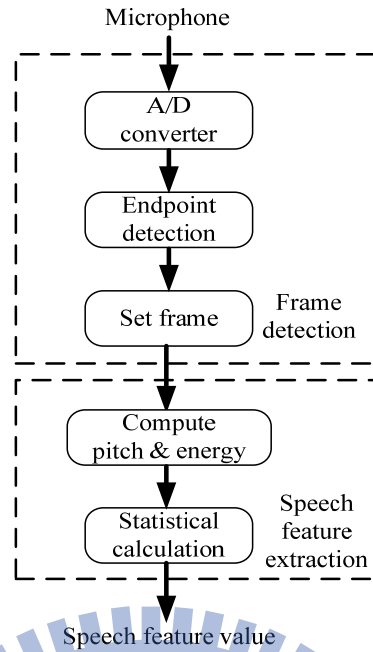


Fig. 3-8: The functional block diagram of facial image processing.

feature extraction processing. The basic idea for estimating emotion by the speech signal is to select features that imply emotion information.

B. Speech Feature Extraction

In this work, contours of pitch and energy are analyzed [29] for human emotion recognition. The pitch contour is obtained by autocorrelation. The maximum point is selected to calculate the pitch values. The energy contour is obtained by calculating the short time energy of each frame. Then the speech feature values can be obtained by computing the statistical quantity of pitch and energy contour. Altogether, twelve speech feature values are obtained for emotion recognition. The elements of speech features are listed in Table 3-2.

3.1.3 Bimodal Information Fusion Algorithm

In order to determine the final result by taking into account both the audio and visual classification results, we developed a bimodal information fusion algorithm to provide a fusion weight for the classifier. According to the principle of SVM, the larger the distance

Table 3-2: The description of speech feature values.

Features	Description
P_{ave}	Average pitch
P_{std}	Standard deviation of pitch
P_{max}	Maximum pitch
P_{min}	Minimum pitch
PD_{ave}	Average of pitch derivation
PD_{std}	Standard deviation of pitch derivation
PD_{max}	Maximum of pitch derivation
E_{ave}	Average energy
E_{std}	Standard deviation of energy
E_{max}	Maximum energy
ED_{ave}	Average of energy derivation
ED_{std}	Standard deviation of energy derivation

between a test sample and the hyperplane, the greater the recognition reliability. Figure 3-9 shows a trained SVM hyperplane and the distance of a test sample to the hyperplane. It can be seen from the figure that the test samples x_1 and x_2 belong to the same class. However, the distance d_1 is smaller than d_2 . Thus, the recognition reliability of test sample x_2 is greater than that of x_1 , because the position of x_2 can resist a larger shift of the hyperplane.

Furthermore, if the training samples are distributed widely, the trained hyperplane will lead to smaller recognition reliability. It may result in a false recognition even the average distance between a test sample and the hyperplane is still large. Figure 3-10 shows two cases

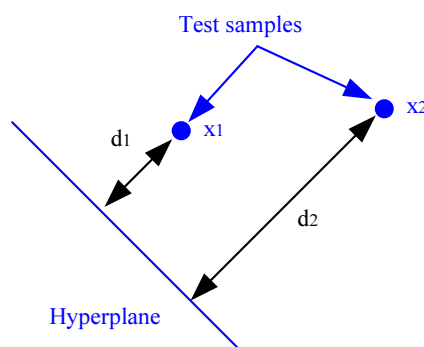


Fig. 3-9: Representing recognition reliability using the distance between test sample and hyperplane.

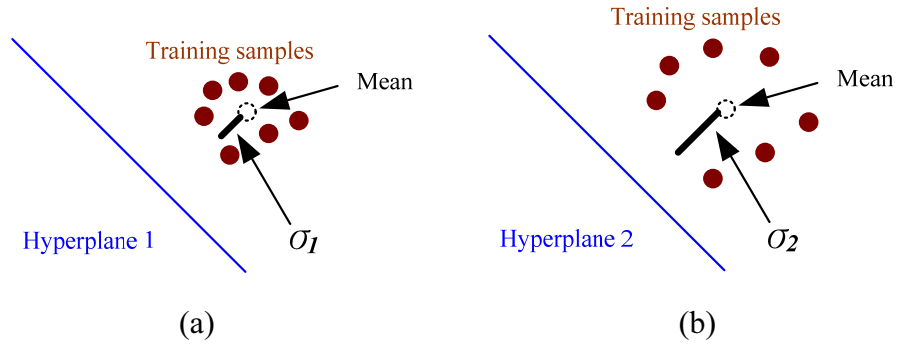


Fig. 3-10: Representing recognition reliability using the standard deviation of training samples. (a) Smaller standard deviation, (b) Larger standard deviation.

of training sample distributions. In Fig. 3-10(a) and (b), the mean values for both distribution are the same, but the standard deviation of hyperplane 1 (σ_1) is smaller than that of hyperplane 2 (σ_2). The recognition reliability of hyperplane 1 is thus greater than that of hyperplane 2, because the training samples are more congregated in the former case. We can conclude that the recognition result is more reliable if the distance between the test sample and hyperplane is larger and the standard deviation of training data set is smaller.

Based on the above observation, we propose the following algorithm of bimodal information fusion:

1) Assume the number of training samples is N for both visual and audio SVM classifiers. Compute the average distance D_{Fave} and D_{Ave} between samples and the hyperplane of facial and speech training data respectively such that:

$$D_{Fave} = \frac{1}{N} \sum_{i=1}^N d_{F_i} \quad (3.1)$$

$$D_{Ave} = \frac{1}{N} \sum_{i=1}^N d_{A_i}, \quad (3.2)$$

where d_{F_i} and d_{A_i} represent the distance between the i_{th} facial and speech training samples and their corresponding respectively.

$$d_{F_i} = \frac{\vec{i}_{F_i} \cdot \vec{H}_F}{|\vec{H}_F|} \quad (3.3)$$

$$d_{A_i} = \frac{\vec{i}_{A_i} \cdot \vec{H}_A}{|\vec{H}_A|}, \quad (3.4)$$

where \vec{i}_{F_i} and \vec{i}_{A_i} represent the i_{th} training sample of facial and speech training data respectively. \vec{H}_F and \vec{H}_A represent the SVM hyperplane of facial and speech data respectively.

2) Compute the standard deviation σ_F and σ_A of facial and speech training data respectively.

$$\sigma_F = \left[\frac{1}{N} \sum_{i=1}^N (d_{F_i} - D_{Fave})^2 \right]^{1/2} \quad (3.5)$$

$$\sigma_A = \left[\frac{1}{N} \sum_{i=1}^N (d_{A_i} - D_{Aave})^2 \right]^{1/2}. \quad (3.6)$$

3) Calculate the distance D_F and D_A between the facial and speech test samples and the corresponding hyperplanes respectively.

$$D_F = \frac{\vec{x}_F \cdot \vec{H}_F}{|\vec{H}_F|} \quad (3.7)$$

$$D_A = \frac{\vec{x}_A \cdot \vec{H}_F}{|\vec{H}_F|}, \quad (3.8)$$

where \vec{x}_F and \vec{x}_A represent the facial and speech test sample respectively.

4) Calculate and normalize the weights of facial classification and speech classification respectively such that:

$$Z_F = \frac{D_F - \sigma_F}{D_{Fave} - \sigma_F} \quad (3.9)$$

$$Z_A = \frac{D_A - \sigma_A}{D_{Aave} - \sigma_A}. \quad (3.10)$$

5) If the classified results of two modalities are not the same, the decision machine compares the magnitude of facial and speech classification weights to obtain a classified

result. If $Z_F \geq Z_A$, adopt the recognition result of facial feature. If $Z_F < Z_A$, then adopt the recognition result of speech feature.

3.1.4 Hierarchical SVM Classifiers

In this work, five facial expressions are categorized according to both the facial and speech information. An SVM hyperplane distinguishes two categories. Therefore two four-stage classifiers need to be constructed as shown in Fig. 3-11. Each stage determines one expression using two emotion categories. The selected emotion category will proceed to the next stage until a final expression is determined. For instance, when an unknown sample appears, the SVM first classifies happiness vs. sadness followed by surprise vs. neutral. After this stage, the corresponding results are further classified at the next stage. For example, the results of the first stage classifiers are assumed to be happiness and surprise (shown as ① and ② in Fig. 3-11). At the second stage, the classifier determines the unknown data as surprise or anger. If the facial image recognition result is surprise but the speech recognition result is anger (shown as ③ and ④), a fusion result is obtained from comparing the weights of both modalities. Here suppose that the weight Z_F of facial image data is larger than the weight Z_A of speech data. So the result of anger (from speech features) vs. surprise (from

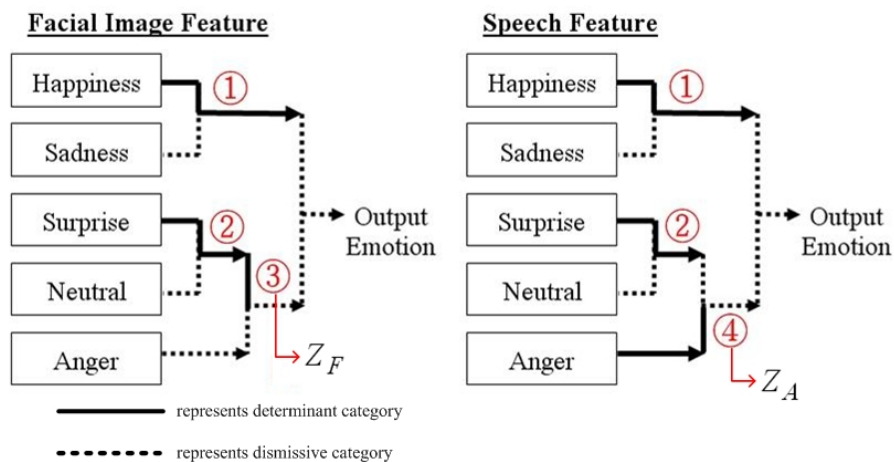


Fig. 3-11: SVM bimodal recognition procedure.

facial features) is classified as surprise. At the last stage, the classifiers determine the unknown data as happiness or surprise as shown in Fig. 3-11. The system will eventually come to a final recognition result.

3.2 Speech-signal-based Emotion Recognition

An embedded speech processing system was designed and produced for real-time speech signal acquisition and processing. Figure 3-12 shows the block diagram of the proposed speech-signal-based emotion recognition system. Speech signals are acquired from a microphone. Using a speech signal pre-processing procedure, the speech voice frames are determined by end-point detection [70]. In the speech feature extraction stage, the fundamental frequency and energy features of a speech frame are extracted to represent the speech signal of interest. After obtaining the features of speech frame, Fisher's linear discriminant analysis (FLDA) is utilized to transfer feature values to a suitable space [71]. The feature values in the transferred space represent significant emotional traits and improve the recognition rate. Finally, a hierarchical support vector machine (SVM) classifies the emotional categories. In order to simplify the design of the emotion recognition system for an

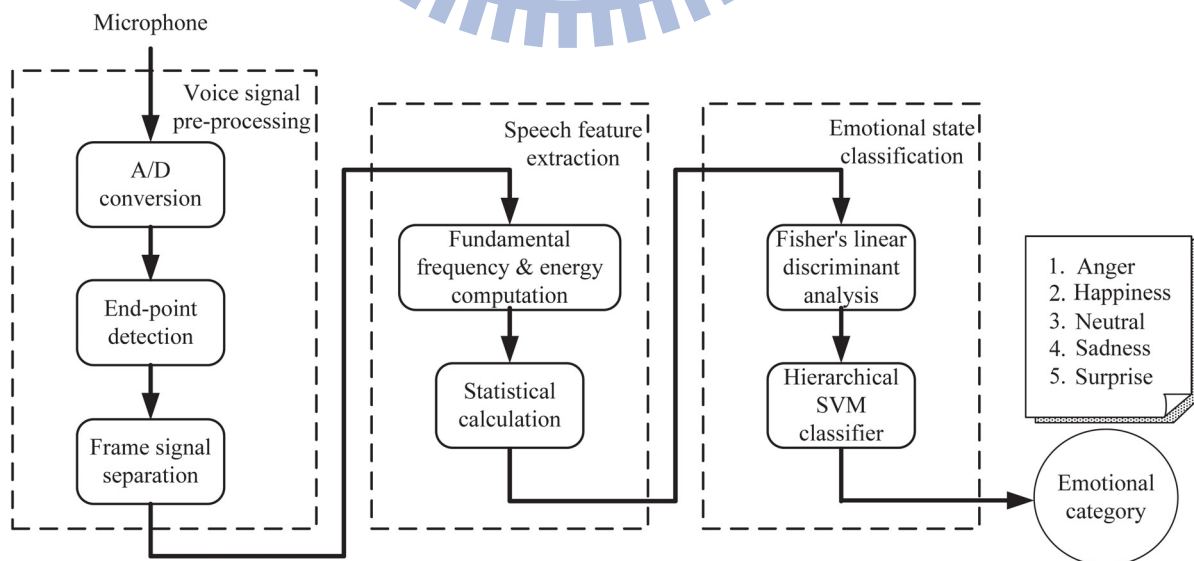


Fig. 3-12: Block diagram of the proposed speech-signal-based emotion recognition system.

entertainment robot, it is assumed that each sentence corresponds to only one emotional category. The detailed design of the emotion recognition system is presented in the following section.

3.2.1 Speech Signal Pre-processing

Before extracting the features of the speech signal for recognition, a voice signal pre-processing stage separates speech frames from the acquired signal. In this design, pre-processing consists of analog to digital conversion, end-point detection and frame signal separation.

Speech signals acquired from the microphone are analog voltage signals. Through amplification and sampling, the analog voltage signal is converted to digital, in a discrete form. Based on the sampling theorem, a sampling frequency is set to be more than twice the bandwidth of the input signals, in order to avoid signal distortion. In general, the spectrum of human speech is less than 4K Hz. The sampling frequency is set to 8K Hz, in this study. Furthermore, a normalization scheme is used to reduce the influence of constantly changing input signals. The normalized speech signal is obtained such that:

$$x(n) = \frac{x_{ori}(n)}{x_{max}} \quad n = 1, 2, \dots, N \quad (3.11)$$

$$x_{max} = \max(x_{ori}(n)) \quad n = 1, 2, \dots, N, \quad (3.12)$$

where $x(n)$ represents the normalized speech signal, $x_{ori}(n)$ represents the original speech signal and x_{max} is the maximum value in the sequence, $x_{ori}(n)$. By dividing with x_{max} , as shown in Equation (3.11), the amplitudes of whole speech signal are normalized between -1 and 1.

In order to extract the emotional features in a voice, a frame size must first be determined for the digitized speech signal. Short-time energy, which is an acoustic feature that correlates the sampled amplitude in each voice frame, is calculated such that:

$$E(k) = \sum_{m=0}^{N-1} |x(n+m)|, \quad (3.13)$$

where $E(k)$ is the short-time energy in the k^{th} frame, $x(n)$ represents the normalized speech signal and N is the frame size. The starting and terminal thresholds are then determined for the voice frame, to determine the starting and terminal points respectively by using empirical rules. Once the value of $E(k)$ is greater than the starting threshold, the starting point is determined. However, the terminal point is determined when the value of $E(k)$ is smaller than the terminal threshold. Hence a frame size, N , is determined as the real speech signal. As shown in Figure 3-13, the starting and terminal points of a speech frame are determined by the starting threshold and the terminal threshold, respectively.

The zero-crossing rate (ZCR) is then used for audio frame setting. Zero-crossing rate is a basic acoustic feature. It is equal to the number of zero-crossings of the waveform within a given frame. Here the zero-crossing rate is defined as the number of times which the speech signals cross the zero value origin of the y -coordinates. In general, the zero-crossing rate of non-speech and environmental noise is lower than that of human speech [72]. The zero-crossing rate is calculated such that:

$$Z(k) = \frac{1}{2} \sum_{m=0}^{N-1} |\text{sgn}(x(n+m)) - \text{sgn}(x(n+m-1))| \quad (3.14)$$

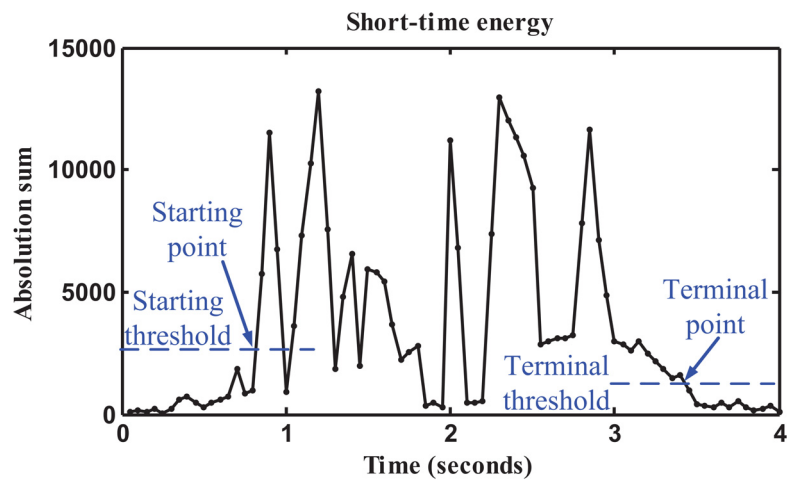


Fig. 3-13: Energy of a speech signal.

$$\text{sgn}[x(n)] = \begin{cases} 1 & \text{if } x(n) \geq 0 \\ -1 & \text{if } x(n) < 0 \end{cases}, \quad (3.15)$$

where $Z(k)$ is the zero-crossing rate of the k^{th} frame. In practice, the short-time energy is used to estimate the starting and terminal points of the whole speech segment, wherein the speech voice occurs. Then, the zero-crossing rate is used to find the real speech signal more precisely. As shown in Figure 3-14, the real speech signal is determined by the ZCR threshold.

In this design, zero-crossing rate and short-time energy are both used to detect the starting and terminal points of non-speech. Figure 3-15 shows the four rules to find the real human speech signal:

- (1) If $E(k)$ is lower than the terminal threshold, it belongs to non-speech.
- (2) If $E(k)$ is higher than the starting threshold, the starting point of the human speech signal is determined.
- (3) If $E(k)$ is lower than the starting threshold and $Z(k)$ is higher than the ZCR threshold of the zero-crossing rates, this is determined as the starting point of the human speech signal.

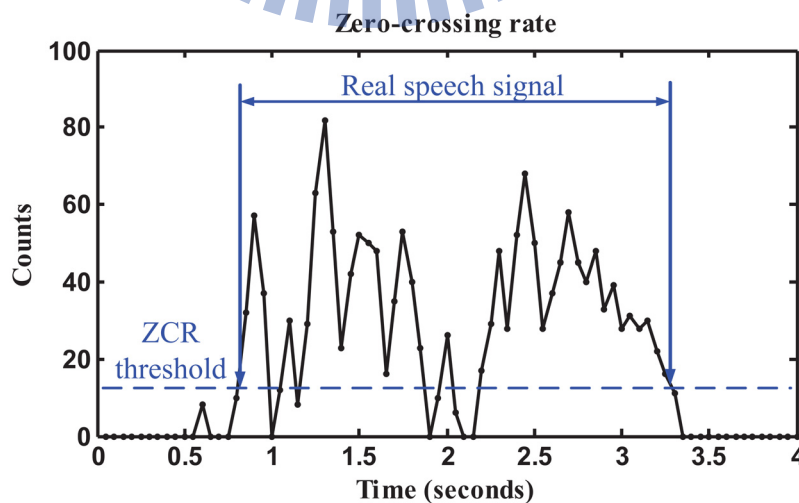


Fig. 3-14: Zero-crossing rate of a speech signal.

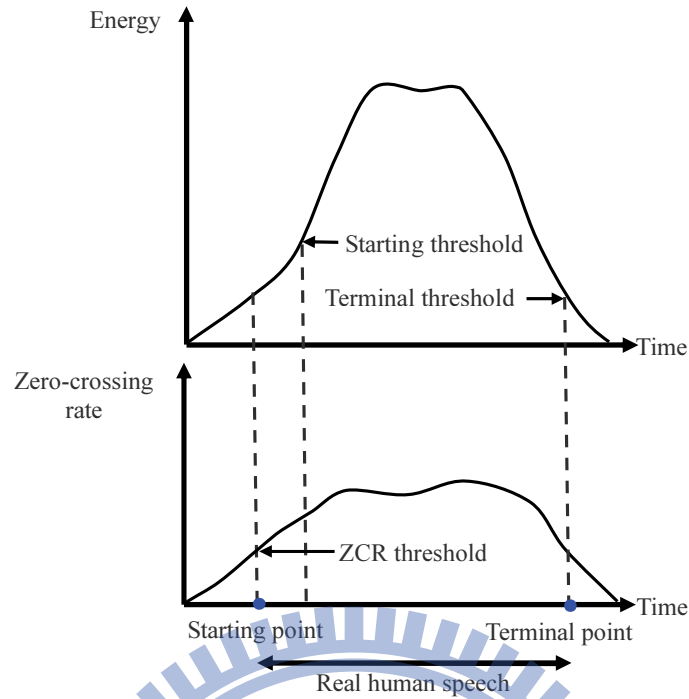


Fig. 3-15: Example of real human speech detection.

- (4) If $E(k)$ is lower than the terminal threshold, after the starting point, it is determined that this is the terminal point of the human speech signal.

Using the above rules, the starting and terminal points of speech signals are determined. The boundary of real human speech is also determined. Figure 3-16 shows an example of end-point detection.

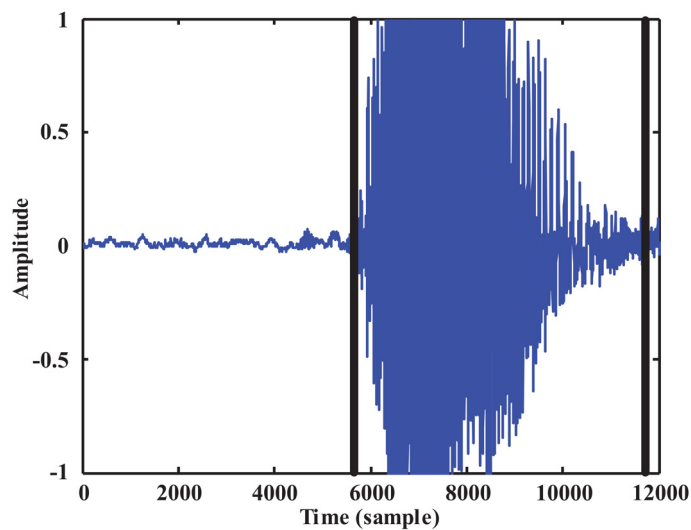


Fig. 3-16: An example of end-point detection.

After obtaining the end-points of the actual human speech signal, suitable presentation of the speech signal is required, before the feature extraction step. In order to reduce the variation between adjacent frames, the overlapping part of the signal is used to avoid discontinuity. This study uses a Hamming window to emphasize the medium signal and to restrain both side signals [73]. Figure 3-17 shows the frame-signal separation using a Hamming window. It can be seen that there are overlaps between frames. The Hamming window is represented such that:

$$Window(n) = \begin{cases} 0.54 - 0.46 \cos\left(\frac{2\pi n}{N-1}\right) & n = 0, 1, \dots, N-1 \\ 0 & otherwise \end{cases}, \quad (3.16)$$

where N is the length of the frame and n is the sample point in a frame. Figure 3-18 shows the procedure for speech signal extraction in each frame. Figure 3-18(a) shows an example of an original speech signal in a frame. Figure 3-18(b) depicts the Hamming window. Figure 3-18(c) is the extracted result for the original speech signal multiplied by the Hamming window. This study uses the first 128 samples to determine the energy threshold values and then divides a frame into several 32 ms periods, for further feature extraction.

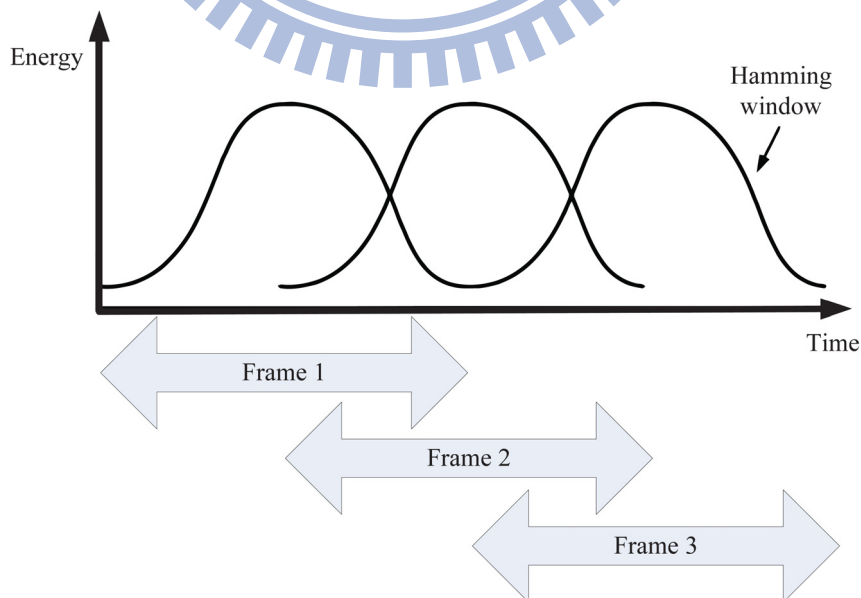


Fig. 3-17: Frame-signal separation using a Hamming window.

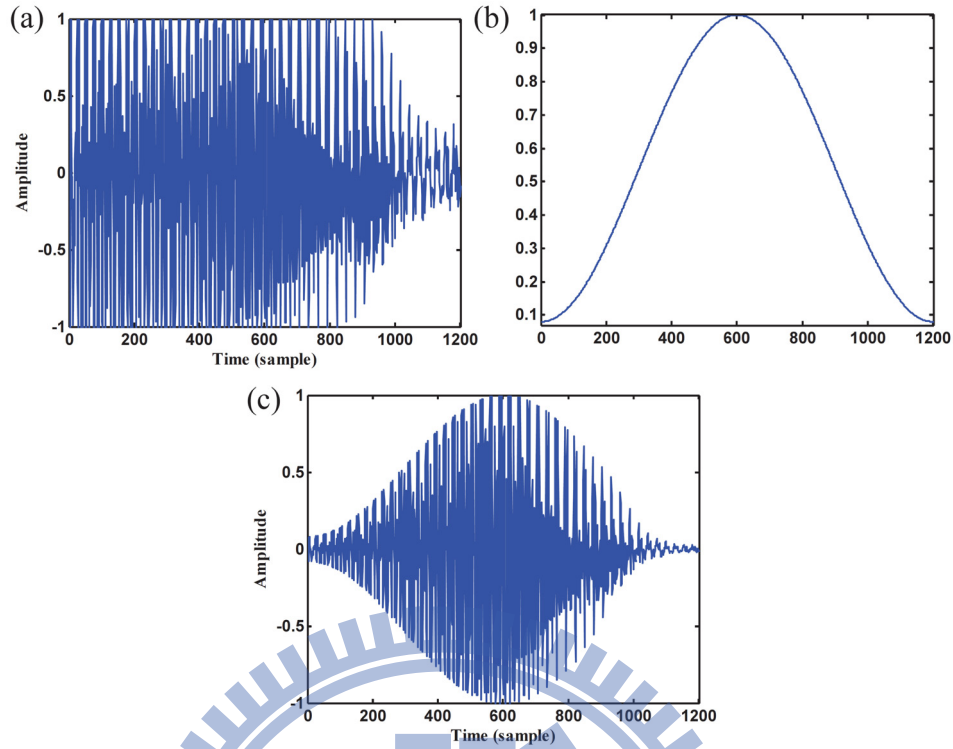


Fig. 3-18: Procedure for speech signal extraction in each frame. (a) A frame of original speech signal. (b) Hamming window. (c) Result of original speech signal multiplied by Hamming window.

3.2.2 Feature Extraction

After the speech signal is obtained for each frame, useful features are extracted from the speech signal. In this work, the contours of the fundamental frequency and energy [29] are used for human emotion recognition. Several methods can be used to extract the fundamental frequency from a speech signal [66]. In this design, the contour of the fundamental frequency is obtained using an autocorrelation function. The fundamental frequency is determined by the maximum autocorrelation value. The autocorrelation function is defined such that:

$$R(d) = \sum_{d=0}^{N-1-d} x(n) \cdot x(n+d), \quad (3.17)$$

where d is the shifting parameter. The value of d that maximizes $R(d)$ over a specified range is selected as the period of the fundamental frequency of the sample points. Figure 3-19 shows the original time response of the speech signal and results for feature extraction of the fundamental frequency. The energy contour is obtained by calculating the short-time energy

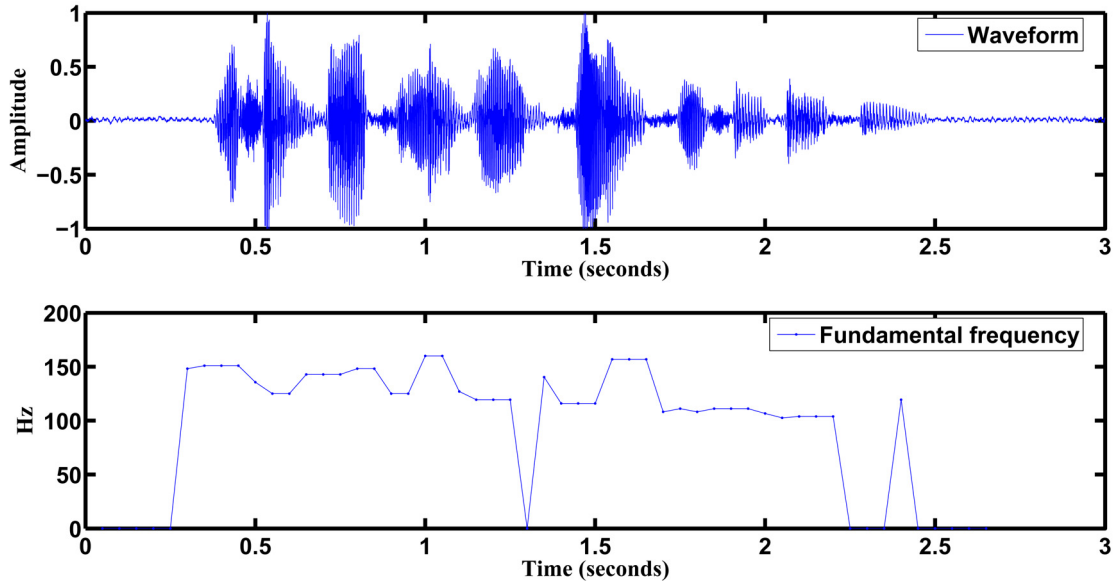


Fig. 3-19: The original time response of the speech signal and the results for feature extraction of the fundamental frequency.

of each frame in Equation (3.13).

After obtaining the short-time energy and fundamental frequency values, the statistical values of fundamental frequency and energy features are calculated, including average, standard deviation, maximum, minimum and median. These statistical values are listed in Table 3-3. Fourteen statistical values are defined in this study. Based on observation, the selected statistical features are sufficient to express variations in emotion and produce satisfactory results.

Table 3-3: The description of speech feature values.

Fundamental frequency (F_0)	Energy
1. Average of F_0	9. Average energy
2. Standard deviation of F_0	10. Standard deviation of energy
3. Maximum of F_0	11. Maximum energy
4. Minimum of F_0	12. Median energy
5. Median of F_0	13. Average of energy derivation
6. Average of F_0 derivation	14. Standard deviation of energy derivation
7. Standard deviation of F_0 derivation	
8. Maximum of F_0 derivation	

3.2.3 Emotional State Classification

After obtaining the statistical features from the speech signal, a suitable classification procedure is required to recognize the emotion categories. In order to increase the discriminability of the feature values, Fisher's linear discriminant analysis (FLDA) [74] is used to find a suitable subspace in which to discriminate emotional categories. An SVM [67, 75] has been an effective method for designing recognition systems. This study uses both FLDA and SVM to classify the emotional categories.

FLDA is a popular method for pattern recognition, to find a linear combination of features which separate two or more classes of objects. It projects the original high-dimensional data onto a low-dimensional space. All of the classes are well separated by maximizing the Raleigh quotient [76]. In FLDA, one assumes there are r training sample vectors, given by $\{ts_i\}_{i=1}^r$, for p classes: C_1, C_2, \dots, C_p , and that there are r_j samples for the j^{th} class, such that:

$$r = \sum_{j=1}^p r_j. \quad (3.18)$$

Let μ be the mean of all of the training samples, such that:

$$\mu = \frac{1}{r} \sum_{i=1}^r ts_i, \quad (3.19)$$

and μ_j be the mean of the j^{th} class, such that:

$$\mu_j = \frac{1}{r_j} \sum_{ts_i \in C_j} ts_i, \quad (3.20)$$

where the within-class scatter matrix S_w and the between-class scatter matrix S_b are defined as follows:

$$S_w = \sum_{i=1}^{r_j} \sum_{j=1}^p (ts_i - \mu_j)(ts_i - \mu_j)^T \quad (3.21)$$

$$\mathbf{S}_B = \sum_{j=1}^p r_j (\boldsymbol{\mu}_j - \boldsymbol{\mu})(\boldsymbol{\mu}_j - \boldsymbol{\mu})^T . \quad (3.22)$$

The goal is to find a transform vector \mathbf{w} such that the Raleigh quotient is maximized. The Raleigh quotient is defined such that

$$q = \frac{\mathbf{w}^T \mathbf{S}_B \mathbf{w}}{\mathbf{w}^T \mathbf{S}_W \mathbf{w}} , \quad (3.23)$$

\mathbf{w} can be defined by solving a generalized eigen problem, as specified by $\mathbf{S}_B \mathbf{w} = \lambda \mathbf{S}_W \mathbf{w}$, where λ is a generalized eigenvalue. An $L \times M$ matrix, \mathbf{W} , can be found to transform the original L -dimensional data into a M -dimensional space. It is expected that the p classes can be well separated in this M -dimensional space. In this work, M is selected as 12 from practical test. Since voice signals are noisy and direction sensitive, the FLDA is used to efficiently discriminate the speech features. In this study, each emotional sentence is represented as fourteen statistical features which are listed in Table 3-3. Then these fourteen statistical features are projected into a subspace by using the transformation matrix \mathbf{W} obtain the new twelve feature values. Afterward each emotional sentence is transformed into twelve feature values for recognition.

SVM is a two-class classifier for a set of related supervised learning methods that analyze data and recognize patterns. The SVM model represents examples as points in space. It determines a hyperplane, so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category, based on the side of the gap on which they fall. In this study, five classes of emotional categories are classified by using SVM. In order to utilize this

two-class classifier to classify five categories, a hierarchical SVM is adopted [70]. The hierarchical SVM classifier is illustrated in Fig. 3-20. In this design, five emotional states are categorized. An SVM hyperplane can distinguish two categories. Therefore a four-stage classifier is developed, as shown in Fig. 3-20. Each stage determines one emotional state from the two and the selected one proceeds to the next stage, until a final emotional state is determined. When unknown emotional speech is imported into the SVM, as shown in Fig. 3-20, the SVM first classifies neutral vs. happiness, then classifies anger vs. surprise. After these stages, the corresponding results are further classified at the next stage. For example, the results of the first and second stage classifiers are assumed to be happiness and surprise (shown as ① and ② in Fig. 3-20). At the third stage, the classifier determines the unknown data as surprise or sadness. If the classification result is surprise (shown as ③), then the classifier determines that the unknown data is happiness or surprise, in the final stage (shown as ④). The system eventually produces a recognition result.

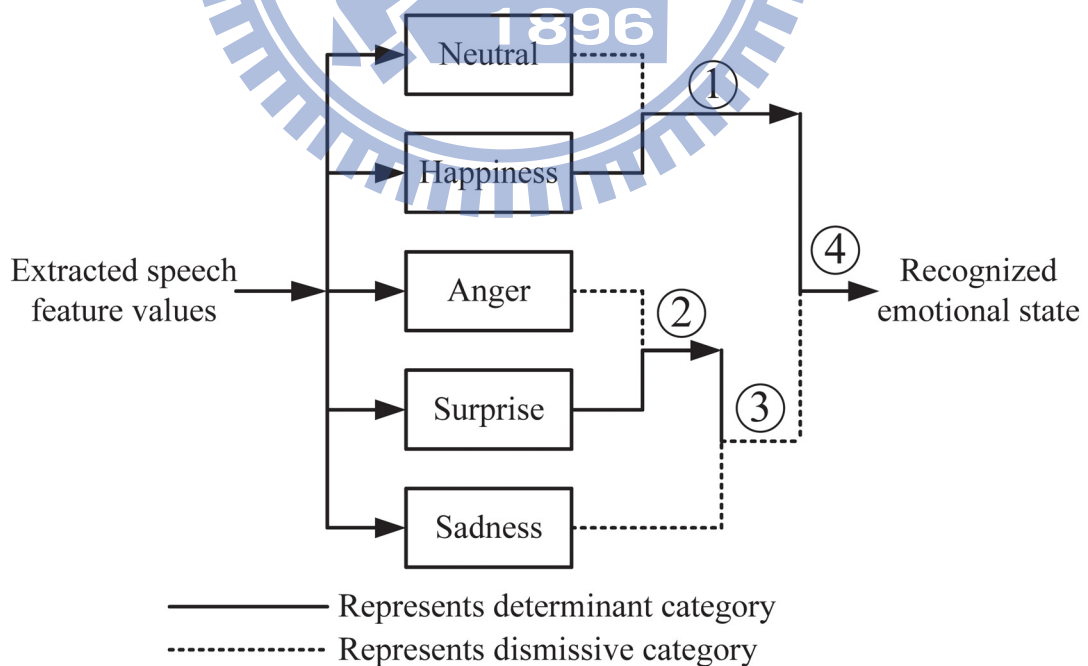


Fig. 3-20: Structure of the hierarchical SVM classifier.

3.2.4 Implementation of the Emotion Recognition Embedded System

The developed algorithms were implemented on a DSP-based embedded system [77], to facilitate the experimental study of an entertainment robot. The embedded system consists of a microphone and a DSK6416 DSP board from Texas Instruments. The selection of the DSK6416 as the main processing unit is because of its high performance in fixed-point calculation, with a 1 GHz clock rate. Figure 3-21 shows the TMS320C6416 DSK codec interface [78-79]. The DSK uses a Texas Instruments AIC23 stereo codec for input and output of audio signals. The codec samples an analog signal from a microphone and converts the signal into digital data, so that it can be processed by the DSP. The DSP chip and codec communicate via two serial channels; one controls the codec's internal configuration registers and the other is responsible for digital audio samples. As shown in Fig. 3-21, the McBSP1 is used as the unidirectional control channel; the McBSP2 is used as the bi-directional audio-data channel. The codec has a 12 MHz system clock. The internal sample rate subdivides the 12 MHz clock to generate common frequencies, including 48 KHz, 44.1 KHz and 8 KHz; a frequency of 8 KHz is selected to sample the user's speech signal, in this study. As a user speaks into the microphone, the embedded system acquires speech signals and begins to recognize the user's emotional state. The recognition results are transmitted via

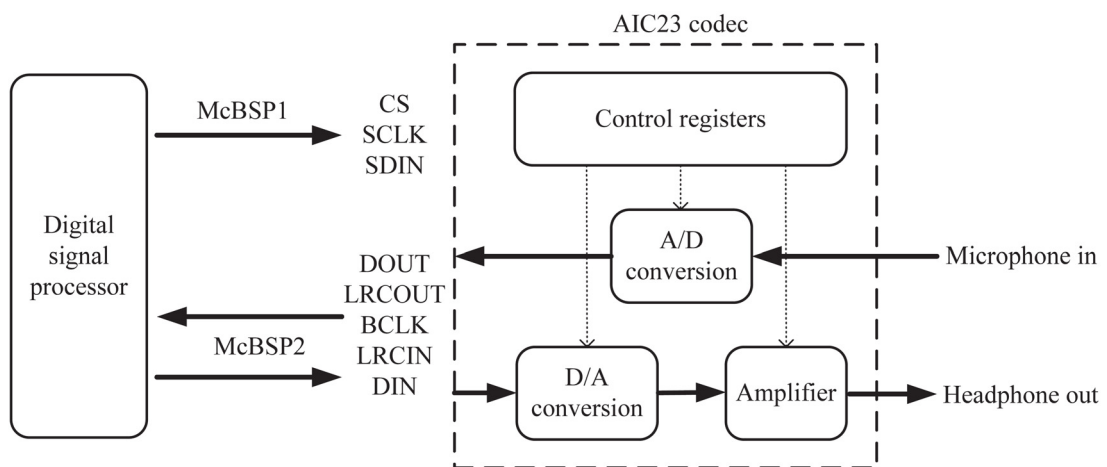


Fig. 3-21: The TMS320C6416 DSK codec interface.

RS-232 serial link to a host computer (PC), where intelligent responses are generated to react to the received speech signal.

In order to test the emotion recognition system in practical scenarios of human-robot interaction, the embedded speech processing system is integrated within the self-built entertainment robot. Figure 3-22 shows an interaction scenario for a user and the entertainment robot. The control architecture of this robot is depicted in Fig. 3-23. The DSP-based system is installed at the back of the entertainment robot. Seven Radio Controlled (RC) servos are used to control the movement of the ears, head, hands and legs of the entertainment robot. A motor servo controller, from Pololu Robotics and Electronics Inc. [80], controls the RC servos in the robot. The DSP-based emotion recognition system estimates emotion categories and determines, in real time, a suitable response for the entertainment robot. Some interesting studies [81-82] have utilized microphone arrays to avoid using a headset. Their methods improve the speech recognition system to cope with noise and direction sensitivity problems. In this study, we focus on the integration of emotional speech recognition algorithm and entertainment robot. In order to reduce the influence of the sound of robot motion or surrounding interference, a headset is used in the experiments, as shown in Fig. 3-22.

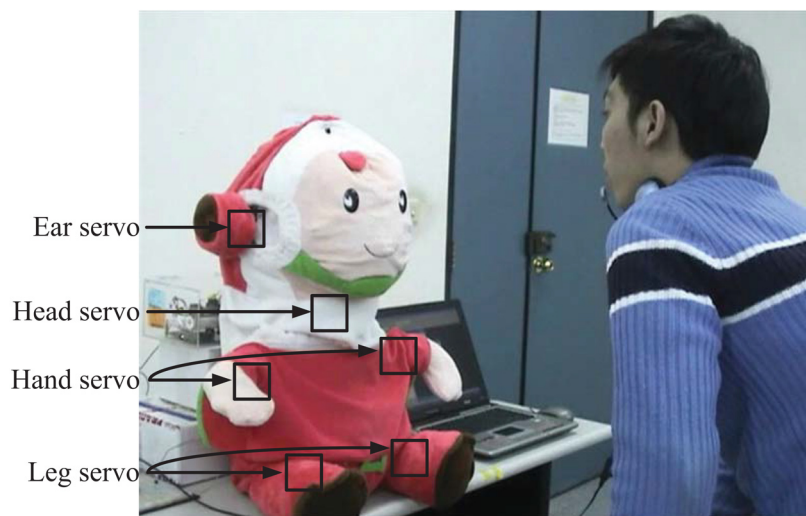


Fig. 3-22: Interaction scenario for a user and the entertainment robot.

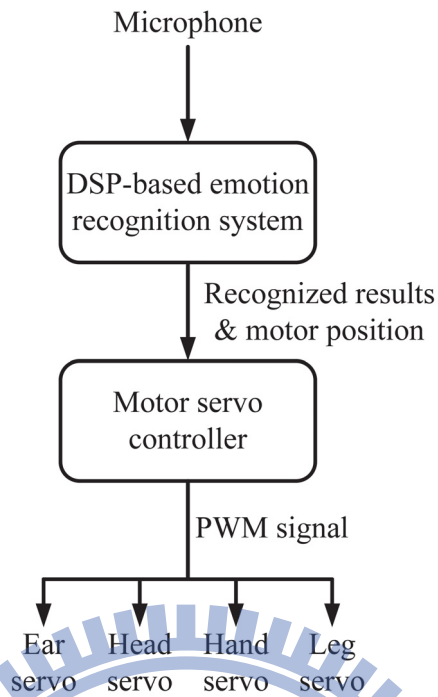


Fig. 3-23: Control architecture of the entertainment robot.

3.3 Summary

In this chapter, two human emotion recognition methods, including bimodal information fusion algorithm, speech-signal-based emotion recognition are proposed and presented. All of these emotion recognition methods will enhance the interaction between human and robot in a natural manner.

Chapter 4

Experimental Results

In this chapter, the experimental results of robotic emotion generation and human emotion recognition are presented and discussed. For the robotic emotion generation, both anthropomorphic robotic head and artificial face simulator were employed to evaluate the results of human-robot interaction. In the part of human emotion recognition, the experimental results of three kinds of emotion recognition methods, which are described in Chapter 4, are presented.

4.1 Experimental Results of Robotic Emotion Generation

The developed robotic emotion generation system has been tested and evaluated for autonomous emotional interaction. We first implemented the proposed AEIS on a self-constructed anthropomorphic robotic head for experimental validation. The robotic head, however, has some hardware limitations in completing the evaluation experiments of mood transition system. A face simulator was adopted for testing the effectiveness of proposed human-robot interaction design.

4.1.1 Experiments on an Anthropomorphic Robotic Head

In order to verify the developed algorithms for emotional human-robot interaction, an embedded robotic vision system [77] has been integrated with an anthropomorphic robotic head with 16 degree-of-freedom. The DSP-based vision system was installed at back of the robotic head and the CMOS image sensor was put on the right eye to capture facial images. The system architecture of the robotic head is depicted in Fig. 4-1. A Qwerk platform [83]

works as an embedded controller. It receives estimated emotional intensity of a user from the vision system and output corresponding pulse width modulation signals to 16 RC servos to generate corresponding robotic facial expression. Figure 4-2 shows several basic facial expressions of the robotic head.

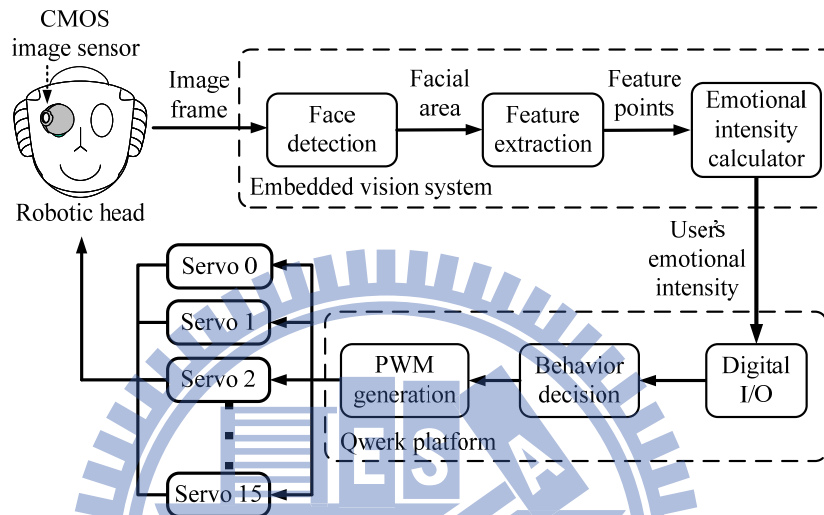


Fig. 4-1: Architecture of the self-built anthropomorphic robotic head.

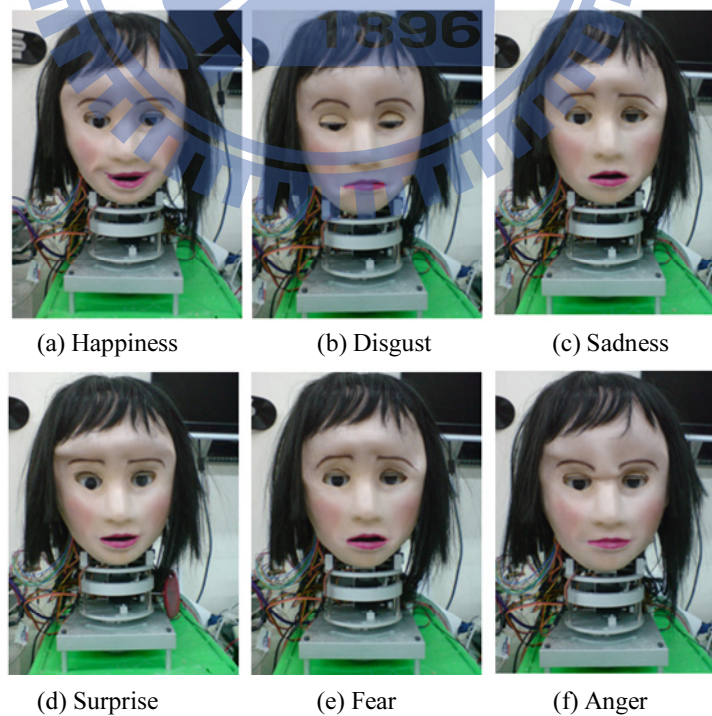


Fig. 4-2: Examples of facial expressions of the robotic head.

In the experiment, a user presented his facial expressions in front of the robotic head as shown in Fig. 4-3. The robot responded to the user with different degrees of wondering as the user presented various intensities of surprise. A video clip of this experiment can be found in [84].

4.1.2 Experimental Setup for the Artificial Face Simulator

A virtual-conversation scenario was set up for testing the effectiveness of proposed human-robot interaction design. As shown in Fig. 4-4(a), in the virtual-conversation test, a subject spoke to the artificial face (on the screen) while the talker's facial expression was detected by a web camera. The subject in the experiment is a student of the authors' Institute. Table 4-1 lists the conversation dialogue and corresponding subject's facial expressions during the test. In the dialogue, the subject complained about her job with sad and angry facial expressions in the beginning. Then the subject talked about the coming Christmas vacation. Her mood varied from angry to happy state. After acquiring facial images, the user emotional state recognizer transferred the user's facial expressions into sets of emotional intensity every 0.5 seconds. The duration of this conversation is around 36 seconds. There are 73 sets of emotional intensity values detected from the user in this conversation scenario. In order to

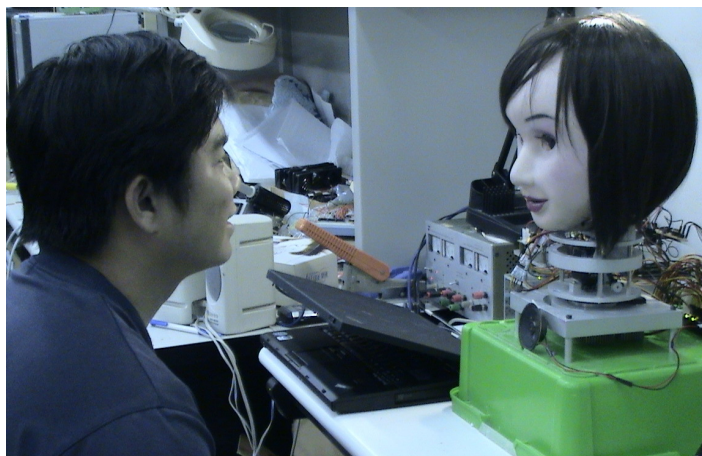


Fig. 4-3: Interaction scenario of a user and robotic head.



Fig. 4-4: Experiment setup: interaction scenario with an artificial face.

Table 4-1: List of the conversation dialogue and corresponding subject's facial expressions.

Sentence #	Dialogue	User's emotional state
1	Hi, Robot. How are you feeling today?	Neutral
2	I feel so bad today. I screwed up my job.	Sad
3	Do you know I feel very sad now? I really hope it was not happened.	Sad
4	I am really angry at myself for my mindless mistake.	Angry
5	However, in a few days it will be Christmas. I think that I can get relaxed during the vacation.	Neutral
6	I am planning to go to Tokyo with my boyfriend. We hear of a carnival will take place this year.	Happy
7	Ha! I can't wait to go to the trip.	Happy

observe the robotic emotional behavior purely due to individual personality and mood transition and avoid undesirable effect caused by error from user emotional state recognition, the detected user emotional intensities are regulated to more reasonable ones manually. Table 4-2 shows part of the regulated user emotional intensities when the subject uttered sentence 1 and 2. These sets of emotional intensity are utilized again as input to test the response of the artificial face with different robot personalities and moods.

Table 4-2: Regulated user emotion intensity of conversion sentence 1 and 2.

Sentence #	$UE_k (k=1,2, \dots, 15)$
1	(0.5,0.2,0,0.3), (0.5,0.3,0,0.2), (0.5,0.4,0,0.1), (0.6, 0.4,0,0), (0.8,0.2,0,0), (1,0,0,0), (1,0,0,0).
2	(0.9,0,0,0.1), (0.8,0,0,0.2), (0.7,0,0,0.3), (0.6,0,0,0.4), (0.5,0,0,0.5), (0.4,0,0,0.6) , (0.4,0,0,0.6), (0.3,0,0,0.7).

4.1.3 Evaluation of Robotic Mood Transition Due to Individual Personality

It is desirable that a robot behaves differently in different interaction scenarios. For example, to keep attention from students in education applications, the robot needs to behave more friendly and funny. Hence the openness and agreeableness scales are designed higher. One can design the desired personality by adjusting the corresponding *Big Five* factors. In this experiment, two opposite robotic individual personalities were designed respectively for RobotA (with more active trait) and RobotB (with more passive trait). The *Big Five* factors were applied to model these two personalities. Table 4-3 lists the assigned scales corresponding to both opposite personalities. As we know, people belonging to active trait are usually open minded and interact with others more frequently. Hence the openness and agreeableness scales of RobotA are higher than those of RobotB and these two higher scales

Table 4-3: Definition of personality scales using Big Five factors.

	RobotA (Active trait)	RobotB (Passive pessimist)
Openness	1	0.3
Conscientiousness	0.5	0.5
Extraversion	0.1	0.1
Agreeableness	0.5	0.2
Neuroticism	0.1	0.3
(P_α, P_β)	(0.34, 0.24)	(0.20, -0.07)

lead the personality parameters (P_{α} , P_{β}) to more positive tendency. Furthermore, a more passive pessimist has the tendency to experience negative thinking in general. Therefore the neuroticism factor of RobotB is higher than that of RobotA. The higher neuroticism factor of RobotB leads its personality more negative tendency on arousal (β axis). After trait values have been identified, the robot personality parameters (P_{α} , P_{β}) are determined by using (2.4) and (2.5). And the proposed robotic mood transition model is built accordingly.

To evaluate the effectiveness of the proposed emotional expression generation scheme based on individual personality, we conducted two sessions of experiments by using the artificial face as shown in Fig. 4-4(b). In the experiments, the same input sets were presented to RobotA and RobotB with the regulated user emotional intensities, respectively with above-mentioned conversation. The robotic mood states were observed as the same user spoke to RobotA and RobotB. Accordingly, the artificial face reacted with different facial expressions resulting from mood state transition. Table 4-4 and Table 4-5 list the calculated robotic mood states (RM_k) and simulated facial expressions corresponding to RobotA and RobotB respectively. Video clips of this experimental can be found in [85].

Figure 4-5 depicts the mood transition of RobotA as the above conversation was performed. The initial mood state of RobotA was set at neutral state (0.61,-0.47), referring to Fig. 2-2. The mood transition trajectories moved from the fourth quadrant to third, second and first quadrant in the end. The corresponding facial expressions varied from neutral (#1) to boredom (#2), sadness (#3), anger (#4), surprise (#5), happiness (#6) and excitement (#7) in the end. The sharp turning point (#5) in Fig. 4-5 indicates that RobotA recognized the subject's emotional state varied rapidly from anger to happiness. Figure 4-6 shows the mood transition of RobotB as the same emotional conversation was performed. The initial mood state of RobotB was also set on neutral state. The corresponding facial expressions varied from neutral (#1) to sleepiness (#2, #3), boredom (#4), sadness (#5), boredom (#6) and then near neutral in the end. Compared with Fig. 4-5, the robotic mood transition of passive trait is

Table 4-4: Facial expressions for the RobotA.





































k	1	5	9	13	17	21
RM _k	(0.61, -0.47)	(0.81, -0.57)	(0.79, -0.98)	(0.64, -0.98)	(0.41, -0.81)	(0.15, -0.54)
Facial expression						
k	25	29	33	37	41	45
RM _k	(-0.14, -0.17)	(-0.47, 0.29)	(-0.70, 0.46)	(-0.99, 0.82)	(-0.92, 0.90)	(-0.58, 0.90)
Facial expression						
k	49	53	57	61	65	69
RM _k	(-0.25, 0.90)	(0.09, 0.90)	(0.42, 0.90)	(0.76, 0.90)	(1, 0.99)	(1, 1)
Facial expression						

Table 4-5: Facial expressions for the RobotB.

k	1	5	9	13	17	21
RM _k	(0.61, -0.47)	(0.73, -0.44)	(0.71, -0.33)	(0.62, -0.32)	(0.49, -0.36)	(0.34, -0.44)
Facial expression						
k	25	29	33	37	41	45
RM _k	(0.17, -0.54)	(-0.02, -0.66)	(-0.15, -0.71)	(-0.33, -0.81)	(-0.31, -0.83)	(-0.11, -0.83)
Facial expression						
k	49	53	57	61	65	69
RM _k	(0.09, -0.83)	(0.28, -0.83)	(0.48, -0.83)	(0.67, -0.83)	(0.91, -0.85)	(1, -0.93)
Facial expression						

basically in the regions of boredom, sad and neutral emotion. It stayed almost destructive no matter what kind of the subject's emotional states came into play. On the contrary, the robotic mood transition of active trait scattered in whole emotional space. These features manifest the difference in characters between active and passive traits. This experiment reveals that the proposed mood transition scheme is able to realize robotic emotional behavior with different personality trait. Video clips of the mood transition for RobotA and RobotB can be found in [86].

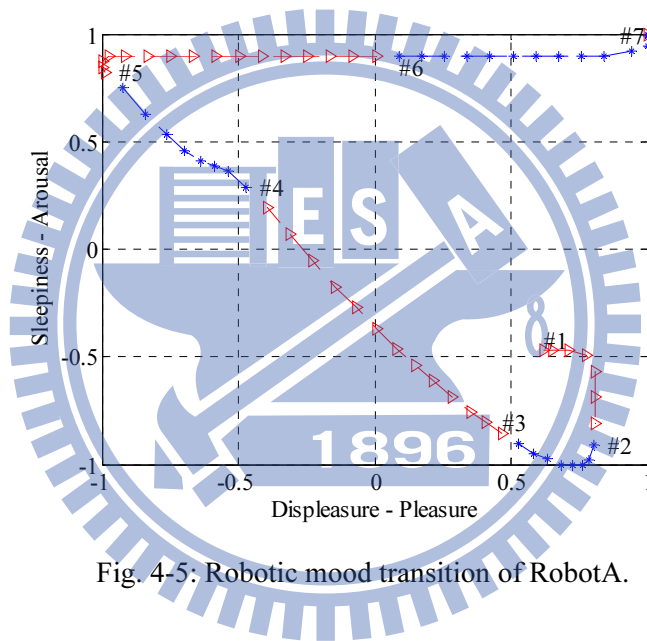


Fig. 4-5: Robotic mood transition of RobotA.

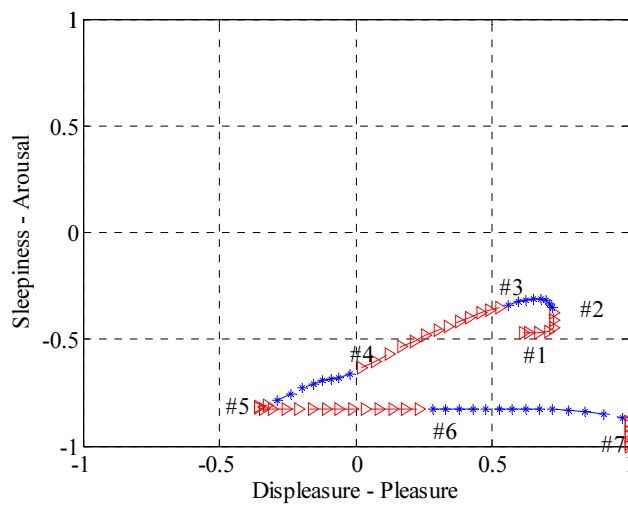


Fig. 4-6: Robotic mood transition of RobotB.

Figure 4-7 shows the variation of seven fusion weights while the subject uttered to RobotA. In the emotional conversation, the subject spoke seven dialogues as shown in Table 4-1. The corresponding fusion weights variations of these seven dialogues are shown by seven sectors in Fig. 4-7. In dialogue #1, the neutral facial expressions dominate the output behavior; this is reasonable since the subject's emotional state is neutral. In dialogue #2 and #3, the weights of sadness gradually increase while the transitions of subject's emotional states are from neutral to sad. Next, the sad weight decreases and the surprise weight increases as the subject feels angry progressively (dialogue #4). In the meantime, the fear weight also increases to respond to the subject's angry expression. After the subject turned to be happy, the surprise and fear weights decrease (dialogue #5) and happy weight increases to dominate the output behavior. After the subject turned to be happy, the surprise and fear weights decrease (dialogue #5) and happy weight increases to dominate the output behavior.

Figure 4-8 shows the variation of seven fusion weights as the subject uttered to RobotB with the same emotional conversation. In dialogue #3 and #4, the weights of sadness gradually increase while the transitions of subject's emotional states are from neutral to sad and angry. After the subject's emotional states become happiness, the sad weight decreases

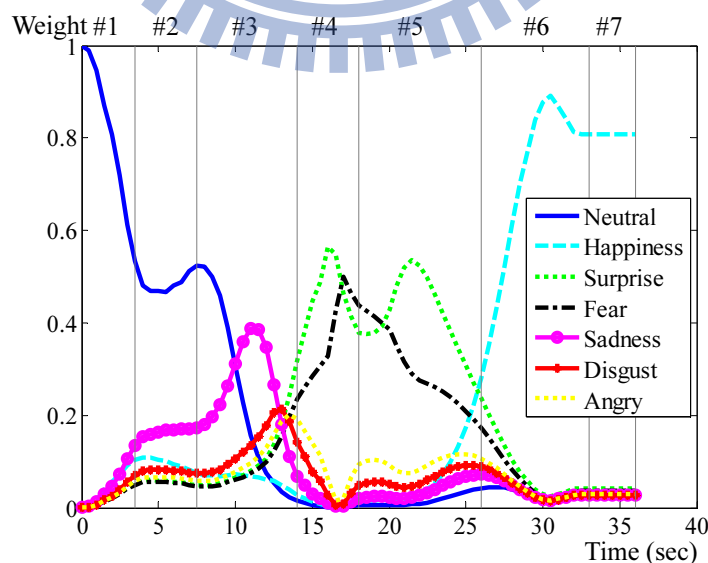


Fig. 4-7: Weights variation for RobotA (active trait).

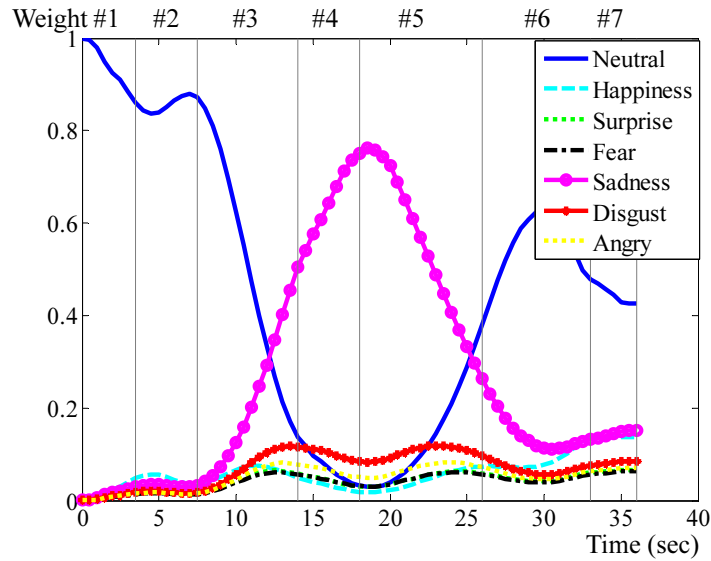


Fig. 4-8: Weights variation for RobotB (passive trait).

(dialogue #5) and neutral weight increases to dominate the output behavior. Compared with RobotA in Fig. 4-7, the personality of passive trait leads to less behavior variations and gets into sadness emotion easily although the subject's emotional states become happiness. These features match the emotional tendency for both active and passive traits.

4.1.4 Evaluation of Emotional Interaction Scheme

In this experiment, questionnaire evaluation for the robot mood transition design was conducted for the emotional conversation performed by the same subject with RobotA, RobotB and RobotC respectively. Here the emotional response of RobotC was designed such that it is irrelevant to the proposed emotional interaction method. RobotC just follows facial expressions as recognized from the subject. The emotional conversation with RobotA, RobotB and RobotC were recorded on three video clips [85] for questionnaire evaluation. We used the *Big Five* factors to evaluate the effectiveness of the proposed robotic emotional expression generation system.

Twenty subjects of age 20~40 were invited to watch the videos of virtual conversation

with RobotA, RobotB and RobotC. The invited subjects were asked to answer questionnaires (see Appendix A) after watching the above videos. In the questionnaire, a subject is asked to give scores from agreeing to disagreeing about the emotional interactions in the videos. We then average the scores using scales (0-1) for the RobotA, RobotB and RobotC respectively. The summary of the experimental results is shown in Fig. 4-9. In the current design, facial expressions of the animation simulator are presented by direct control of pure mood transition. Unlike wording wisdom of human, the readability of facial expressions is related to very different underlying semantics [87-89]. Although the difference between the designed facial animation and human facial expression is obvious, the current design allows an observer to answer the questionnaires more straightforwardly. The major characteristics of designed robotic trait (active and passive) are openness, agreeableness and neuroticism. By observing the openness and agreeableness factors in Fig. 4-9, both factors are evaluated higher for RobotA than those of RobotB. It reveals that RobotA is recognized to have more tendencies to

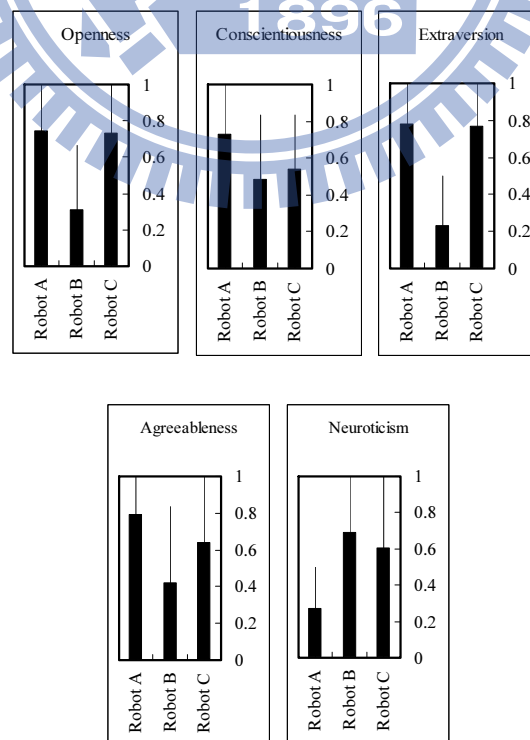


Fig. 4-9: Questionary result of psychological impact.

react and interact with human than RobotB. Moreover, the neuroticism factor of RobotB is evaluated to be higher than that of RobotA. It indicates that the passive pessimist is indeed more inclined to experience negative thoughts than active trait. These results conform to the designed personality in Table 4-3.

As mentioned, RobotC only copies the subject’s facial expressions without any mood transition discussed in this work. In other words, the detected *Big Five* factors of RobotC only show the subject’s personality. In order to verify the difference between robots with the proposed mood transition scheme (RobotA and RobotB) and without it (RobotC), the same 20 subjects answered the questionnaire after watching the videos in [85]. In the questionnaire, a subject is asked to give scores from agreeing to disagree about the degree of natural or artificial interactions in the videos. The summary of the experimental results is shown in Fig. 4-10. Based on the item of natural vs. artificial in Fig. 4-10, RobotA and RobotB both behave more naturally than RobotC. It shows that the proposed mood transition method enables the robot to behave in a human-like manner.

Table 4-6 shows the average values of 20 questionnaire surveys and Table 4-7 shows the corresponding standard deviation of questionnaire result. In Table 4-6, the personality parameters of RobotA and RobotB are estimated as (0.68, 0.19) and (0.43, -0.22), respectively. By comparing with the designed personality in Table 4-3, we see that the personality parameters of RobotA and RobotB are (0.34, 0.24) and (0.20, -0.07), respectively. It is seen

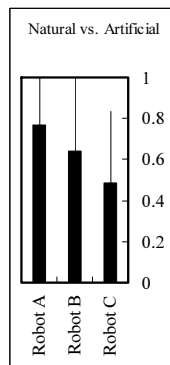


Fig. 4-10: Questionnaire result of Natural vs. Artificial.

Table 4-6: Estimation of personality parameters by questionnaire survey.

	RobotA	RobotB
Openness	0.74	0.31
Conscientiousness	0.73	0.48
Extraversion	0.78	0.23
Agreeableness	0.79	0.42
Neuroticism	0.27	0.69
(P_α, P_β)	(0.68, 0.19)	(0.43, -0.22)

Table 4-7: Standard deviation of questionnaire results.

	RobotA	RobotB	RobotC
Openness	0.16	0.20	0.24
Conscientiousness	0.14	0.21	0.23
Extraversion	0.11	0.13	0.20
Agreeableness	0.15	0.21	0.26
Neuroticism	0.16	0.30	0.29

that both P_α values (0.34 and 0.20) of designed RobotA and RobotB are proportional to the estimated P_α values (0.68 and 0.43) in Table 4-6, respectively. These results are represented as shown in Fig. 4-11. It reveals that both the designed and estimated mood transition velocities of RobotA are about 1.6 times (0.68/0.43 and 0.34/0.20) those of RobotB on the P_α - P_β axes. In another word, both designed and estimated RobotA are happier easily than RobotB with a similar ratio. Furthermore, both of the designed and estimated P_β values of RobotB are negative. It indicates that both the designed and estimated RobotA will tend to arousal and RobotB will tend to sleepiness while the same user's emotional intensity is imported. Hence the estimated results of robot personality parameters are consistent with the designed personality scales in Table 4-3. Based on the experimental results, it can be concluded that a robot can be designed with a desired personality and differently designed robotic personalities give distinct interactive behaviors. Moreover, the emotional robots behave more human-like interaction.

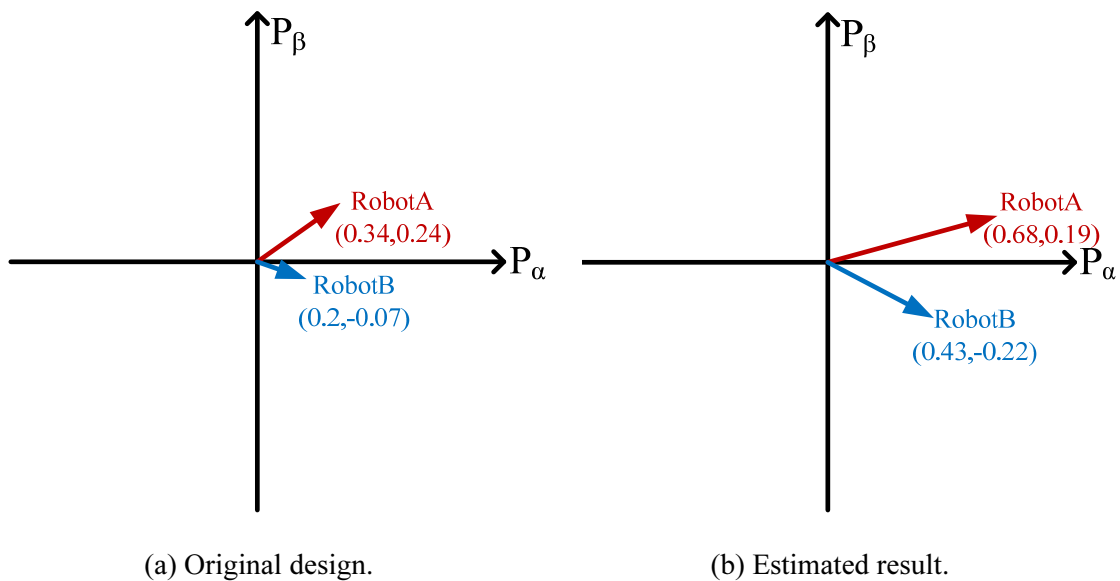


Fig. 4-11: Representation of robot personality parameters.

4.2 Experiments on Bimodal Information Fusion Algorithm

In contrast to many existing visual-only or audio-only databases for benchmark testing [90], there is hardly a database that combines both visual and audio information. Martin *et al.* [91] built an audio-visual emotion database by using a digital video camera. However, the resolution of the camera is too high to be applied for practical pet-robot scenarios, where very often low-cost vision sensors are adopted. Therefore, we built our own database from lab members using off-the-shelf CMOS image sensor and PC microphone.

A DSP-based system has been designed and constructed for the experiments, for both building the database and experimental evaluation. As shown in Fig. 3-2, a user presents his facial expressions in front of the CMOS image sensor and speaks to the microphone. After acquiring both facial and speech signals, the DSP system begins to process the visual and audio information. There are five emotional expressions in the built-up database as described. Figure 4-12 shows part of the database. Currently, the database includes fourteen persons and every one of them expresses their emotions ten times in each emotion category. So there are 140 data samples. In the off-line experiments, we randomly selected 70 data samples as training samples and the other 70 data samples were used as test samples.



Fig. 4-12: Examples of database.

4.2.1 Off-line Experimental Results

Table 4-8 shows the experimental results of five emotional categories using only the speech features. The average recognition rate is 73.7%. Table 4-9 shows the experimental results of five emotional categories using image features. The average recognition rate is 81.7%. The recognition rates of using the proposed bimodal information fusion algorithm to combine both visual and speech features are shown in Table 4-10. The recognition rate of the

Table 4-8: Experimental results using speech features.

Output \ Input	Anger	Happiness	Neutral	Sadness	Surprise	Recognition Rate
Anger	48	12	5	3	2	68.6%
Happiness	8	43	6	10	3	61.4%
Neutral	5	9	48	5	3	68.6%
Sadness	2	6	3	59	0	82.9%
Surprise	0	1	7	1	61	87.1%
Average						73.7%

Table 4-9: Experimental results using image features.

Output Input	Anger	Happiness	Neutral	Sadness	Surprise	Recognition Rate
Anger	53	3	7	6	1	75.7%
Happiness	2	57	11	0	0	81.4%
Neutral	9	7	48	6	0	68.6%
Sadness	1	1	6	62	0	88.6%
Surprise	1	1	2	0	66	94.3%
Average						81.7%

Table 4-10 Experimental results using information fusion.

Output Input	Anger	Happiness	Neutral	Sadness	Surprise	Recognition Rate
Anger	60	3	4	2	1	85.7%
Happiness	3	59	6	2	0	84.3%
Neutral	6	9	52	2	1	74.3%
Sadness	1	1	3	65	0	92.9%
Surprise	0	2	0	0	68	97.1%
Average						86.9%

combined bimodal information is 86.9%. A 5% improvement of the recognition rate is achieved relative to the facial feature and 13% improvement relative to the speech features. It can be seen that the recognition rate of the combined bimodal approach is higher than any single mode approach.

4.2.2 On-line Experimental Results

Further, on-line experiments were carried out using the developed DSP-based emotion recognition system. The training of SVM hyperplane was performed off-line on a PC using the constructed database. The trained parameters of the hyperplane were then transferred and stored in the DSP system for on-line test. In the test, a person presents his/her face in front of

the CMOS image sensor and speaks to the microphone, and DSP system will return the emotion category in two seconds. We invited five new persons to join the on-line experiments. Every person expressed ten times the emotion category with facial expression and voice. The recognition result of using only image information is shown in Table 4-11. The average recognition rate is 74.4%. Table 4-12 shows the bimodal emotion recognition rate of the on-line test. The average recognition rate is 77.6%. The experimental results verify that the proposed method can work effectively in on-line applications. The recognition rate of on-line test is lower than the off-line result. This is mainly due to the image noise in the on-line test. Also, the test samples are new faces and voices, the recognition rate is thus lower than the off-line results.

Table 4-11: On-line experimental results using only image features.

Output Input	Anger	Happiness	Neutral	Sadness	Surprise	Recognition Rate
Anger	40	1	8	0	1	80%
Happiness	3	38	9	0	0	76%
Neutral	1	10	34	4	1	68%
Sadness	1	4	14	30	1	60%
Surprise	0	2	3	1	44	88%
Average						74.4%

Table 4-12: On-line experimental results using information fusion.

Output Input	Anger	Happiness	Neutral	Sadness	Surprise	Recognition Rate
Anger	41	1	7	0	1	82%
Happiness	1	38	9	0	2	76%
Neutral	2	10	35	2	1	70%
Sadness	1	1	14	34	0	68%
Surprise	0	2	2	0	46	92%
Average						77.6%

4.3 Experiments on Speech-signal-based Emotion Recognition

The performance of the proposed emotional voice recognition system was evaluated using a self-built database. Furthermore, experimental study of the proposed system was performed by integrating the DSP-based system into an entertainment robot.

4.3.1 Experiments Using the Self-built Database

The proposed emotion recognition system was tested using a speech database built in the ISCI lab of National Chiao Tung University. There are five categories of emotional speech in the database: happiness, sadness, surprise, anger and neutral. For each category, there are three kinds of different sentences. In order to express the emotion in a natural way, each subject was asked to narrate expressive sentences, in Chinese, to imitate an actual interactive scenario. Table 4-13 lists the meaning of each sentence, in English. Currently, the database includes various emotional utterances from five persons. Each person recorded each sentence six times, so there are 90 utterances per emotion category and 450 utterances in total, in this database. In the following experiments, 45 data samples were randomly selected as training data for each emotional category and the other 45 data samples were used as test data. Part of

Table 4-13: Meaning of sentence content for five emotional categories.

Emotional category	Content of sentence
Anger	<ol style="list-style-type: none">1. How can you do that without my agreement?2. It's none of your business.3. What you are doing is wrong!
Happiness	<ol style="list-style-type: none">1. It's almost new year!2. I will go abroad on vacation tomorrow.3. I won the lottery!
Neutral	<ol style="list-style-type: none">1. It's a sunny day.2. I have something to do later.3. Are you hungry?
Sadness	<ol style="list-style-type: none">1. My cat is lost.2. I got a cold.3. Everything went without a hitch today.
Surprise	<ol style="list-style-type: none">1. Are you serious?2. I can't believe that it really happened.3. Ah! My notebook is lost.

the voice clips of the database can be found in [92].

In order to compare the effect of speech features, fundamental frequency and short-time energy features, the emotion is first evaluated between any two emotional categories. Figure 4-13 shows the experimental results of the SVM classification of any two emotional categories. There are ten combinations of any two emotional expressions. It is seen that the recognition rates for these nine combinations are higher than 85%. The recognition rate of neutral vs. sadness (A vs. E) is the lowest, mainly due to the small prosodic variation between neutral and sad speech utterances. The other recognition rates lie between 85% and 96%. The average recognition rate is 89.2%. This indicates that the proposed statistical features can represent emotional characteristics properly.

The hierarchical SVM classifier (shown as Fig. 3-21) was then employed to recognize five emotional categories. In the experiments, the SVM classifier was trained using a set of 45 data samples for each emotional category. These 45 data samples came from five persons, with each person contributing three samples of each emotional sentence. The other 45 data samples were tested for recognition of the emotion category. The test results are presented in Table 4-14. The average recognition rate for the five emotional expressions is 73.78%. It is

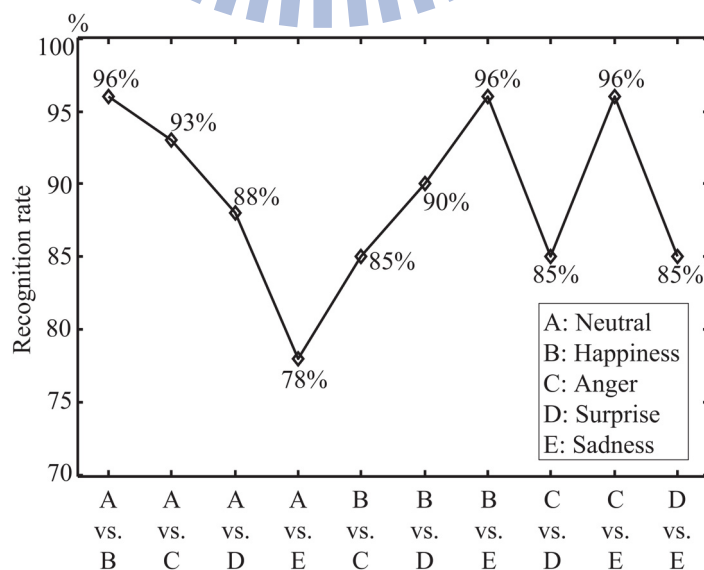


Fig. 4-13: Experimental results of recognition rate for any two emotional categories.

noted that anger can be classified as surprise. It is due to the similar speech rates and tones of these two kinds of sentences in the self-built database. Moreover, the accent and noise of the voice influence the classification results a lot. We will take these factors into consideration in future work.

4.3.2 Experiments with the Entertainment Robot

In this study, we aim to develop an entertainment robot suitable as a children’s toy. In such a robotic application, fast response to natural speech signal is required. Therefore, a simple entertainment robot is built to verify the proposed natural speech signal emotion recognition algorithm. The complete emotion recognition system was integrated into the self-constructed entertainment robot. Figure 4-14 shows a block diagram of the implemented interaction control system on the robot.

In the experiment, a user speaks in front of the robot, as shown in Fig. 3-22. After acquiring the speech signals, the emotion recognition system begins to process the audio information. When no human speech is detected, the robot manifests a bored behavior by turning its head to look around. When a user says “hello” to the robot, with neutral emotion, the robot raises its hands to respond to the user. If a happy emotion from the user is detected, the robot rotates its ears and raises its hands to show a happy gesture. When the user expresses anger to the robot, the robot puts its hands down to portray fear. However, the robot

Table 4-14: Experimental results of recognizing five emotional categories.

Output Input	Anger	Happiness	Neutral	Sadness	Surprise	Recognition rate
Anger	30	0	3	4	8	66.67%
Happiness	1	37	3	4	0	82.22%
Neutral	1	6	35	3	0	77.78%
Sadness	0	4	6	30	5	66.67%
Surprise	5	2	1	3	34	75.56%
					Average	73.78%

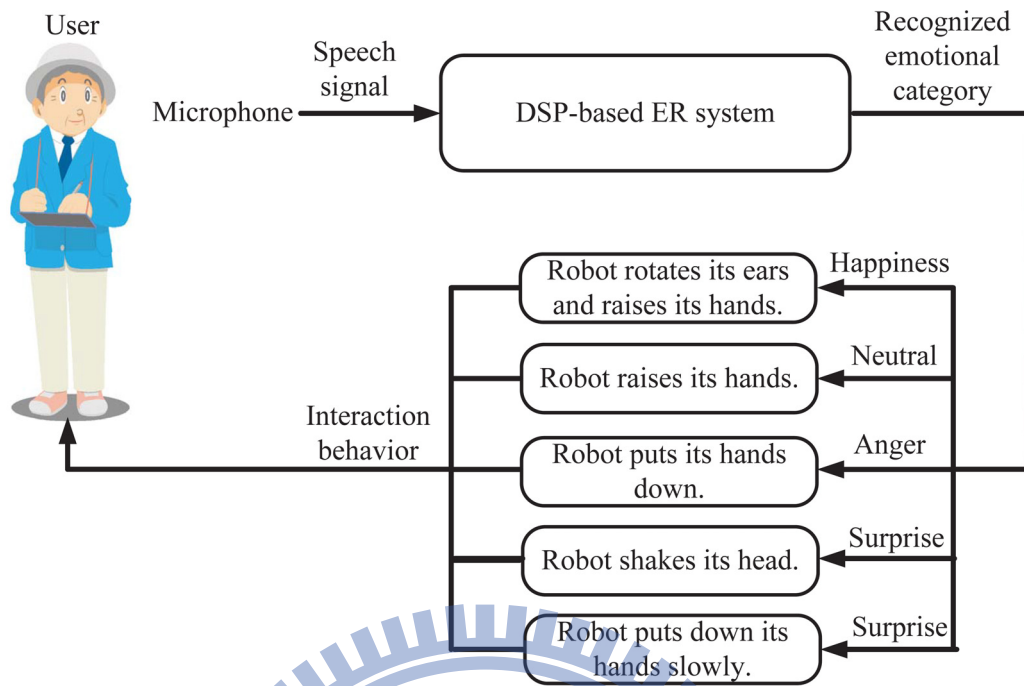


Fig. 4-14: Block diagram of the emotional interaction system.

shakes its head if surprise is detected from the user. Figure 4-15 shows the interaction responses of the robot when the user said, “I am angry!”, with an angry tone. After that, the user used a surprise tone to the robot. As shown in Fig. 4-16, the robot shook its head in response to the recognized emotional state. The experimental results verify that the proposed emotion recognition system allows the robot to interact with a user in a natural and friendly manner. A video clip of the experimental results can be found in [93]. In the future, a fast system will be further studied to recognize human’s emotional speech and interact in a more



Fig. 4-15: Interactive response of the robot as the user says, “I am angry!” (a) The robot puts down its hands to portray fear. (b) The robot continues to put down its hands to the lowest position. (c) The robot raises its hands back to the original position.



Fig. 4-16: Interactive response of the robot, when the user speaks in a surprised tone. (a) The robot shakes its head to the right. (b) The robot shakes its head to the left. (c) The robot puts its head back to the original position.

humanlike manner. Some suitable psychological findings will also be considered to apply to the emotional robotic system.

4.4 Experiments on Image-based Emotional State Recognition

In this design, the user's emotional state (UE_k^n) is used as input to the system. In order to obtain UE_k^n , an image-based facial expression recognition module has been designed and implemented. The facial expression recognition module consists of face detection stage, feature extraction stage and emotional intensity analyzer. The method of facial feature extraction is described in 3.1.1. After obtaining the facial feature points, twelve significant feature values, which are distances between two selected feature points. In order to reduce the influence of distance between a user and the camera, these feature values are normalized for emotion recognition. Thus, every facial expression is presented as a feature set.

To recognize user's emotional states, we further developed an image-based method to extract facial expression intensity. Four feature vectors, namely, \bar{F}_{Neu} , \bar{F}_{Ha} , \bar{F}_{Ang} and \bar{F}_{Sad} are defined to represent the standard *neutral*, *happy*, *angry* and *sad* expressions. Dissimilarities between current feature set of a user ($\bar{F}_{User,k}$) and the standard facial expressions are calculated such that:

$$d_{N,k} = \left\| \bar{F}_{User,k} - \bar{F}_{Neu} \right\| \quad (4.1)$$

$$d_{H,k} = \|\bar{F}_{User,k} - \bar{F}_{Ha}\| \quad (4.2)$$

$$d_{A,k} = \|\bar{F}_{User,k} - \bar{F}_{Ang}\| \quad (4.3)$$

$$d_{S,k} = \|\bar{F}_{User,k} - \bar{F}_{Sad}\| \quad , \quad (4.4)$$

where $d_{N,K}$, $d_{H,K}$, $d_{A,K}$ and $d_{S,K}$ represent respectively, the dissimilarities between the feature set of user and the defined standard neutral, happy, angry and sad expression at sampling instant k . $\| \cdot \|$ represents the Euclidean distance. In our design, the intensity of user's emotion is recognized as the standard facial expression while the dissimilarities between the current feature set and standard facial expression is small. Therefore, the user's emotional intensities UE_k^n are calculated such that:

$$ue_{N,k}^n = \frac{d_{N,k}^{-1}}{d_{N,k}^{-1} + d_{H,k}^{-1} + d_{A,k}^{-1} + d_{S,k}^{-1}} \quad (4.5)$$

$$ue_{H,k}^n = \frac{d_{H,k}^{-1}}{d_{N,k}^{-1} + d_{H,k}^{-1} + d_{A,k}^{-1} + d_{S,k}^{-1}} \quad (4.6)$$

$$ue_{A,k}^n = \frac{d_{A,k}^{-1}}{d_{N,k}^{-1} + d_{H,k}^{-1} + d_{A,k}^{-1} + d_{S,k}^{-1}} \quad (4.7)$$

$$ue_{S,k}^n = \frac{d_{S,k}^{-1}}{d_{N,k}^{-1} + d_{H,k}^{-1} + d_{A,k}^{-1} + d_{S,k}^{-1}} \quad , \quad (4.8)$$

where $ue_{N,k}^n$, $ue_{H,k}^n$, $ue_{A,k}^n$ and $ue_{S,k}^n$ represent respectively, the n^{th} user's emotional intensities at sampling instant k for neutral, happy, angry and sad expressions. By using this procedure, the user's emotional state is represented as a set of four emotional intensities.

In this section, Cohn-Kanade AU-Coded Facial Expression Database [94] is used to verify the proposed method of emotional state recognition. Twenty-four sets of facial images of different basic facial expressions were selected as training data. Each set contains 7 facial images of a particular emotion with various facial expressions. 60 face images of different basic facial expressions were selected as test data. To compare the system with ground truth, we choose the strongest emotion as recognition results. The result of this experiment is shown in Table 4-15. The average recognition rate is 90%.

Table 4-15: Test result of emotion state recognition.

Output \ Input	Neutral	Anger	Happiness	Sadness	Recognition Rate
Neutral	13	1	0	1	87%
Anger	0	15	0	0	100%
Happiness	2	0	13	0	87%
Sadness	1	1	0	13	87%

Figure 4-17 shows an example of emotional state recognition. In this example, neutral, happy, angry and sad facial expressions are used as testing samples. In Fig. 4-17(a), fourteen dot marks represent the extracted feature points for facial expression recognition. The emotional intensities are obtained using (4.5)-(4.8). As shown in Fig. 4-17(a), the ratio of the neutral component amounts to 54%, which dominates the facial expression, although the other emotion components also contribute to the facial expression. Similar results are obtained as shown in Figs. 4-17(b)-(d).

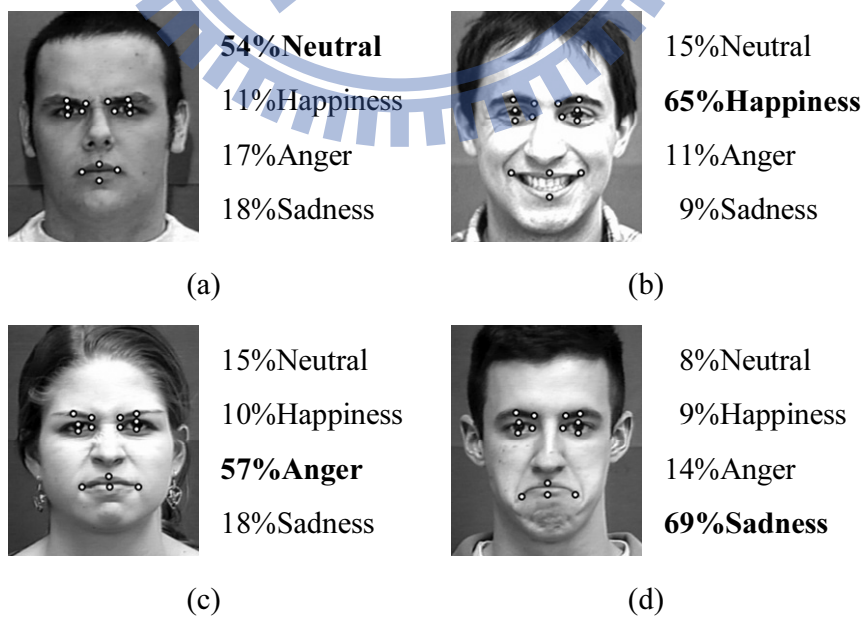
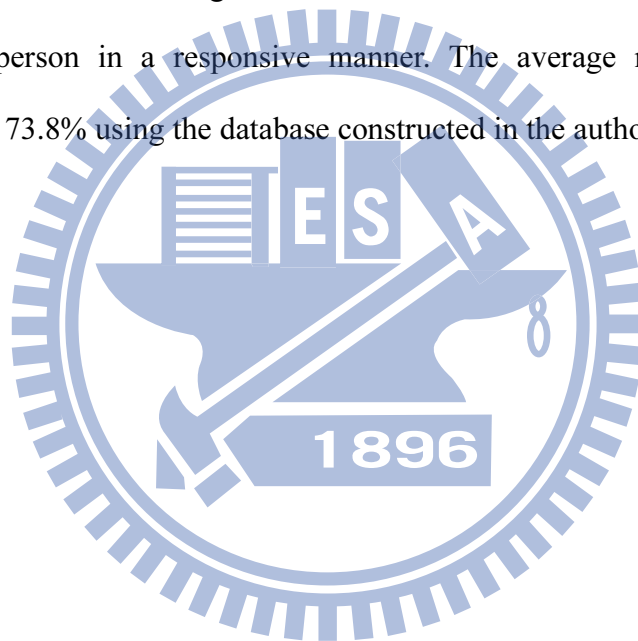


Fig. 4-17 Examples of user emotional state recognition.

4.5 Summary

In the part of robotic emotion generation, experimental results reveal that the simulated artificial face interacts with people in a manner of mood transition and with robotic personality. The questionnaire investigation confirms positive results on the evaluation of responsive robotic facial expressions generated by the proposed design. In the part of human emotion recognition, the experimental results of proposed bimodal emotion recognition system show that an average recognition rate of 86.9% is achieved, a 5% improvement compared to using only image information. On the other side, the experimental results of speech-signal-based emotion recognition for the entertainment robot show that the robot interacts with a person in a responsive manner. The average recognition rate for five emotional states is 73.8% using the database constructed in the authors' lab.



Chapter 5

Conclusions and Future Work

5.1 Dissertation Summary

In this work, a robotic mood transition model for autonomous emotional interaction has been developed. An emotional model is proposed for mood state transition exploiting a robotic personality approach. By adopting *Big Five* factors to represent robot personality in the 2-D emotional model, one is able to generate facial expressions in a more natural manner. The behavior fusion architecture with a designed rule table provides a robot the capability to generate emotional interactions. Experimental results on the artificial face show that the robot interacts with people with suitable mood transition and a kind of robotic personality. The questionnaire investigation confirms positive results on the evaluation of responsive robotic facial expressions generated by the proposed design.

For the bimodal information fusion algorithm, the proposed bimodal fusion scheme and statistically-determined fusion weights computed from individual modality effectively increase the recognition accuracy. Practical experiments have been carried out using a stand-alone robotic vision system. With a self-built database of fourteen persons, the proposed system achieves a recognition rate of 86.9%. For the proposed speech-signal-based emotion recognition, the emotion recognition system developed classifies five emotional categories, in real time. Experimental results using an entertainment robot show that the robot can interact with a user in a responsive manner, using the developed speech signal recognition system. Using a database built in the lab, the proposed system achieves an average recognition rate of 73.8% for five emotional states.

5.2 Future Directions

Some directions deserve further study in the future:

- 1) More comparisons with other emotional models will be further studied. It will be interesting to investigate different models for robotic emotion generation and evaluate their emotional intelligence with practical experiments.
- 2) For human emotion recognition, it is suggested to focus on the development of robust algorithms to deal with more natural visual and audio signals. Methods to extract more reliable features of both visual and audio modalities will also be investigated to improve the performance. The direct fusion of both visual and audio features is considered for the future to overcome the incomplete problem.
- 3) Because the voice signal must be acquired using the embedded system, it is difficult to establish a benchmark to evaluate the developed recognition algorithm. In the future, a method to extract key phrases in an utterance will be investigated, to increase the recognition rate. The emotional state can be estimated more directly from the speech signal, than from the extracted statistical features of the whole voice frame.
- 4) In this study, all participants in our experiment are aware of the test. It belongs to intrusive testing. Other types of testing can be studied in the future.

Appendix A

Evaluation Questionnaire of Emotional Interaction

Part I. Evaluation of Big Five Personality Traits

1. Openness:

Open mindedness, interest in culture.

	<u>Low openness</u> Conservative, lack of interests, non-artistic, non-analytic				<u>High openness</u> Curious, broad range of interests, creative, imaginative, nontraditional		
	-3	-2	-1	0	1	2	3
Robot A	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Robot B	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Robot C	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

2. Conscientiousness:

Organized, persistent in achieving goals.

	<u>Low Conscientiousness</u> Lack of goals, unreliable, lazy, careless, lax, hedonism, casual, lack of motivation at work				<u>High Conscientiousness</u> Organizational power, reliable, enthusiastic, self-regulation, punctuality, moral principles, orderly, enthusiasm, perseverance		
	-3	-2	-1	0	1	2	3
Robot A	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Robot B	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Robot C	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

3. Extraversion:

Preference for and behavior in social situations.

Low Extraversion

Reserved, indifference,
not enthusiastic, serious,
task-oriented, shy, quiet

High Extraversion

Highly social, active,
talkative, people-oriented,
optimistic, love-enjoyable,
kind

	-3	-2	-1	0	1	2	3
Robot A	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Robot B	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Robot C	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

4. Agreeableness:

Tend to be compassionate and cooperative rather than suspicious and antagonistic towards others

Low Agreeableness

Serious, rough, suspicious,
uncooperative, vengeful, ruthless,
irritable, hypocritical

High Agreeableness

Warm, good-natured, reliable, willing
to help, forgivable, willing to believe,
straight

	-3	-2	-1	0	1	2	3
Robot A	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Robot B	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Robot C	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

5. Neuroticism:

Tend to experience negative thoughts.

Low Neuroticism

Calm, relaxing, non-emotional,
courageous, safe, self-satisfied

High Neuroticism

Anxious, nervous, emotional,
insecure, non-adaptive,
depressed

	-3	-2	-1	0	1	2	3
Robot A	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Robot B	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Robot C	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Part II. Evaluation of Feasibility for Social Robot

1. Artificial vs. Natural

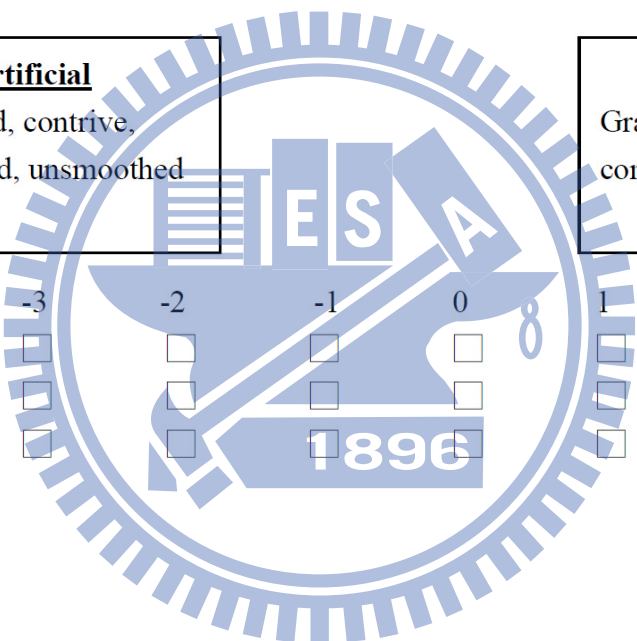
Artificial

Deliberated, contrive,
exaggerated, unsmoothed

Natural

Graceful, genuine,
comfortable, smooth

	-3	-2	-1	0	1	2	3
Robot A	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Robot B	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Robot C	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>



Bibliography

- [1] M. Fujita, "On Activating Human Communications with Pet-type Robot AIBO," *Proceedings of IEEE*, Vol. 92, No. 11, pp. 1804-1813, 2004.
- [2] H. H. Lund, "Modern Artificial Intelligence for Human-robot Interaction," *Proceedings of IEEE*, Vol. 92, No. 11, pp. 1821-1838, 2004.
- [3] S. G. Roh, K. W. Yang, J. H. Park, H. Moon, H. S. Kim, H. Lee, H. R. Choi, "A Modularized Personal Robot DRPI: Design and Implementation," *IEEE Trans. on Robotics*, Vol. 25, No. 2, pp. 414-425, 2009.
- [4] NEC's KOTOHANA Emotion communicator,
<http://thefutureofthings.com/pod/1042/necs-kotohana-emotion-communicator.html>
- [5] C. Breazeal, "Emotion and Sociable Humanoid Robots," *International Journal of Human-Computer Studies*, Vol. 59, pp. 119-155, 2003.
- [6] C. Breazeal, D. Buchsbaum, J. Gray, D. Gatenby and B. Blumberg, "Learning From and About Others: Towards Using Imitation to Bootstrap the Social Understanding of Others by Robots," *Journal of Artificial Life*, Vol. 11, pp.1-32, 2005.
- [7] MIT Media Lab, personal robot group,
<http://robotic.media.mit.edu/projects/robots/mds/headface/headface.html>
- [8] T. Wu, N. J. Butko, P. Ruvulo, M. S. Bartlett and J. R. Movellan, "Learning to Make Facial Expressions," in *Proc. of IEEE 8th International Conference on Development and Learning*, Shanghai, China, 2009, pp. 1-6.
- [9] N. Mavridis and D. Hanson, "The IbnSina Center: An Augmented Reality Theater with Intelligent Robotic and Virtual Characters," in *Proc. of IEEE 18th International Symposium on Robot and Human Interactive Communication*, Toyama, Japan, 2009, pp. 681-686.
- [10] N. Mavridis, A. AlDhaheri, L. AlDhaheri, M. Khanji and N. AlDarmaki, "Transforming IbnSina into an Advanced Multilingual Interactive Android Robot," in *Proc. of IEEE GCC Conference and Exhibition*, Dubai, United Arab Emirates, 2011, pp. 120-123.
- [11] T. Hashimoto, S. Hiramatsu, T. Tsuji and H. Kobayashi, "Realization and Evaluation of Realistic Nod with Receptionist Robot SAYA," in *Proc. of the 16th IEEE International Symposium on Robot and Human interactive Communication (RO-MAN 2007)*, Jeju Island, Korea, 2007, pp. 326-331.

- [12] T. Hashimoto, S. Hiramatsu, T. Tsuji and H. Kobayashi, "Development of the Face Robot SAYA for Rich Facial Expressions," in *Proc. of International Joint Conference on SICE-ICASE*, Pusan, Korea, 2006, pp. 5423-5428.
- [13] D. W. Lee, T. G. Lee, B. So, M. Choi, E. C. Shin, K. W. Yang, M. H. Back, H. S. Kim and H. G. Lee, "Development of an Android for Emotional Expression and Human Interaction," in *Proc. of International Federation of Automatic Control*, Seoul, Korea, 2008, pp. 4336-4337.
- [14] M. S. Siegel, "Persuasive Robotics: How Robots Change Our Minds," Massachusetts Institute of Technology, PhD Thesis, 2009.
- [15] N. Mavridis, M. Petychakis, A. Tsamakos, P. Toulis, S. Emami, W. Kazmi, C. Datta, C. BenAbdelkader and A. Tanoto, "FaceBots: Steps Towards Enhanced Long-Term Human-Robot Interaction by Utilizing and Publishing Online Social Information," *Springer Paladyn Journal of Behavioral Robotics*, Vol. 1, No. 3, pp. 169-178, 2011.
- [16] H. Miwa, T. Okuchi, K. Itoh, H. Takanobu and A. Takanishi, "A New Mental Model for Humanoid Robots for Human Friendly Communication," in *Proc. of IEEE International Conference on Robotics and Automation*, Taipei, Taiwan, 2003, pp. 3588-3593.
- [17] H. Miwa, K. Itoh, M. Matsumoto, M. Zecca, H. Takanobu, S. Rocella, M.C. Carrozza, P. Dario and A. Takanishi, "Effective Emotional Expressions with Emotion Expression Humanoid Robot WE-4RII," in *Proc. of the 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Sendai, Japan, 2004, pp. 2203-2208.
- [18] D. Duhaut, "A Generic Architecture for Emotion and Personality," in *Proc. of IEEE International Conference on Advanced Intelligent Mechatronics*, Xi'an, China, 2008, pp. 188-193.
- [19] L. Moshkina, S. Park, R. C. Arkin, J. K. Lee and H. Jung, "TAME: Time-Varying Affective Response for Humanoid Robots," *International Journal of Social Robotics*, Vol. 3, pp.207-221, 2011.
- [20] C. Itoh, S. Kato and H. Itoh, "Mood-transition-based Emotion Generation Model for the Robot's Personality," in *Proc. of IEEE International Conference on Systems, Man and Cybernetics*, St Antonio, TX, USA, 2009, pp. 2957-2962.
- [21] S. C. Banik, K. Watanabe, M. K. Habib and K. Izumi, "An Emotion-Based Task Sharing Approach for a Cooperative Multiagent Robotic System," in *Proc. of IEEE International Conference on Mechatronics and Automation*, Kagawa, Japan, 2008, pp. 77-82.
- [22] J. C. Park, H. R. Kim, Y. M. Kim and D. S. Kwon, "Robot's Individual Emotion Generation Model and Action Coloring According to the Robot's Personality," in *Proc. of*

IEEE International Symposium on Robot and Human Interactive Communication, Toyama, Japan, 2009, pp. 257-262.

- [23] H. R. Kim and D. S. Kwon, "Computational Model of Emotion Generation for Human-Robot Interaction Based on the Cognitive Appraisal Theory," *International Journal of Intelligent and Robotic Systems*, Vol. 60, pp. 263-283, 2010.
- [24] D. Lee, H. S. Ahn and J. Y. Choi, "A General Behavior Generation Module for Emotional Robots Using Unit Behavior Combination Method," in *Proc. of IEEE International Symposium on Robot and Human Interactive Communication*, Toyama, Japan, 2009, pp. 375-380.
- [25] M. J. Han, C. H. Lin and K. T. Song, "Autonomous Emotional Expression Generation of a Robotic Face," in *Proc. of IEEE International Conference on Systems, Man and Cybernetics*, St Antonio, TX, USA, 2009, pp. 2501-2506.
- [26] Y. Tian, T. Kanade and J.F. Cohn, "Recognizing Action Units for Facial Expression Analysis," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 23, No. 2, pp. 97-115, 2001.
- [27] M. Pantic and L.J.M. Rothkrantz, "Automatic Analysis of Facial Expressions: The State of the Art," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 22, No. 12, pp. 1424-1445, 2000.
- [28] D. Ververidis, C. Kotropoulos and I. Pitas, "Automatic Emotional Speech Classification," in *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing*, Montreal, Quebec, Canada, 2004, pp. 593-596.
- [29] B. Schuller, G. Rigoll and M. Lang, "Speech Emotion Recognition Combining Acoustic Features and Linguistic Information in a Hybrid Support Vector Machine - Belief Network Architecture," in *Pro. of IEEE International Conference on Acoustics, Speech, and Signal Processing*, Montreal, Quebec, Canada, 2004, Vol. 1, pp. 577-580.
- [30] P. S. Aleksic and A. K. Katsaggelos, "Audio-Visual Biometrics," *Proceedings of IEEE*, Vol. 94, No. 11, pp. 2025-2044, 2006.
- [31] L. C. De Silva, T. Miyasato and R. Nakatsu, "Facial Emotion Recognition Using Multi-modal Information," in *Proc. of IEEE International Conference on Information, Communications and Signal Processing*, Singapore, 1997, pp. 397-401.
- [32] L. C. De Silva, "Audiovisual Emotion Recognition," in *Proc. of IEEE International Conference on Systems, Man and Cybernetics*, The Hague, The Netherlands, 2004, pp. 649-654.

- [33] H. J. Go, K. C. Kwak, D. J. Lee and M. G. Chun, "Emotion Recognition from the Facial Image and Speech Signal," in *Proc. of SICE Annual Conference*, Fukui, Japan, 2003, pp. 2890-2895.
- [34] Y. Wang and L. Guan, "Recognizing Human Emotion from Audiovisual Information," in *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing*, Philadelphia, PA, USA, 2005, pp. 1125-1128.
- [35] J. C. Platt, *Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods*, MIT Press, Cambridge, MA, 2000.
- [36] O. W. Kwon, K. Chan, J. Hao and T. W. Lee, "Emotion Recognition by Speech Signals," in *Proc. of 8th European Conference on Speech Communication and Technology*, Geneva, Switzerland, 2003, pp. 55 125-128.
- [37] T. L. Nwe, S. W. Foo, and L. C. De Silva, "Speech Emotion Recognition Using Hidden Markov Models," *Speech Communication*, Vol. 41, No. 4, pp. 603-623, 2003.
- [38] K. H. Hyun, E. H. Kim, and Y. K. Kwak, "Improvement of Emotion Recognition by Bayesian Classifier Using Non-zero-pitch Concept," in *Proc. of IEEE International Workshop on Robot and Human Interactive Communication*, Nashville, USA, 2005, pp. 312-316.
- [39] T. L. Pao and Y. T. Chen, "Mandarin Emotion Recognition in Speech," in *Proc. of IEEE Workshop on Automatic Speech Recognition and Understanding*, St. Thomas, Virgin Islands, 2003, pp. 227-230.
- [40] D. Neiberg, K. Elenius and K. Laskowski, "Emotion Recognition in Spontaneous Speech Using GMMs," in *Proc. of International Conference on Spoken Language Processing*, Pittsburgh, Pennsylvania, USA, 2006, pp. 809-812.
- [41] M. You, C. Chen, J. Bu, J. Liu and J. Tao, "Emotional Speech Analysis on Nonlinear Manifold," in *Proc. of IEEE International Conference on Pattern Recognition*, Hong Kong, China, 2006, pp. 91-94.
- [42] M. You, C. Chen, J. Bu, J. Liu and J. Tao, "A Hierarchical Framework for Speech Emotion Recognition," in *Proc. of IEEE International Symposium on Industrial Electronics*, Montreal, Quebec, Canada, 2006, pp. 515-519.
- [43] Z. J. Chuang and C. H. Wu, "Emotion Recognition Using Acoustic Features and Textual Content," in *Proc. of IEEE International Conference on Multimedia and Expo*, Taipei, Taiwan, 2004, pp. 53-56.
- [44] C. Busso, S. Lee and S. Narayanan, "Analysis of Emotionally Salient Aspects of

- Fundamental Frequency for Emotion Detection,” *IEEE Trans. on Audio, Speech, and Language Processing*, Vol. 17, No. 4, pp. 582-596, 2009.
- [45] B. Yang and M. Lugger, “Emotion Recognition from Speech Signals Using New Harmony Features,” *Signal Processing*, Vol. 90, No. 5, pp. 1415-1423, 2010.
- [46] C. Li, Q. Zhou, J. Cheng, X. Wu and Y. Xu, “Emotion Recognition in a Chatting Robot,” in *Proc. of 2008 IEEE International Conference on Automation and Logistics*, Qingdao, China, 2008, pp. 1452-1457.
- [47] E. H. Kim, K. H. Hyun, S. H. Kim and Y. K. Kwak, “Improved Emotion Recognition with a Novel Speaker-independent Feature,” *IEEE Trans. on Mechatronics*, Vol. 14, No. 3, pp. 317-325, 2009.
- [48] J. S. Park, J. H. Kim and Y. H. Oh, “Feature Vector Classification Based Speech Emotion Recognition for Service Robots,” *IEEE Trans. on Consumer Electronics*, Vol. 55, No. 3, pp. 1590-1596, 2009.
- [49] K. T. Song, M. J. Han and J. W. Hong, “Online Learning Design of an Image-Based Facial Expression Recognition System,” *Intelligent Service Robotics*, Vol. 3, No. 3, pp. 151-162, 2010.
- [50] M. A. Amin and H. Yan, “Expression Intensity Measurement from Facial Images by Self Organizing Maps,” in *Proc. of IEEE International Conference on Machine Learning and Cybernetics*, Kunming, 2008, pp. 3490-3496.
- [51] M. Beszedes and P. Culverhouse, “Comparison of Human and Automatic Facial Emotions and Emotion Intensity Levels Recognition,” in *Proc. of IEEE International Symposium on Image and Signal Processing and Analysis*, Istanbul, Turkey, 2007, pp. 429-434.
- [52] M. Oda and K. Isono, “Effects of Time Function and Expression Speed on the Intensity and Realism of Facial Expressions,” in *Proc. of IEEE International Conference on Systems, Man and Cybernetics*, Singapore, 2008, pp. 1103-1109.
- [53] K. K. Lee and Y. Xu, “Real-time Estimation of Facial Expression Intensity,” in *Proc. of IEEE International Conference on Robotics and Automation*, Taipei, Taiwan, 2003, pp. 2567-2572.
- [54] K. T. Song and J. Y. Lin, “Behavior Fusion of Robot Navigation Using a Fuzzy Neural Network,” in *Proc. of IEEE International Conference on Systems, Man and Cybernetics*, Taipei, Taiwan, 2006, pp. 4910-4915.
- [55] Grimace project, available online at: <http://grimace-project.net/>

- [56] P. Ekman and W. V. Friesen, *The Facial Action Coding System: A Technique for The Measurement of Facial Movement*, Consulting Psychologists Press, San Francisco, 1978.
- [57] D. G. Myers, *Theories of Emotion*, NY: Worth Publishers, New York, 2004.
- [58] S. F. Locke, MIT Meter Measures the Mood of Passers-By, available online at: <http://www.popsoci.com/technology/article/2011-11/mit-meter-measures-mood-passers>
- [59] R. R. McCrae and P. T. Costa, "Validation of the Five Factor Model of Personality across Instruments and Observers," *Journal of Personality and Social Psychology*, Vol. 51, 81-90, 1987.
- [60] P. T. Costa and R. R. McCrae, "Normal Personality Assessment in Clinical Practice: The NEO Personality Inventory," *Journal of Psychological Assessment*, Vol. 4, 5-13, 1992.
- [61] A. Mehrabian, "Analysis of the Big-five Personality Factors in Terms of the PAD Temperament Model," *Australian Journal of Psychology*, Vol. 48, No. 2, pp. 86-92, 1996.
- [62] L. R. Goldberg, "The Development of Markers for the Big-Five Factor Structure," *Psychological Assessment*, Vol. 4, pp.26-42, 1992.
- [63] J. A. Russell and G. Pratt, "A Description of the Affective Quality Attributed to Environments," *Journal of Personality and Social Psychology*, Vol. 38, No. 2, 311-322, 1980.
- [64] J. A. Russell and M. Bullock, "Multidimensional Scaling of Emotional Facial Expressions: Similarity from Preschoolers to Adults," *Journal of Personality and Social Psychology*, Vol. 48, 1290-1298, 1985.
- [65] J. A. Russell, "A Circumplex Model of Affect," *Journal of Personality and Social Psychology*, Vol. 39, No. 6, 1161-1178, 1980.
- [66] Frederick Jelinek, *Statistical Methods for Speech Recognition*, MIT Press, Cambridge, MA, 1999.
- [67] N. Christianini and J.S. Taylor, *An Introduction to Support Vector Machines*, MIT Press, Cambridge, MA, 2000.
- [68] P. Viola and M. Jones, "Rapid Object Detection Using a Boosted Cascade of Simple Features," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, Kauai Marriott, Hawaii, 2001, pp. 511-518.
- [69] J. H. Lai, P. C Yuen, W. S. Chen, S. Lao and M. Kawade, "Robust Facial Feature Point Detection Under Nonlinear Illuminations," in *Proc. of IEEE ICCV Workshop on Recognition, Analysis and Tracking of Faces and Gestures in Real-time Systems*,

- Vancouver , Canada, 2001, pp.168-174.
- [70] M. J. Han, J. H. Hsu, K. T. Song, and F. Y. Chang, “A New Information Fusion Method for SVM-based Robotic Audio-visual Emotion Recognition,” in *Proc. of IEEE International Conference on Systems, Man and Cybernetics*, Montreal, Canada, 2007, pp. 2656-2661.
- [71] H. C. Kim, D. J. Kim and S. Y. Bang, “Face Recognition Using LDA Mixture Model,” in *Proc. of IEEE International Conference on Pattern Recognition*, Quebec, Canada, 2002, pp. 925-928.
- [72] K. M. Yan, “Development of A Home Robot Speech Recognition System,” National Chiao Tung University, Master Thesis, 2002.
- [73] B. Gold and N. Morgan, *Speech and Audio Signal Processing: Processing and Perception of Speech and Music*, John Wiley & Sons, New York, USA, 2000.
- [74] S. Mika, G. Ratsch, J. Weston, B. Scholkopf and K. R. Muller, “Fisher Discriminant Analysis with Kernels,” in *Proc. of IEEE International Workshop on Neural Networks for Signal Processing*, Madison, WI, USA, 1999, pp. 41-48. *Nature of Statistic Learning Theory*
- [75] V. Vapnik, *The Nature of Statistical Learning Theory*, Springer, New York, USA, 1995.
- [76] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*, Wiley, New York, USA, 1973.
- [77] H. Andrian and K. T. Song, “Embedded CMOS Imaging System for Real-Time Robotic Vision,” in *Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems*, Edmonton, Alberta, Canada, 2005, pp. 3694-3699.
- [78] TMS320C6416 DSK technical reference, available online at:
http://c6000.spectrumdigital.com/dsk6416/V1/docs/dsk6416_TechRef.pdf
- [79] TLV320AIC23B data manual, available online at:
<http://www.ti.com/lit/ds/symlink/tlv320aic23b.pdf>
- [80] Pololu serial 8-servo controller, available online at:
<http://www.pololu.com/products/pololu/0727/>
- [81] J. M. Valin, S. Yamamoto, J. Rouat, F. Michaud, K. Nakadai and H. G. Okuno, “Robust Recognition of Simultaneous Speech by A Mobile Robot,” *IEEE Trans. on Robotics*, Vol. 23, No. 4, pp. 742-752, 2007.
- [82] H. Nakajima, K. Nakadai, Y. Hasegawa and H. Tsujino, “Blind Source Separation with

Parameter-free Adaptive Step-size Method for Robot Audition,” *IEEE Trans. on Audio, Speech, and Language Processing*, Vol. 18, No. 6, pp. 1476-1485, 2010.

- [83] Qwerk Platform, Charmed Labs, available online at:
http://www.charmedlabs.com/index.php?option=com_content&task=view&id=29
- [84] <http://isci.cn.nctu.edu.tw/video/AnthropomorphicRobot/>
- [85] <http://isci.cn.nctu.edu.tw/video/RoboticMoodTransition/>
- [86] <http://isci.cn.nctu.edu.tw/video/RoboticMoodTransitionAnalysis/>
- [87] M. E. Hoque, R. E. Kaliouby and R. W. Picard, “When Human Coders (and Machines) Disagree on the Meaning of Facial Affect in Spontaneous Videos,” in *Proc. of 9th International Conference on Intelligent Virtual Agents*, Amsterdam, Netherlands, 2009, pp. 337-343.
- [88] M. E. Hoque, L-P. Morency and R. W. Picard, “Are You Friendly or Just Polite? – Analysis of Smiles in Spontaneous Face-to-face Interactions,” in *Proc. of 4th International Conference on Affective Computing and Intelligent Interaction*, Memphis, TN, USA, 2011, pp. 135-144.
- [89] M. E. Hoque and R. W. Picard, “Acted vs. Natural Frustration and Delight: Many People Smile in Natural Frustration,” in *Proc. of IEEE 9th International Conference on Automatic Face and Gesture Recognition*, Santa Barbara, CA, USA, 2011, pp. 354-359.
- [90] M. Pantic, Michel Valstar, R. Rademaker and L. Maat, “Web-based Database for Facial Expression Analysis,” in *Proc. of IEEE International Conference on Multimedia and Expo*, Amsterdam, 2005, pp. 317-321.
- [91] O. Martin, I. Kotsia, B. Macq and I. Pitas, “The eNTERFACE'05 Audio-Visual Emotion Database,” in *Proc. of IEEE International Conference on Data Engineering Workshops*, Atlanta, 2006.
- [92] Intelligent System Control Integration Laboratory, Emotional Utterance Voice Clip, available online at: <http://isci.cn.nctu.edu.tw/JCIE/VoiceClip/>
- [93] Intelligent System Control Integration Laboratory, Experimental Video Clip, available online at: <http://isci.cn.nctu.edu.tw/JCIE/VideoClip/>
- [94] Cohn-Kanade AU-Coded Facial Expression Database, available online at:
<http://www.pitt.edu/~jeffcohn/CKandCK+.htm>

Vita

姓名：韓孟儒

性別：男

生日：中華民國 65 年 7 月 21 日

籍貫：台北市



論文題目：中文：機器人情感模型及情感辨識設計

英文：Design of Robotic Emotion Model and Human

Emotion Recognition

學/經歷：

1. 民國 87 年 6 月 國立台北科技大學電機工程技術系畢業
2. 民國 92 年 6 月 國立中興大學電機工程研究所畢業
3. 民國 92 年 9 月 國立交通大學電控工程研究所博士班
4. 民國 99 年 11 月 工業技術研究院機械與系統研究所副研究員

Publication List

Journal paper

- [1] Meng-Ju Han, Chia-How Lin and Kai-Tai Song, “Robotic Emotional Expression Generation Based-on Mood Transition and Personality Model,” *IEEE Transactions on Systems, Man and Cybernetics, Part B*, to appear, November 5, 2012 (Accepted).
- [2] Kai-Tai Song, Meng-Ju Han and Shih-Chieh Wang, “Speech-Signal-Based Emotion Recognition and Its Application to Entertainment Robots,” *Journal of the Chinese Institute of Engineers*, to appear, September 20, 2012 (Accepted).
- [3] Kai-Tai Song, Meng-Ju Han and Jung-Wei Hong, “Online Learning Design of an Image-Based Facial Expression Recognition System,” *Intelligent Service Robotics*, Vol. 3, No. 3, pp. 151-162, 2010.
- [4] Meng-Ju Han, Jing-Huai Hsu, Kai-Tai Song and Fuh-Yu Chang, “A New Information Fusion Method for Bimodal Robotic Emotion Recognition,” *Journal of Computers*, Vol. 3, No. 7, pp. 39-47, 2008.

Patent

- [1] 宋開泰、韓孟儒、王仕傑、林家合、林季誼，“表情檢測裝置及其表情檢測方法”，中華人民共和國發明專利證書號：ZL 2009 1 0141299.1.
- [2] 宋開泰、韓孟儒、許晉懷、洪濬尉、張復瑜，“情緒辨識與對新辨識資訊之學習方法”，台灣發明專利證書號：I365416.
- [3] Kai-Tai Song, Meng-Ju Han, Jing-Huai Hsu, Jung-Wei Hong and Fuh-Yu Chang, “Method of Emotion Recognition,” 美國專利公開號：20080201144.
- [4] 陳豪宇、韓孟儒、吳至仁、林泓宏、康哲儒、楊谷洋、宋開泰、蔡文祥、莊仁輝，“移動式取像系統及其控制方法”，台灣專利公開號：200905617.
- [5] 宋開泰、韓孟儒、王仕傑、林家合、林季誼，“表情偵測裝置及其表情偵測方法”，台灣專利公開號：201039251；美國專利公開號：20100278385.
- [6] 宋開泰、韓孟儒、王仕傑、江銘峰、林家合，“人臉偵測裝置及其人臉偵測方法”，台灣專利公開號：201040846；中國專利申請號：200910141418.3；美國專利公開號：20100284619.
- [7] 宋開泰、韓孟儒、林嘉豪，“機器人自主情感表現裝置以及表現機器人自主情感之方

法”，台灣專利公開號：201123036；美國專利公開號：20110144804.

- [8] 宋開泰、韓孟儒、王仕傑，“人臉辨識方法及應用此方法之系統”，台灣專利公開號：201123030；美國專利公開號：20110150301.

Conference paper

- [1] Meng-Ju Han, Chia-How Lin and Kai-Tai Song, “A Design for Smooth Transition of Robotic Emotional States,” in *Proc. of IEEE International Conference on Advanced Robotics and Its Social Impacts*, Seoul, Korea, 2010, pp. 13-18.
- [2] Yi-Wen Chen, Meng-Ju Han, Kai-Tai Song and Yu-Lun Ho, “Image-Based Age-Group Classification Design Using Facial Features,” in *Proc. of IEEE International Conference on System Science and Engineering*, Taipei, Taiwan, 2010, pp. 548-552.
- [3] Kai-Tai Song, Shih-Chieh Wang, Meng-Ju Han and Ching-Yi Kuo, “Pose-Variant Face Recognition Based on an Improved Lucas-Kanade Algorithm,” in *Proc. of IEEE International Conference on Advanced Robotics and Its Social Impacts*, Tokyo, Japan, 2009, pp. 87-92.
- [4] Meng-Ju Han, Chia-How Lin and Kai-Tai Song, “Autonomous Emotional Expression Generation of a Robotic Face,” in *Proc. of IEEE International Conference on Systems, Man and Cybernetics, San Antonio, Texas, USA, 2009*, pp. 2501-2506.
- [5] Kai-Tai Song, Meng-Ju Han, Fu-Hua Jen and Jen-Chao Tai, “Facial Expression Recognition and Its Application to Emotional Interaction of a Robotic Head,” in *Proc. of the 10th International Conference on Automation Technology*, Tainan, Taiwan, 2009, pp. 493-498.
- [6] Kai-Tai Song, Meng-Ju Han and Shuo-Hung Chang, “Pose-Variant Facial Expression Recognition Using an Embedded Image System,” in *Proc. of International Symposium on Precision Mechanical Measurements*, Hefei, China, 2008.
- [7] Shih-Chieh Wang, Meng-Ju Han and Kai-Tai Song, “Human Emotion Recognition of a Pet Robot Using Natural Speech Information,” in *Proc. of National Symposium on System Science and Engineering*, Ilan, Taiwan, 2008.
- [8] Meng-Ju Han, Jing-Huai Hsu, Kai-Tai Song and Fuh-Yu Chang, “A New Information Fusion Method for SVM-Based Robotic Audio-Visual Emotion Recognition,” in *Proc. of IEEE International Conference on Systems, Man and Cybernetics*, Montreal, Canada, 2007, pp. 2656-2661.
- [9] Kai-Tai Song, Meng-Ju Han, Fuh-Yu Chang and Shuo-Hung Chang, “A Robotic Facial

Expression Recognition System Using Real-Time Vision System,” in *Proc. of the 8th International Conference on Measurement Technology and Intelligent Instruments*, Sendai, Japan, 2007, pp. 63-66.

- [10] Jung-Wei Hong, Meng-Ju Han, Kai-Tai Song and Fuh-Yu Chang, “A Fast Learning Algorithm for Robotic Emotion Recognition,” in *Proc. of the 7th IEEE International Symposium on Computational Intelligence in Robotics and Automation*, Jacksonville, Florida, USA, 2007, pp. 25-30.
- [11] Meng-Ju Han, Jing-Huai Hsu, Kai-Tai Song and Fuh-Yu Chang, “Embedded Emotion Recognition System Using Key Feature Sets,” in *Proc. of 2006 CACS Automatic Control Conference*, Taipei, Taiwan, 2006, pp. 1048-1053.
- [12] Meng-Ju Han and Kai-Tai Song, “Block-Based Motion Detection for Lateral Driving Assistance,” in *Proc. of 2005 CACS Automatic Control Conference*, Tainan, Taiwan, 2005, pp. I-Three 145-150.

Domestic Journal Publications

- [1] Kai-Tai Song and Meng-Ju Han, “On Robot Locomotion,” *Scientific People*, No. 57, pp. 76-79, Nov. 2006. (in Chinese)

Honors

- [1] 韓孟儒, 林振暘, 林志昇, 吳巧敏, 江信毅, 孫宗暘, “勇猛保鏢,” 2008 智慧型機器人產品創意競賽 - 新光保全組：第一名。
- [2] 王仕傑, 韓孟儒, 宋開泰, “Human Emotion Recognition of a Pet Robot Using Natural Speech Information,” 2008 中華民國系統科學工程會議論文競賽：優等獎。
- [3] 韓孟儒, 王仕傑, “具表情辨識功能之寵物機器人,” 2007 受邀於韓國浦項(Pohang)「9th Intelligent Robot Contest」展示。