

## Research on Multimodal Emotion Recognition Platform Construction

---

### Wenqiang Liu<sup>1</sup>

*Brain Cognitive Computing Lab, School of Information Engineering, Minzu University of China  
Beijing, 100081, China  
Email: liuwqleo@163.com*

### Huiping Jiang<sup>2</sup>

*Brain Cognitive Computing Lab, School of Information Engineering, Minzu University of China  
Beijing, 100081, China  
Email: jianghp@muc.edu.cn*

### Yao Lu<sup>3</sup>

*Brain Cognitive Computing Lab, School of Information Engineering, Minzu University of China  
Beijing, 100081, China  
Email: 5206661314@163.com*

With the development of BCI technology, simple non-behavioral human-computer interaction has been achieved. However, due to the low signal-to-noise ratio and time-varying nature of EEG signals, the existing signal processing algorithms is inadequate for online applications. This has become the bottleneck that restricts the development of man-machine integrated system based on brain-computer interface technology. We realized that a single state of mind measurement technology would be very difficult to achieve efficient brain-computer integrated system, therefore, the multi-modality measurement fusion technology is an inevitable trend. At present, researches on affective recognition are mainly based on single signal, while the multi-modal emotion recognition combining with multiple signals needs to be further studied. This article describes a multi-modal synchronous acquisition cognitive experiment platform. The platform uses three independent signal acquisition and analysis systems of EEG, eye movement and expression monitoring, and realizes the accurate synchronization of three types of signals based on EEG with the self-developed synchronization module. The EEG, eye movement and facial expression information can be effectively and efficiently standardized to form a large-scale multi-ethnic cognitive database so as to facilitate the standardization of cognitive experiments, especially affective recognition.

*ISCC 2017  
16-17 December 2017  
Guangzhou, China*

---

<sup>1</sup>Speaker

<sup>2</sup>Corresponding Author Acknowledgements. Supported by National Natural Science Foundation of China Youth Project Funding (No: 61503423), the Leading Talent Program of State Ethnic Affairs Commission, and Double First-class Special Funding of MUC.

<sup>3</sup>The Second Author

## 1. Introduction

### 1.1 Emotion Recognition

As the emotion is an important message that people convey in the process of communication and exchange, the emotional state affects emotion as an important message that people convey in the process of communication and exchange. The emotional state changes affect the people's perception and decision-making. The emotion recognition is an important research field of pattern recognition, which introduces the affective dimension into human-computer interaction. The emotional expression modalities include facial expressions, speech, gestures, physiological signals and words, etc. The emotion recognition is an essential a multimodal fusion problem.

The emotional state has a significant effect on human cognition. With the development of computer science and computer science, the emotional computing came into being. Professor Picard [1] first put forward the calculation of emotion. She defined the calculation of emotion as "the calculation of the factors that are related to emotion, caused by emotion or can affect emotion." The goal of emotional computing is to enable the computer capable of expressing emotions, emotion recognition and emotion understanding. Among them, the emotion recognition is the basis of emotion calculation and the key point of emotional cognition research.

The BCI is a device that relies on muscle communication and control to transmit information and commands between the brain and the outside world [2]. The purpose of the system is to uncover the rules of specific brain thinking activities and establish the communication of specific, relevant information between the intent and behavior of brain-thinking activities. The purpose of brain computer interface (BCI) is to establish the connection between brain thinking activities' intention and behavior, realize the direct interaction and control of human-computer information and have a broad application prospect in brain cognition and biofeedback training. With non-linear, non-stationary and vulnerable to the outside interference , other characteristics of the brain reflect in an integrated manner the electrical activity and brain functional status of the carrier with high temporal resolution and real-time response.

The emotional recognition has always been an important direction in artificial intelligence research. With the development of science and detection technology, people recognize that many human physiological indexes can be applied to the research of affective computation. Recognition based on non-physiological signals and recognition based on physiological signals are two major methods of emotion recognition. In the current researches on affective recognition, most researches focus on expression, speech, attitude or single physiological signals (such as ECG, EMG, respiration, etc.). Non-physiological signals mainly include facial expressions and phonetic intonation. Although this method is simple and easy to use, its reliability is low and its authenticity is poor; therefore, the physiological signal becomes a relatively objective indicator for emotion measurement. For EEG signals, in comparison with other non-physiological signals, the authenticity of emotion can be easily reflected and the recognition rate is high. Therefore, EEG has been widely applied to the field of emotion calculation. Some researches have achieved some results. For example, when the emotion recognition is performed by expression, Gwen Littlewort [3] et al. adopted the Gabor transform and multi-classifier fusion method, the recognition rate can reach as high as 91.5%; in the use of

EEG signal recognition, the recognition rate can reach 80% [4]. However, how to combine these different emotional signals to exert their respective advantages and establish a higher recognition rate system remains to be further studied.

### 1.2 Emotion Recognition Based on Facial Expression

Face expression recognition (FER) to be studied is how to automatically, reliably and efficiently use the facial expression of the message. Facial expression recognition is also an important part of the theoretical study of artificial psychology. Human face expression is rich in human behavior information. Its research can further understand the human corresponding psychological state.

Face recognition mainly depends on facial expression information contained in the facial expression information, and the use of computer facial expression information feature extraction and analysis, according to people's understanding and thinking to be classified and understood, and then from the face information to analyze the people's emotions. Commonly used identification methods include global and local identification technology, deformation extraction and motion extraction, geometric features and facial features, etc.. Luo Fei [5] and others, on the basis of Gabor wavelet feature facial expression recognition research, validated the facial expression recognition system, which can more accurately identify the facial expression and the highest accuracy rate of 80%.

### 1.3 Eye-based Emotion Recognition

Eye movement refers to the movement of the eye, including three basic types: gaze, eye movements and following the movement. Based on the analysis of such indicators as fixation time, fixation frequency, eye latency, eye movement distance and other indicators, we can obtain cognitive characteristics such as the information selection mode of visual processing of testers. Studies have shown that eye movements can not only record the cognitive process and the level of eye movement characteristics, but also a physiological indicator of emotion. Therefore, the eye movement analysis is of great importance in the recognition of emotional images and emotion recognition.

Feng Xiaohang [6] and others used eye movement recording technology to explore the four basic emotional eye movement characteristics. They found that the students identified the basic emotional happiness, fear, sadness and anger. The total time of fixation was significantly different to identify anger and sad mood. The total fixation time as required is longer while the total fixation time for fear identification is the shortest.

### 1.4 EEG-based Emotion Recognition

EEG (EEG) is produced in the cerebral cortex. Different thoughts and emotional states reflect different EEG signals. EEG signals were collected through pretreatment, feature extraction, feature selection and classification steps to obtain emotional recognition.

Rezz Khosrowabadi et al. [8] designed an Affective Brain-Computer Interface (ABCI) that can extract four kinds of features. EEG signals from 10 healthy subjects were collected, and the features were extracted by using Kernel Density Estimate and Gaussian Mixture Model. Bayesian network and random tree were used to classify the EEG signals. The results show that ABCI uses Gaussian Classification model extraction features and One-Rule algorithm for classification to achieve the best results and the recognition rate as high as 90%. The neural networks are generally divided into feedforward type, feedback type, self-organizing type and

random type. Xie Songyun, Guo Rong and Zhao Yingjun published EEG classification method based on neural network [8] used BP neural network and PNN neural network to classify the eigenvectors when eyes were open and closed, the correct rate of BP is 77% with PNN recognition rate of 73%, indicating that the BP model is better than the PNN neural network model.

### 1.5 Multi-modal Emotion Recognition

Although single mode recognition emotion has achieved good effect, it still has some limitations. It is difficult to separate some similar affection pairs, for example, when the voice information is only used, it is very difficult to separate angry and surprised due to their similarities in acoustic characteristics caused. At present, more and more researches begin to shift their attention to multi-modal emotion recognition and achieve some results. Wagner et al [9] achieved a 92% fusion recognition rate by integrating the physiological parameters of four channels of electromyography, electrocardiogram, skin resistance and respiration. Southeast University, Huang Cheng Wei et al [10] fused voice signals and ECG signals to establish multi-modal classifier recognition rate as high as 90%.

In order to achieve a better recognition effect, we can explore a multi-modal emotion recognition method that combines more signals. By using the existing laboratory conditions, a multimodal cognitive platform integrating three expressions of facial expression, eye movement and brain electricity was constructed to recognize the emotion. In order to achieve expression, eye tracking and multi-modal integration of EEG, the experimental process of the three instruments was recorded in the expression of video, eye movements and multi-channel EEG signals for accurate corresponding to or synchronizing.

## 2. Synchronous Cognitive Platform

### 2.1 EEG Acquisition System

EEG acquisition system by using the United States NeuroScan's 64-lead EEG / ERP EEG recorder, consists of SynAmps electrophysiological amplifier and acquisition and processing software SCAN. SynAmp uses editable software for great flexibility of data acquisition. The design's start-up and time-lock functions can ensure that all channels in the system have no phase deviation. All channels use high-quality 24-bit A/D analog digital chips to process the data to be transferred to the SCAN computer over the USB 2.0 interface. Its electrode cap adopts the 10-20 system electrode placement method as stipulated by the International Electro-EEG while simultaneously acquiring a sampling rate of 20,000Hz per guide to ensure that all channels can be collected synchronously.

SCAN includes acquisition and analysis of two parts with over 25 online functions and 50 data processing tools, including statistical analysis, topographic maps, filtering, correction of artifacts (multiple methods), principal component analysis and independent component analysis.

### 2.2 Eye Movement Acquisition System

The eye movement acquisition system uses the iView X Hi-Speed 1250 system, which includes an iView PC test computer, a Stimulus PC image display computer and a test tray. The system combines advanced camera technology and eye trimming algorithms to measure horizontal and vertical eye movements and pupil size, and is ergonomically designed to achieve a better man-machine interface. The hi-speed eye tracker features high sampling rate (1250 Hz)

to ensure a higher accuracy and suitability for reading research and other fields requiring more precision.

During the experiment, the subject placed the chin on the bracket with the infrared light source and the camera installed, and looked at the picture on the screen of the computer in front of the Stimulus PC. The testers control the images displayed on the Stimulus PC through the iView PC test computer, capture and analyze the subject's eye movement data. The recorded data allows automatic online analysis and can also be exported to ASCII data format. It is easy to use standard processing software post-processing. The combined BeGaze software allows for further data processing.

### **2.3 Facial Expression Video Capture System**

The system used two cameras to capture the experimental video data. One camera was placed on the wall, the entire laboratory environment was monitored and the other camera was placed on the computer in front of the subjects to record the subjects' facial expressions. The system software used VC ++ 6.0 + and OpenCV1.0 to develop and the acquisition card function library provided by Daheng to save the two videos in avi format.

### **2.4 Synchronization Module Design**

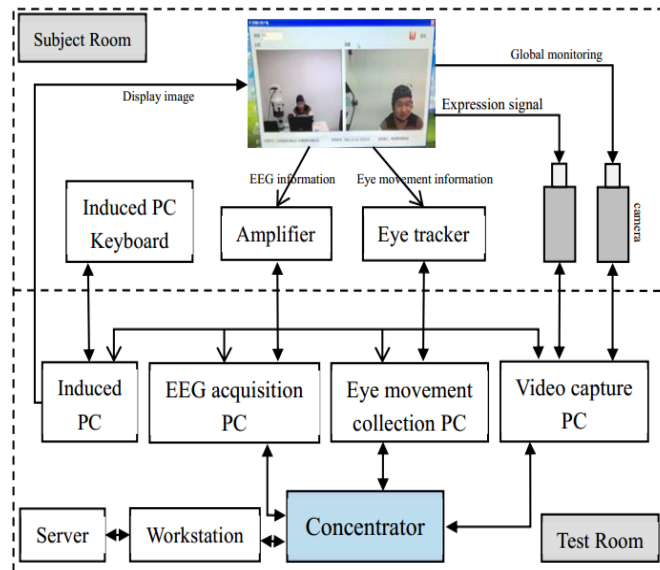
EEG acquisition system, eye movement acquisition system and video surveillance system are three separate systems with each containing an independent acquisition device and processing software. The middle of the process is transparent to the outside world. Therefore, to achieve the integration of the three, it is necessary to input and output part docking, and some also needs manual operation. For example, EEG data can be saved as an ASCII text file after the edit function is completed, and the eye movement data can also be saved as an ASCII text file by BeGaze after simple processing. Therefore, the overall platform of experimental software running on the server is different from the terminal control platform for centralized control of the custom communication interface in the image lab, and the wizard mode should be adopted. At the beginning of the experiment, various preparations are prompted. When iViewX, E-Studio, Scan and other software complete part of the intermediate links by prompts on the screen, how to use these software operations and specify the operation results to save the location, and then perform the import platform to extract the results saved to the database shall be known.

## **3. Synchronous Module and Platform Construction**

### **3.1 Synchronization Platform Architecture**

The system consists of image acquisition system, EEG acquisition system, eye movement acquisition system and induction PC, as shown in Figure 1. First of all, the stimulus information such as sound, picture or video was induced on the screen of PC, the EEG information of the subject was collected through the electrode cap and sent to EEG to collect PC and recorded by SynAmp amplifier. Eye tracker and eye will be tested by eye movement signal acquisition and PC acquisition of eye movements; at the same time, two cameras will video information collected to the video capture PC. The four PC units into the next seat, the plane (graphics workstation) to coordinate their work. EEG acquisition system, eye movement acquisition system and video capture system together constitute a distributed information acquisition

system. All collected information is stored on a multi-ethnic, large-scale cognitive database of ORACLE and DELL disk storage arrays.



**Figure 1:** Synchronous Platform System Structure

As the three sets of data acquisition platform can only be used to support the original definition of the interface software, it is difficult for the process of starting and stopping to control at the same time. In addition, the mark information of the EEG signal is a benchmark of the three synchronizations, and the mark information corresponding to the EEG mark needs to be marked in the video file.

### 3.2 Process Control

If you manually operate three sets of data acquisition software to start or stop recording, it will inevitably lead to a considerable time difference, which will not be conducive to synchronization. The solution to this problem is to set up a lab LAN that interconnects three PCs and servers through a hub. Design a client server program to establish TCP/IP connection between the server and three collecting PCs. The server sends the "start" (stop) command through the port to the three clients, and the client program calls the Windows internal function control acquisition software to start (stop) recording. In this way, it is possible to ensure the unity of the starting point of time within an acceptable error range.

### 3.3 Video Tag Information

In order to facilitate the research, we need to mark key information of the participants in the video information as a reference for synchronization. During the work of the platform, the subjects observed the stimuli induced by the induced PC and operated the keyboard. The induction PC and video captured PC by using client/server mode program through the fixed port communication. The induction PC client program used the "hook" technology to capture key information which was thensent to the video capture PC at real time (ie video capture program). The video captured program by using openCV and Daheng provided video capture card library. Key information was written to the currently collected video frame, that is, "mark."

## 4. Concrete Realization of the Synchronization Module

### 4.1 Development Tools and Environment

C++ is derived from the C language, both inherited the C language features such as efficient portability, but also introduce the object-oriented features that make the development of human-computer interaction type of application easier and faster.

Microsoft Visual C++ is the Microsoft's development of the Win32 environment program, the object-oriented visual integrated programming system. It not only has the automatic generation of the program framework, flexible and convenient class management, code integration and interface design interaction, besides, you can develop a variety of programs and other advantages as generated by a simple set of program framework to support the database interface, OLE2, WinSock network and 3D control interface, which have been widely used in the program development.

### 4.2 Program Flow Chart

The program consists of two parts, namely, synchronous acquisition and synchronization analysis. Figure 2 and Figure 3 show the process.

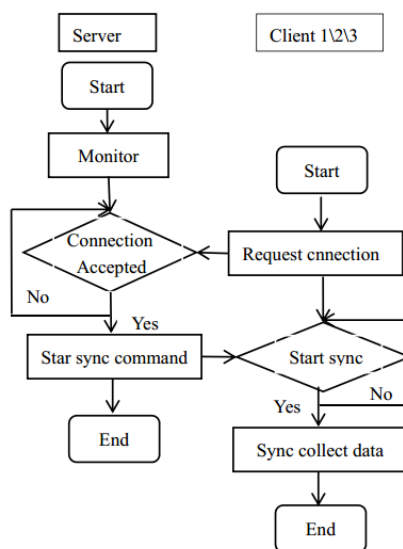


Figure 2: Synchronous Acquisition Flow Chart

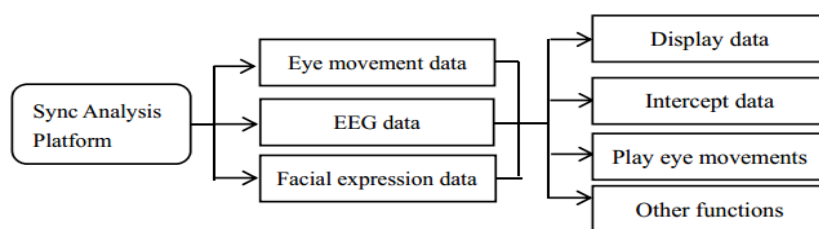


Figure 3: Synchronization Analysis Flow Chart

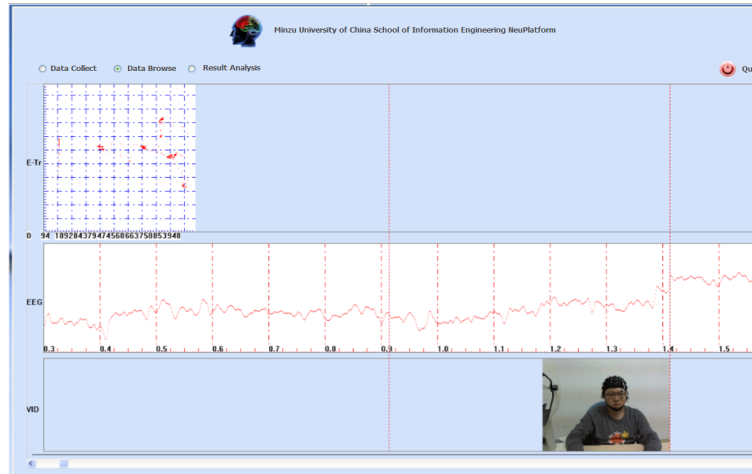
Figure 2 in the client 1\2\3 said the eye movement, EEG and expression acquisition PC. Each of these three clients sent a connection request to the server. When the server received the connection and connects successfully, it also issued a "start" command to the three clients at the same time, and the three devices started data acquisition at the same time to synchronize the data collection.

Figure 3 shows the synchronization analysis platform including display data, intercept data, play eye movements and other functions.

POS (ISCC 2017) 015

### 4.3 Main Function Introduction

Before reading these data, you need to convert the data into a text file to facilitate reading and writing in the VC. For EEG files, first read the EEG or CNT file through MATLAB, and then save the voltage amplitude data in the file as a text file. For the eye movement file, the eye movement data as collected can be saved as a text file directly through the eye movement data acquisition software. The text file contains a lot of experimental information, according to need to read the trajectory of eye coordinates, pictures and other information. The video data is read directly from the saved avi file without further processing. The program is shown in Figure 4.



**Figure 4:** Program Interface

#### 4.3.1 Data Read and Synchronized Display

The data display area is divided into three parts, followed by eye track, EEG waveform and video frame. Read the three kinds of data, visually displayed to the experimenter. When the experimental staff click the mouse, the program can be based on the location of the mouse to EEG signal time as a benchmark to show the corresponding location of eye movement data, the corresponding video frame. Facilitate researchers to observe and research.

#### 4.3.2 Data Interception

There are two vertical dashed lines on the program interface, indicating the start and end of the interception data. When the mouse clicks the interface, the dotted line is moved to the clicked position. According to the abscissa of the mouse position, the eye movement trajectory and the starting point of the video are calculated, and the eye movement area shows all the eye movement trajectories before the time corresponding to the position, and the video area displays a frame corresponding to the position and is displayed at the dotted line position. Click again to determine the end of the data. Do not change the interface display for this time. Click on the right "intercept" button according to the starting point. The starting point of eye movement and EEG data (voltage amplitude) were saved to the time-named text file.

#### 4.3.3 Zoom Function

The zoom function is designed to accurately select the location and observation of the interception of data.

#### 4.3.4 Playback Function

Determine the starting and ending points according to the two dotted lines on the interface. Click the "Play" button and the program will play three kinds of signals. The experimenter can observe the three signals in the same time period.



## 5. Conclusion

Based on multi-modal signal system that has entered the clinical application of the research stage, how to improve multi-modal signal recognition accuracy and speed become the key issues to practical system. A system based on multi-modal signal is proposed. In order to meet the requirements of system recognition accuracy and speed, it is necessary to design a reasonable experimental paradigm, select the key technologies such as effective fusion and synchronization, and collect the correct signal.

This article describes the design and implementation of a multimodal synchronization cognitive platform. This platform realizes simultaneous acquisition and analysis of EEG, eye movements and facial expression data, and can be applied to the research of multi-modal emotion recognition. Among the current researches on affective computation, the research on single signal based on expression and physiological information has been more in-depth, while the affective recognition of multi-modal integration is less. One of the reasons lies in the limitations of the research platform. The multi-modal cognitive platform can be set up to effectively EEG, eye movements and facial expression signal fusion and accurate synchronization. A large-scale multi-ethnic cognitive database is constructed by a standardized method, which is suitable for multi-modal emotion recognition research. On this basis, various extensions can also be made to meet the needs of emotion recognition based on other signals.

## References

- [1] Picard R W. *Affective computing*[M]. MIT Press, 1997.
- [2] Mason S G, Bashashati A, Fatourechi M, et al. *A comprehensive survey of brain interface technology designs*. [J]. Annals of Biomedical Engineering, 2007, 35(2):137-69.
- [3] FERRARIS E, FASSA L. *Polysilicon fatigue tested monitoring based on the 2nd harmonic of the device current measurement* [C]. In: Proceedings-2005 International Conference on MEMS, 2005:14-19.
- [4] Johannes Wagner, Jonghwa Kim, Elisabeth Andr'. *From Physiological Signals to Emotions: Implementing and Comparing Selected Methods for Feature Extraction and Classification*[J]. In IEEE International Conference on Multimedia & Expo., 2005.
- [5] Luo fei. *For facial expression recognition based on Gabor little potter character research* [D]. Chongqing university of posts and telecommunications, 2009.
- [6] Feng Xiaohang etc. *College students' basic emotion recognition eye movement research*. Journal of Qingdao university teachers college, 2012, 2
- [7] Reza Khosrowabadi, Abdul Wahab, Kai Keng Ang and Mohammad.H Baniasad. Affective computation on EEG correlates of emotion from musical and vocal stimuli: *International Joint Conference on Neural Networks*, 2009
- [8] Deng Pei, penny ur (Liang Ningjian. *The application of artificial neural network model in EEG research overview of academic conference* [C] the national psychology, 2014.
- [9] Wagner J, Kim J, Andre E. From physiological signals to emotions: *implementing and comparing selected methods for feature extraction and classification*[C] //Proceedings of the 2005 IEEE International Conference on Multimedia & Expo. Amsterdam, the Nether lands, 2005: 940- 943
- [10] Huang Chengwei Jin Yun, qing-yun wang, Zhao Li, zou rong. *Based on speech signal and multimodal emotion recognition of ecg signal*. Journal of southeast university (natural science edition), 2010, 5