2013 10th IEEE International Conference on Control and Automation (ICCA)
Hangzhou, China, June 12-14, 2013

System Design for Monitoring Infant Speech Emotion†

Lin Zhong, Yanxiong Li*, Xin Wei, Guanglong Li, Zili Wang, Yewen Jiang.

Abstract—**To alleviate worries of busy younger parents, this paper designs a system for monitoring infant speech emotion. The system distinguish infants' cry from other voice (e.g. parents' voice) by using speech emotion recognition. When the inputted speech signal is recognized as infant cry, it judges cry reasons (fever, urination, hunger) according to the signals from sensors and then take proper propitiatory measures (playing music or parents' voices) to ease the infants. Finally, it sends the guardians a message or calls the guardians about cry reasons by the GSM module, so that humane guardianship for infant is realized. Experimental results show that the system realizes the aforementioned functions, and is a good partner for infant and an under servant for younger parents.**

I. INTRODUCTION

Nowadays, most of young parents are fully occupied by their work and often feel exhausted. As a result, they have no enough energy to look after their infants. Their tiredness and carelessness will make infant get ill or lead to some serious accidents. Hence, it is beneficial for young parents to develop a system for infant monitor.

There have been some infant monitoring products in market. But most of them only realize simple functions with high prices [1]. For example, some products simply judge infant state according to the external information, e.g. body temperature or diaper humidity. Those products have a disadvantage of low accuracy [2]. Some products use video processing technologies for infant monitor [3]. However, hardware of high performance is required for video signal processing. As a result, the expenses of those products are too high for common consumers.

To alleviate worries of busy younger parents, this paper designs a system for monitoring infant speech emotion. This system distinguishes infants' cry from other voice (e.g. parents' voice) by using speech emotion recognition. Next, it judges cry reasons (fever, urination, hunger) using the signals of sensors and then takes proper propitiatory measures (playing music or parents' voices) to ease the infants. Finally, it informs the guardians of infant's states by sending a message or giving a ring.

* Corresponding author: Yanxiong Li (eeyxli@scut.edu.cn)

† This work was supported by the following projects: National Natural Science Foundation of China (No. 61101160), the Natural Science Foundation of Guangdong Province, China (No. 10451064101004651), the Fundamental Research Funds for the Central Universities, South China University of Technology, China (No. 2011ZM0029), the Foundation for Zhujiang Technological New-Star of Guangzhou City, the Foundation of China Scholarship Council, the Innovative Experimental Project for Chinese's University Students (No. 201210561046) and the Innovative Experimental Project for Guangdong Province's University Students (No. S1010561034).

All authors are with School of Electronic and Information Engineering, South China University of Technology, Guangzhou, China.

The rest of the paper is organized as follows. Section 2 presents system design. Experiments and discussions are given in Section 3, and finally conclusions are drawn in Section 4.

II. SYSTEM DESIGN

The system consists of three components. The first component is called control module, which collects voice signals, identifies the emotional states of infants and works collaboratively with other two modules as well. The second component is sensor module, which collects the causes of the infant's cry. The third component is processing module, which responds to the infant's cry. Figure 1 explicitly describes the architecture of the system.

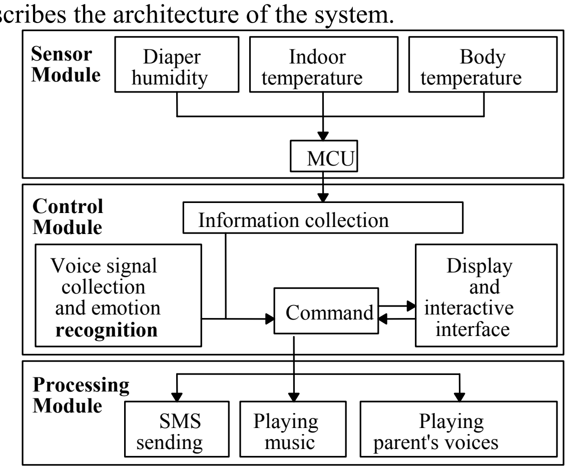


Figure 1. System architecture

A. Hardware design

1) Control module

The control module functions as the identifier of voice and emotion, and is the controller of other two modules. Due to the intensive computation for identifying speech emotion, we employ a highly cost-effective chip, the ARM9-architectured S3C2440 with a NAND Flash memory chip named K9G8G08UOA, as main processor for obtaining better performance.

The displaying and operational functions are realized by the TFT (Thin Film Transistor) screen, the screen driver chip ILI9325 and the touchable chip ADS7843. They combine and present an interactive interface to display the states of infants (body temperature and humidity of diaper) and the external state of the environment, meanwhile responding to the commands from users.

2) Sensor module

The Sensor module is designed for collecting infants' states and indoor temperature. It consists of three sub-modules for monitoring: body temperature, diaper humidity and environmental temperature.

The first two sub-modules above employ a cost-effective SHT11 chip integrated with temperature and humidity monitoring functions. SHT11 is highly integrated, accurate and reliable, which minimizes the data collection sensor and won't discomfort the infant if placed properly in the diaper.

The sensor in chip SHT11 detects diaper humidity and the infant's body temperature. When diaper humidity or body temperature exceeds a threshold, the wireless transmission module sends the control module the corresponding information which subsequently triggers the control module to respond.

Compared with the two sub-modules above, the sub-module for monitoring environmental temperature is less strict with accuracy, thus a relatively cheaper temperature sensor DHT11 is employed. The sensor module communicates with the control modules via wireless module NRF24L01.

3) Processing module

The processing module is designed to inform parents and play music and parents' voices. Both music and voices can be played by the audio playback sub-module. The GSM (Global System for Mobile communications) transmission sub-module adopts the widely used Siemens TC35 module and sends SMS to parents when their infants are cry, informing them with the causes of cry. In this way, the infant will be fine. Audio playback sub-module is composed of the regular amplifier circuit LM386 and speakers [4]. The music and parents' voices come from the memory chip on the control module (It plays them in advance). The processing module communicates with the control module via serial port.

B. Software design

Flow chart of the software is given in Figure 2. After initialization, the system tries to identify the infant's emotion status by analyzing the collected voice signal. If the infant is cry, the system will analyze the causes of cry with signals from the sensor and then comfort the infant, meanwhile informing parents using SMS. If the infant is not cry, the system will not take measures and continue to collect and analyze voice signals.

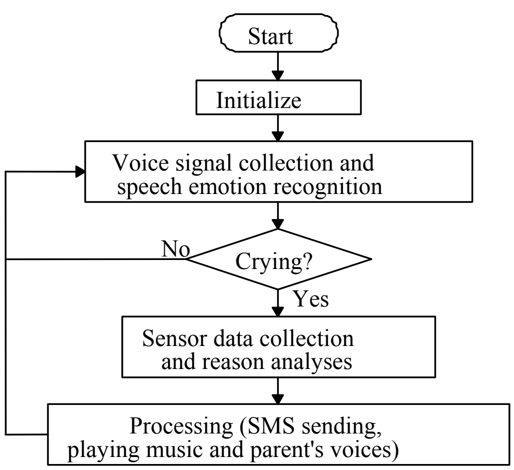


Figure 2. Flow chart of the software

1) Speech emotion recognition method

The previous studies on the processing of infant cry focuses on classification of infant cry rather than discrimination between adult voice and infant cry, and are model-based methods for cry classification. That is, MFCCs (Mel-Frequency Cepstral Coefficients) are used as features and HMM (Hidden Markov Model) is used as classifier [5].

It is known that the model-based methods not only need a lot of training data for model training [6], but also is time-consuming for running on the micro-embedded system. It is reported that the HMM-based methods are not fit for implement on the micro-embedded system due to its computational complexity.

In order to solve the problems of the aforementioned HMM-based method, we proposed a low-complexity method for distinguishing adult voice from infant cry based on spectrum analysis [7]. FFT (Fast Fourier Transformation) is used to transform the audio data from time domain to frequency domain. Then the statistical difference of spectrum between infant cry and adult voice is analyzed. Finally, the dominant frequencies of speech signal are used as an effective feature to distinguish infant cry from adult voice. The proposed method flowchart is shown in Figure 3. The readers are referred to [7] for details of this method.

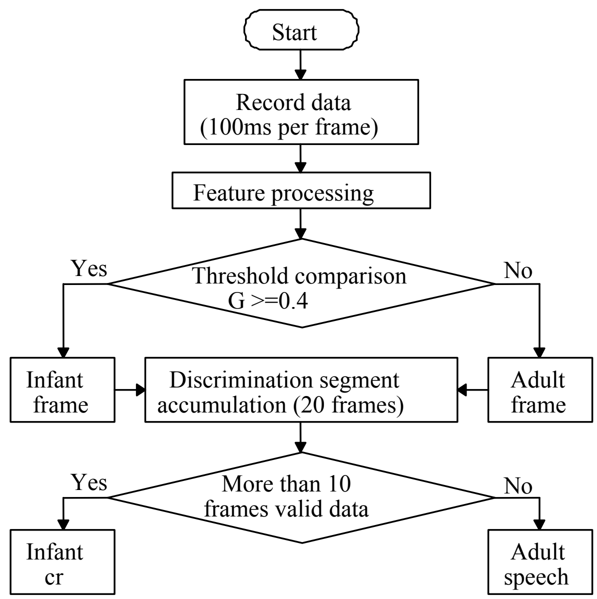


Figure 3. The proposed method flowchart

2) Control program

The control program is divided into two parts. The first part is the control program for the control module, while the second part is the MCU (Micro Control Unit) control program for the sensor module. The functions of the first part are as follows. First, it collects voice signal from peripheral circuit and recognize speech. Second, it analyzes the data from the sensor and figure out the causes of infant's cry. Third, it makes appropriate responses when the infant is cry (e.g. playing music and parents' voice, sending SMS, etc.). Finally, it drives TFT screen displays, control the touchable screen chip, respond to user's operation and provide a friendly interactive interface.

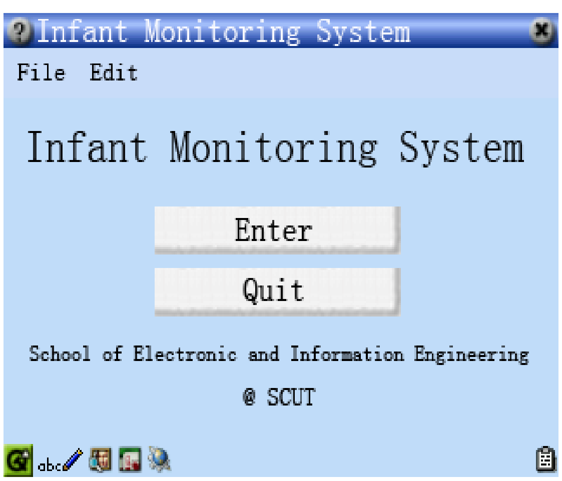
On the ARM9 platform, control module was controlled by the Linux operating system, whose touch screen drivers and serial communication drivers simplify most of the driver program [8]. The control program is designed as a four-thread program. Each of the four threads controls one of the above four functions. That is, the first thread controls the microphone, collects voice signal, frames the voice signal, identifies each frame, and sends control signal when the infant is cry. The second thread receives and constantly updates sensor data through wireless module NRF24L01. Then the thread immediately transmits the data to the main thread for analyzing cry causes. The third thread waits to receive control signals from the main thread, and controls the audio output and informs the audio playback sub-module to play. Meanwhile, it triggers the GSM sub-module to send SMS by controlling the serial interface. The fourth thread receives messages from the main thread, displays them on the screen with the Qt (graphical user interface software development kit) interface program and feedbacks touch operations to the main thread by calling touch screen driver program.

The second part, i.e. the MCU control program for the sensor module, transmits the data collected by the sensor to the wireless transmission module by controlling related sensors and wireless module NRF24L01.

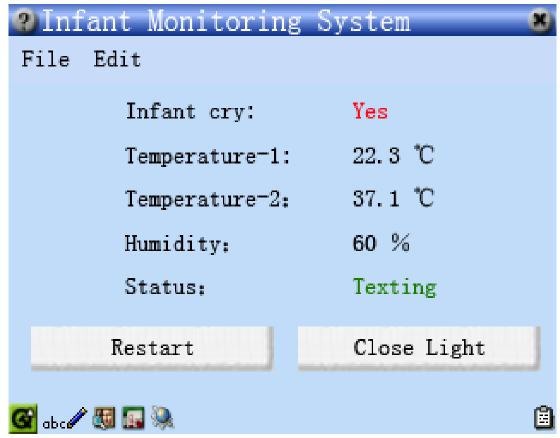
III. EXPERIMENTS

In our experiments, we play infant's cry segments 250 times at the rate of 1 time/min and 1s/time. The cry segments were collected from Children's Hospital of Bai-Yun District in Guangzhou. We also ensure that the microphone is within 0.5 meter from the sound source and the voice segment lasts at least 10 seconds. The sensor module is placed in the infant's diaper.

241 times out of 250 are able to identify the infant's cry, obtaining 96.4% of identification rate. After recognition, system can play the music and parents' voice and send them a message of the ambient temperature, the infant's body temperature and diaper humidity. The information above also displays on the screen at the same time. Figure 4 shows the interactive interface.



(a)



(b)

Figure 4. (a) Qt effect diagram (initializing); (b) Qt effect diagram (monitoring)

Figure 4 (b) shows that the infant is crying, the ambient temperature is 22.3 ℃, the infant's body temperature is 37.1 ℃, diaper humidity is 60%, and the monitoring system is sending messages to the parents. After sending message, the system waits for restarting monitoring by pressing the "Restart" button. If the "Close Light" button is pressed, the system will turn off the screen to save power.

We use a group of 100 infant cry segments and adult's voice to train the HMM module, while another 250 cry segments and adults are used as testing data. Table 1 and Table 2, respectively, present the results (confusion matrix) of distinguishing between infant's cry and adult voices by using our proposed method [7] and the HMM-based method. Table 3 shows the performance comparison of these two methods.

TABLE I. CONFUSION MATRIX OF THE PROPOSED METHOD [7]

Item	***Infant cry***	***Adult voice***
Infant cry	96.4%	3.6%
Adult voice	1.2%	98.8%

TABLE II. CONFUSION MATRIX OF THE HMM-BASED METHOD

Item	***Infant cry***	***Adult voice***
Infant cry	98.1%	1.9%
Adult voice	0.7%	99.3%

TABLE III. PERFORMANCE COMPARISON OF THESE TWO METHODS

Item	***The proposed method [7]***	***The HMM based method***
Discrimination accuracy	96.4%	98.1%
Operation time	Completely real-time	4.5s delays

The results show that the proposed method greatly reduced the time complexity with a high recognition rate. It should be noted that recognition rates will gradually decline as the microphone moves away from the sound source. The system can hardly identify cries shorter than one second, for these cry segments are regarded as noise and thus are ignored. The body temperature collected by the sensor module will differ greatly if the sensor module is placed in different locations of the diaper.

IV. CONCLUSIONS

This paper primarily focused on the design of an infant monitoring system based on speech emotion recognition. From the results of the experiments, we have realized all the expected functions of the system. The overall performance of the system is stable. However, we are also facing challenges that the microphone can't exceed a certain distance from the voice source, and that collecting infant's body temperature is somewhat difficult. We will erase these drawbacks in future work.

REFERENCES

[1] Qianinfo, “Chinese infant monitoring products marketing research,” (in Chinese) *Qianinfo Consulting*, vol. 1, pp. 1-13, 2010.

[2] N. Bharathi, P. Neelamegam, “FPGA-based temperature monitoring and control system for infant incubator,” *Sensors and Transducers*, vol. 143, no. 8, pp. 88-97, Aug. 2012.

[3] W Wang and X. B. Ning, “Audio and video detection processing in infant monitoring system,” (in Chinese) *Journal of Nanjing University*, vol. 39, no. 3, pp. 440-445, Mar. 2003.

[4] G. Gao, Z. Liu, H. Y. Jiao and J. T. Wang, “The design of recording system based on LM386,” *Communications in Computer and Information Science*, vol. 164, pp. 590-597, 2011.

[5] Y. Abdulaziz and S.M.S. Ahmad, “An accurate infant cry classification system based on continuous hidden Markov model,” in 2010 Proc. *Int. Symposium in Information Technology*, vol.3, pp. 1648–1652.

[6] Y. X. Li, Q. H. He, S. Kwong, et al, “Characteristics-based effective applause detection for meeting speech,” *Signal Processing*, vol. 89, no.8, pp. 1625-1633, Aug. 2009.

[7] X Wei, Y. X. Li, L. Zhong, et al, “Distinguishing infant cry from adult voice based on spectrum analysis,” in 2012 Proc. *IEEE Int. Conf. on Audio, Language, and Image Processing*, pp. 333-336.

[8] Y. Wang and J. Y. Zhang, “Research on embedded numerical control system hardware platform based on ARM9,” *Procedia Engineering*, vol. 15, pp. 484-488, 2011.