Automated Portable Cradle System with Infant Crying Sound Detector

[2]Suhaib Azhar, [1,2]Khairunizam W.A.N., [2]Azri A. Aziz, [1]Zuradzman M. Razlan, [1]D. Hazry and [2]M. Farhan Kamil

[1]*Centre of Excellence for Unmanned Aerial Systems (COEUAS)*
[2]*Advanced Intelligent Computing and Sustainability Research Group, School of Mechatronic Universiti Malaysia Perlis, Kampus Pauh Putra, 02600 Arau, Perlis, Malaysia*

A R T I C L E   I N F O

A B S T R A C T

This paper describes the analysis of sound signals specifically infant crying sound through audio signal digital processing for development of automated portable cradle system sound detector. The input sound signals are filtered to a certain frequency range in order to swing the cradle and then the signals are analyzed. In order to analyze the sound signals, the system undergoes a certain process which is audio signal processing. Certain signal processing techniques have been used to observe the waveform of the sound signals. The purpose is to compare between six different types of sounds visually. The results obtained shows that the sound signals are visually distinguishable with one another after applying the processing techniques.

INTRODUCTION

Four decades ago, researchers have been working on acoustic features of speech for the use of differing between individuals' voices. These acoustic patterns reflect both anatomy and behavioral patterns (Farjo, J., *et al*., 2012). From then on, many researchers and inventors discover the new applications from digital signal processing of human voice. Recently, infant crying sound analysis study had been a popular research in order to find the suitable frequency for sound detection (Raina, P.D., and Anagha, M.P., 2012). Previous research works on identifying the cause of infant crying by using frequency analysis(Furui, 1986). Sound signal analysis consists of process to convert a speech waveform into features that are useful for further processing. There are many algorithms and techniques are used.

The objective of sound signal analysis is to differentiate the signals between different sound signals. The voice signals can be differentiated by comparing through certain acoustical features of sound such as the mean and variance of a segmented audio signal (Subramanian, 2004). Several techniques have been used for speech processing and audio feature extraction. The methods mostly engage in either spectral or frequency domain (Molau, S., *et al*., 2001). Firstly, human voice is converted into digital signal form to represent the signals at every discrete time step. The speech that is digitized will be processed to extract the audio features(Manish, P.K., 2003; Alin G. Chitu, *et al*., 2007). The popularly used signal processing techniques for signal pattern comparison are pre-emphasis, frame blocking and windowing(McLoughlin, 2009). However, this system will add in one more technique in prior to pre-emphasis which is normalization. These techniques were implemented using Lab VIEW (Laboratory Virtual Instrument Engineering Workbench) and MATLAB (Matrix Laboratory). The differences between the six sound signals are analyzed.

Automated portable cradle system is an infant crying sound detection system that will respond by swinging the cradle only if the infant crying sound is detected by the system. The other sounds however will not be detected. Basically the purpose of this project is to reduce the time for parent in monitoring their baby. A small microphone module is used as sound sensor to detect the infant crying sound. The infants' sound will become the input signal in order to swing the cradle. Voice consists of sound made by a human being using the vocal folds. The frequencies of infants' crying sounds are between 370 Hz to 420 Hz(L. L. LaGasse, R. Neal, and B. M. Lester, 2005). Therefore, the sound detector of the cradle system should be able to distinguish between different sounds in order to respond to infants' sound frequency only.

The proposed method in this paper is to detect infant crying sound using audio signal processing techniques. The sounds are detected by a microphone attached to the cradle system which will filter the sound signals using a band pass-filter. The filter will block frequencies below a low limit and above a high limit. It will allow certain

Corresponding Author: Suhaib Azhar, Advanced Intelligent Computing and Sustainability Research Group, School of Mechatronic Engineering, Universiti Malaysia Perlis,KampusPauh Putra,02600 Arau, Perlis, Malaysia. E-mail: suhaib99azhar@gmail.com

frequencies passing between the limits. The signals are then used for signal processing techniques. Those techniques are normalization, pre emphasis, frame blocking, and windowing respectively for analyzing the sound signal frequencies.

This research paper is structured as follows: Section 2reviews the research materials and methodologies. Section 3 describes the experimental results. Finally, Section 4 describes discussions and conclusions.

MATERIALS AND METHODS

The experiment was done by acquisition of six different sounds which were infant crying, adult talking, door closed, operated fan, raining, and operating vehicle that were recorded at certain time intervals between 5-8 seconds. Fig.1. shows the mechanical structure of baby cradle which was built by using aluminum frame. The sound detection system attached to the cradle consists of three circuits which were microphone, band pass filter and PIC18F4580 microcontroller.

The microphone will transfer the sound signals received to the band pass filter circuit which filters frequencies to the range of 370-420 Hz. The filtered signals will be processed by the PIC to trigger a relay in order to swing the cradle. The sound signals were also analyzed afterwards after applying the signal processing techniques. The purpose is just to observe the waveform difference between the sound signals visually.
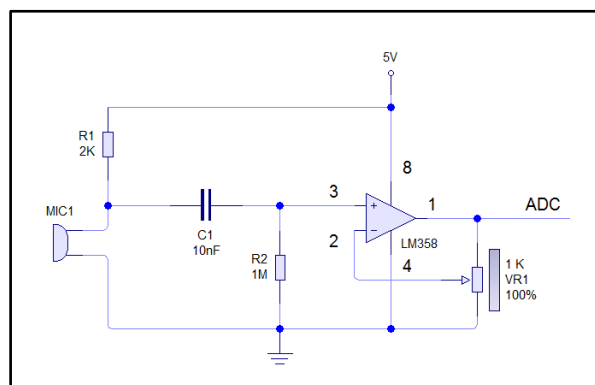

Fig. 1: Baby Cradle Mechanical structure.

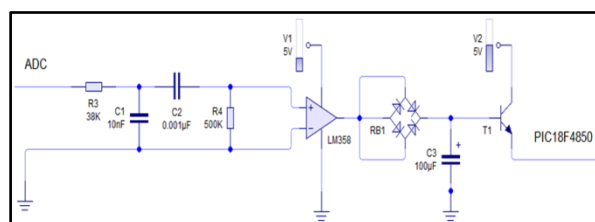**Fig. 2:** Microphone circuit diagram.
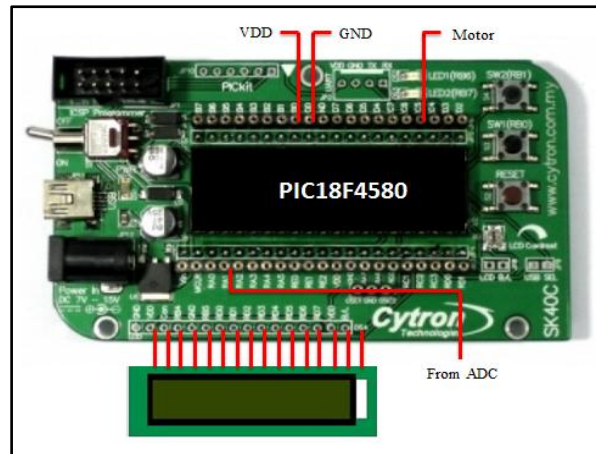

Fig. 3: Band Pass Filter Circuit.

Fig. 4: PIC circuit.

2. Infant Crying Signal Processing:

Signal processing is an important stage for the development of an agilesound recognition system. Processing is applied to enhance the attribute of the sound signal and to improve accuracy of the system. There are four processing techniques that will be used to enhance feature extraction. These include signal normalization, pre-emphasis, frame-blocking and windowing.

For speech processing purpose, each of the sound signals is sampled to 16 kHz. This is because most of the significant voice features of the infants' cry are within 5 kHz bandwidth(M. Hariharan, J. Saraswathy, R. Sindhu, Khairunizam Wan and Sazali Yaakob, 2012). The sampling frequency must be at least twice or larger than the input sound frequency for accurate data sampling.
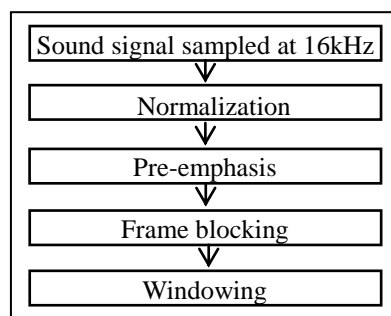

Fig. 5: Speech signal processing block diagram.

3.Recording and Digitizing:

The analogue sound signal is recorded using a microphone. Subsequently, the analogue signal is sampled and quantized. Speech signal are usually represented as functions of continuous variable *t*, which denotes time. The analogue speech signal $S_a(t)$ can be defined as a function varying continuously in time. The processed signals are sampled with a sampling period T_s. Then, we can define a sample of a discrete time signal as

$$S(n) = Sa(nT_s) \tag{1}$$

Which means$t = nT_s$. The signal *S(n)* is called digital signal. According to the sampling period can be defined the sampling frequency as$F_s = T_s^{-1}$. Usually the sampling frequency of the speech signal lies in the range 8000<F_s<22050(L. L. LaGasse, R. Neal, and B. M. Lester, 2005). The sampling frequency of 16 kHz is chosen for a specific reason. The recorded digital signal is of a finite length, which is referred to as N_{total}.

4.Signal Normalization:

Signal normalization is the process of increasing or decreasing the amplitude of a sound signal evenly. The purpose is to reduce disparity between signals that have been recorded in various environments and to avoid the error estimation caused by speakers' volume changes(C. Y. Fook, et al., 2012). The formula equation for normalization is as follows (2).

$$S_n = (S_i - \bar{x}_i)/\partial_i \tag{2}$$

Where S_i and S_n is the ith component of the signal before and after signal normalization respectively. $\bar{x}_i$ and ∂_I is the mean and standard deviation of vector s respectively. Sound signals are converted into signal data of normal distribution with mean equal to zero and variance equal to one (C. Y. Fook, et al., 2012).

5.Signal Pre-emphasis:

The pre-emphasis filter is used to improve the high frequency portion of the signal that was suppressed during the sound recording session and to magnify the high frequency formants. The filter function is as below:

$$y(n) = b(1)x(n) + b(2)x(n-1) + ... + b(nb+1)x(n-nb) - a(2)y(n-1) - ... - a(na+1)y(n-na) \qquad (3)$$

Where$n-1$ is the filter order, which handles both FIR and IIR filters(Oppenheim, A. V., Schafer, R. W., & Buck, J. R., 1999), na is the feedback filter order, and nb is the feed forward filter order.

6. Frame Blocking:

Frame blocking is the process of cutting the sound samples obtained into small frames with length within the range of 10 to 50 ms. The sound signal is divided into frames of N samples. Adjacent frames are being separated by M (M<N).

7. Windowing:

Windowing is to minimize the discontinuities of signals at the start and end of each frame. The window tapers the signal to zero at the start and end of each frame in order to minimize the spectral deformity. However, the window function does not have to be identically zero at the end of interval, as long as product of the window goes sufficiently rapidly towards zero. The window function is described as follows,

$$w(n), \qquad 0 \le n \le N\text{-}1 \qquad (4)$$

Where N is the number of samples of a frame while the signal is the result of windowing,

$$y_I(n) = x_I(n)w(n) \quad 0 \le n \le N\text{-}1 \qquad (5)$$

Hamming window is used in this experiment due to better selectivity for large signals, which has the form:

$$w(n) = 0.54 - 0.46 \cos(2\pi n/(N\text{-}1)), \quad 0 \le n \le N\text{-}1 \qquad (6)$$

RESULTS AND DISCUSSIONS

The six input sound signals are shown in Fig.6 as waveform charts of amplitude(V) versus time (s).
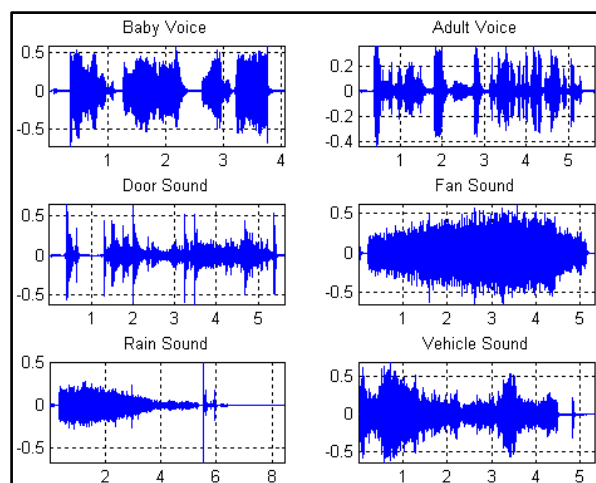

Fig. 6: Input signals of 6 different sounds.

Fig.6 is used for carrying voice signal analysis performance evaluation using signal processing techniques. The input sound signals are the raw input data of sound signal that were not yet digitally processed. It can be seen in Fig.6that each sound signal had its own waveform shape. When comparing between the different sound signals,

it can be seen that the waveform shapes were moderately distinctive. The waveform shape of infant crying voice was very different than the fan sound signal for instance.

The comparison between normalized signals of infant cry and adult voice is shown in Fig.7. The difference can be seen from the shape of the waveform and the value of amplitude. The waveform shapes of both signals differ a lot.
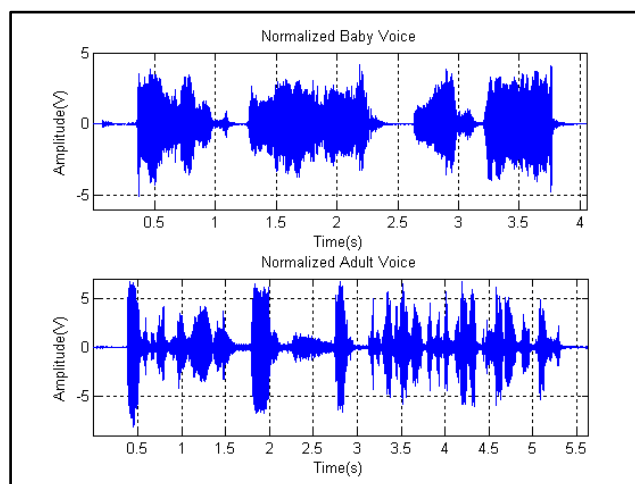

Fig.7: Normalized Signals.

It can be seen that the amplitudes are uniformly increased after normalization. It is obtained that the maximum amplitude value of infant cry signal increased from 0.6V to 4.2Vand the adult voice signal increased from 0.4V to6.7V. This shows that by normalizing, the signals' amplitude increment value is not the same.
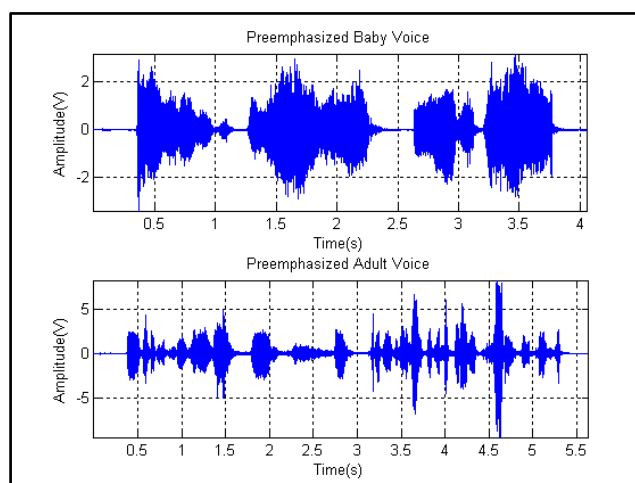

Fig. 8: Pre emphasized signals.

By referring to Fig.8,themaximum amplitude value of infant cry signal decreases from 4.2V to 3.2V which shows that the pre-emphasis process spectrally flattens the infant cry signal and even the spectral energy envelope by amplifying the importance of high frequency components. However, the maximum amplitude value of adult voice increases from 6.7V to 8.2V.

Fig.9 shows the result of frame blocking process for both infant cry and adult voice signals. The sound sample of the signals are segmented or cut into N samples of 800 as shown in Fig.9. It can be seen that the segmented signals from start and end of both frames are in continuous form. Therefore, a hamming window function is applied to the segmented sound signal sample.

From Fig.10, the signal was windowed by using hamming window which causes the signal to have a close similarity to hamming window by multiplying to the applied windows' function on the signal. This was to decrease the spectral deformity by using the window to reduce the signal to zero at the start and end of every frame. Both signals differ a lot when comparing according to waveform shape perspective. The maximum amplitude of infant cry and adult voice signals is also different which are 0.8V and 2.4V respectively.
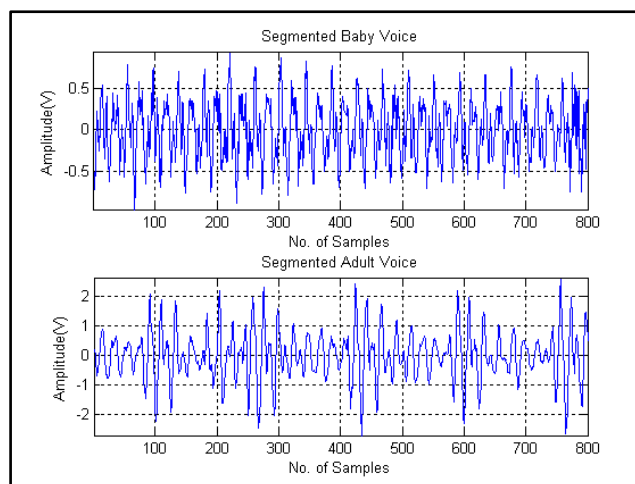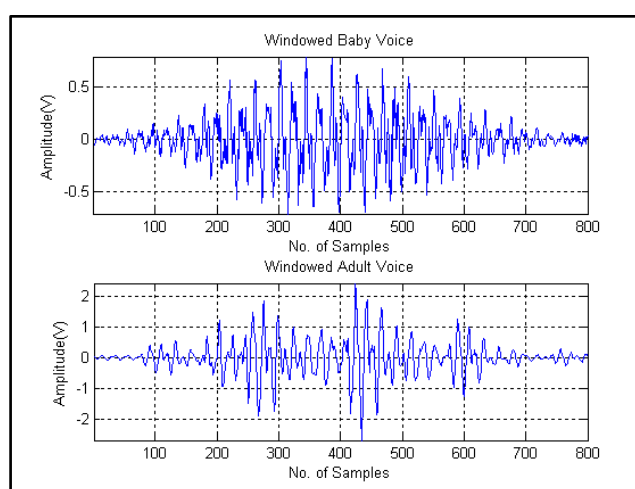
Fig. 9: Frame blocking of signals.

Fig. 10: Windowed signals.

Conclusion:

From the results obtained, the signal processing techniques increase the higher frequency amplitudes while lowers the lower frequency amplitudes uniformly and the windowed signals obtained after applying the Hamming window function were compared. It is known that each windowed signal for different sounds differ moderately.

The drawback of the system is that the system detects by filtering the sound within the range of 370Hz to 420Hz using a band pass filter. This means it will receive any kind of sound that produces a sound frequency of that range. It seems that filtering the sound is not enough to indicate that the sound is an infant crying sound. Besides that, the analysis of sound through signal processing techniques only performed for the purpose of observation and comparison between different types of signals visually. It does not improve much in the accuracy of infant cry detection but rather just an insight of visual comparison.

As a conclusion, the signal processing techniques used such as normalization, pre-emphasis, frame-blocking and windowing successfully improves the quality of signals to differentiate the signals of the six different sounds.

REFERENCES

Alin G. Chitu, *et al*., 2007. Comparison between Different Feature Extraction Techniques for Audio-Visual Speech Recognition. *Multimodal User Interfaces, 1*(1): 7-20.

Fook, C.Y. *et al*., 2012. Comparison of Speech Parameterization Techniques for Classification of Speech Dysfluencies. *Turkish Journal of Electrical Engineering and Computer Sciences, 1*(1): 1983-1994.

Furui, S., 1986. Speaker-independent isolated word recognition using dynamic features of speech spectrum. *Acoustics, Speech and Signal Processing, 34*(1): 52-59.

Hariharan, M., J. Saraswathy, R. Sindhu, Khairunizam Wan and Sazali Yaakob, 2012. Infant cry classification to identify asphyxia using time-frequency analysis and radial basis neural networks. *Expert Systems with Applications,* 39(10): 9515-9523.

Huang, X., A. Acero and H. Hon, 2001. *Spoken language processing: A guide to theory, algorithm, and system development.* Upper Saddle River, NJ, USA: Prentice Hall PTR.

LaGasse, L.L., R. Neal and B.M. Lester, 2005. Assessment of Infant Cry: Acoustic Cry Analysis and Parental Perception. *Mental Retardation and Developmental Disabilities Research Reviews, 11*(1): 83-93.

Manish, P.K., 2003. *Feature Extraction For Speech Recognition.* M.Tech., EE. Bombay: IIT.

McLoughlin, I., 2009. *Applied Speech and Audio Processing: With Matlab Examples.* NY, USA: Cambridge University Press New York.

Molau, S. *et al.*, 2001. Computing Mel-frequency cepstral coefficients on the power spectrum. *Acoustics, Speech, and Signal Processing.1*, pp. 73 - 76. Salt Lake City, UT: IEEE.

Oppenheim, A.V., R.W. Schafer, J.R. Buck, 1999. *Discrete-Time Signal Processing* (2nd ed.). Upper Saddle River, New Jersey, USA: Prentice Hall International, Inc.

Subramanian, H., 2004. *Audio Signal Classification.* M.Tech., EE. Bombay: IIT.