Baby Monitoring Through MATLAB Graphical User Interface

C Shruthi Reddy, Sowmya Ravi, Giriraja C V

Abstract: This paper describes a novel approach to monitor a baby and it's emotion and needs. Feature extraction methods like Magnitude Sum function, Pitch and Energy have been performed to classify the signal. These extraction techniques are proven to be more accurate than the conventional techniques. Although combinations of all three techniques have to be used to achieve 100% accuracy, the computational cost and processing time is less than that of Mel Frequency Cepstral Coefficient. Thus, classification of hungry, tired and in-pain cries were successfully done.

Index Terms: Magnitude Sum Function, Pitch, Energy, Feature Extraction, Baby cries: hungry, tired, in-pain

1 Introduction

IN today's world Signal Processing plays an important role as it is changing the face of the current technology. One such big change that could revolutionize the face of the current scenario is our project. We are dealing with developing a prototype to monitor a baby while its caretaker is away. Different baby cries are recorded and fed into this system. After careful analysis, a baby's cry can be classified into three major categories namely: Tired, Hungry and Pain. A fourth type of trigger is also generated if the baby's cry cannot be recognized. A frame by frame analysis is carried out for a better analysis of the baby's cry. A speech signal rapidly changes. So analysis the signal as a whole would not yield us the right results. Thus, in Speech Processing the signals are divided into frames where analyses of few samples are carried out. In our project, we chose a frame of 250 samples. Different feature extraction techniques can be used to recognize a baby's cry. In our project we extracted four features namely: Magnitude Sum function (MSF), Zero Crossing (ZC), Pitch Extraction and Energy of the signal. A combination of MSF, pitch and energy was implemented in our project. It provided us with 100% accuracy. It requires less computational operators and faster computation is possible. Even though Mel Frequency Cepstral Coefficients (MFCC) closely resembles the human auditory system, there many complications involved. Mainly, the system fails to provide a good accuracy in the presence of noise. Secondly, the computational cost associated with it is large when compared to our method.

2 Feature Extraction

In this section a brief overview of the feature extraction methods are explained. Also, the drawbacks associated with it and the thresholds that needs to set is also explained

2.1 Magnitude Sum Function (MSF):

MSF can be defined as the sum of the absolute value of the input signal. It is given by the following formula:

$$S[m] = \sum_{n=1}^{m} ip(n) \qquad \ldots(1)$$

• *C Shruthi Reddy, Sowmya Ravi, Giriraja C V, Amrita School of Engineering*

If we use this feature individually, then the MSF of a tired signal and that of a noisy signal will almost be the same. This possesses a problem while coding. To overcome this, we use a combination of MSF, pitch and ZC. Refer Table 2 for studying the Magnitude Sum Function for different types of inputs.

2.2 Pitch:

Pitch is the subjective pycho-acoustical feature of sound. It tells about the fundamental frequency of the speech signal. The pitch frequency of a matured human voice will lie in the range of 100-350 Hz and the pitch of a baby's cry lies between 350 and 500 Hz. As the baby's cry cannot be accurately quantified based on pitch, it is used to distinguish between an adult's speech, noise and baby's cry. Pitch can be calculated using many techniques, we employ two of them; one using Short Time AutoCorrelation (STAC) and the other using Cepstrum Domain Analysis. Cepstrum domain method gives better accuracy but sine our aim is to differentiate noise, speech and cry, we employ STAC which gives a tolerance of 5%. A comparison between STAC and Cepstral Domain Analysis is shown in Table 3.

The autocorrelation function is defined as follows

$$AC(k) = \sum x(m)x(m+k) \qquad \ldots(2)$$

Where, the summation varies from (–)inf to (+)inf.

Considering that the input is periodic,

$$AC(k) = AC(k+m) \qquad \ldots(3)$$

The AC function is an even function meaning

$$AC(k) = AC(-k) \qquad \ldots(4)$$

The STAC function can be defined as

$$STAC(k) = \sum x(m)w(n-m)x(m+k)w(n-k-m) \qquad \ldots(5)$$

Where, n is the length of the window, m is the current sample and w (n) is the window function.

Considering eqn. (4)

$$STAC(k) = STAC(-k) = \sum x(m)w(n-m)x(m-k)w(n+k-m) \qquad \ldots(6)$$

If we define

$$H(n) = w(n)w(n+k)$$

Then,

$$STAC(k) = \sum x(m)\, x(m-k) H(n-m) \qquad \ldots(7)$$

K^{th} AC at time n is obtained by filtering x(n)x(n-k)

With a filter with an impulse response of H(n)

2.3 Energy

The energy of a signal is not an accurate parameter to study the different kinds of signal. But since the energy of a noisy signal is said to be lesser than that of a legible speech signal, we can differentiate them easily. Based on our studies, we noted that the noisy signal has a maximum energy of 170J whereas, the weakest tired signal has an energy of approx. 200J. Thus, energy is used to remove the possibility of wrong detection.

3 PROCEDURE

The following flowchart pictorially represents the algorithm

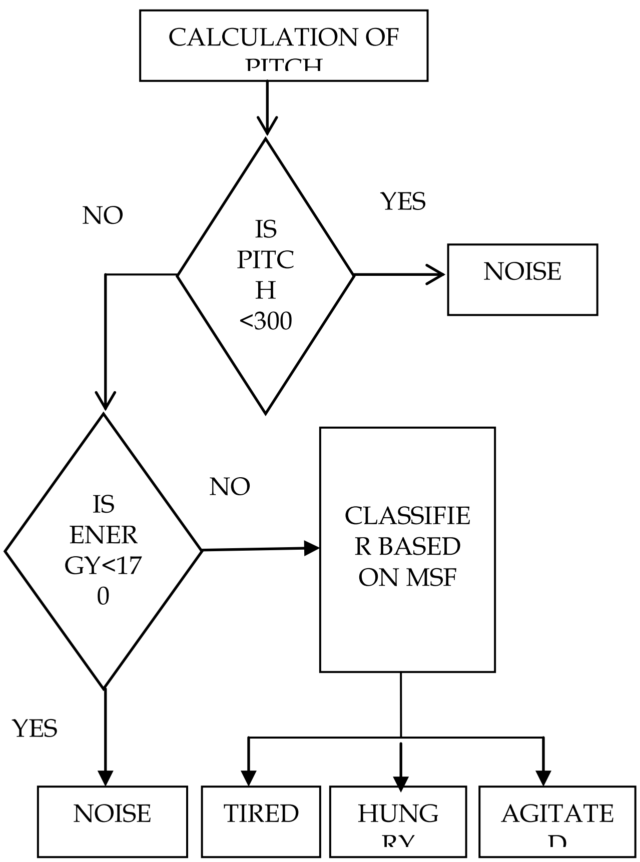

Fig. 1: FLOWCHART

4.1 Steps Involved

- **a)** Read the audio file
- **b)** Frame by frame segmentation for better analysis of input signal
- **c)** Removal background noises for good perceptual quality
- **d)** Removal pauses in the signal to have a more accurate computation of minimum MSF
- **e)** Calculation of pitch
- **f)** Set threshold as 300Hz for pitch. For any signal with values less than this threshold frequency, we conclude that the signal is either noisy or is the speech of an adult.
- **g)** If for a noisy signal, the pitch is greater than 300Hz, it can detected by analyzing the energy of the signal.
- **h)** Suppose the energy of the signal is less than 170J, it is classified as a noisy signal.
  i.e for signals with pitch>300 and energy<170, it is classified as noisy signal
- **i)** After feature extraction, the signal has to be classified. So thresholds are set after careful analysis of different signals.
- **j)** The threshold for classification is mentioned in Table 1.
- **k)** Based on the above parameters, the signal is classified as Hungry, Tired or Pain.

4 HARDWARE IMPLEMENTATION

Graphical User Interface (GUI) is easy for end-user to run any kind user without the need to learn programming. Our application is one such example where the user should install the application and run it. It can be easily created on MATLAB by following a series of steps which is mentioned in the MathWorks website. After building the GUI, software called Terminal is used to connect the Serial Port of the computer to the caretaker's Bluetooth device. An android application was developed using Eclipse to create an appropriate interface for the caretaker. When the baby cries an alarm is triggered and the caretaker is altered. An alternative method is to link MATLAB's SIMULINK to Code Composer Studio (CCS) by following a series of step mentioned in MathWork's website.

Table 1: Threshold of different FE techniques

FEAT-URE	TIRED	HUNGRY	PAIN	NOISY
Pitch	300-450	300-450	300-450	100-500
MSF	<10	10 to 30	>30	0 to 10
Energy	>170	>250	>300	<150

Table 2: MSF Comparison

SL.NO	MAX MSF	MIN MSF	SIGNAL TYPE
1.	13.97	7.73	Tired
2.	13.91	5.59	Tired
3.	14.53	5.58	Tired
4.	45.91	2.64	Agitated
5.	57.15	0.73	Agitated
6.	40.89	3.78	Agitated
7.	26.86	3.05	Hungry
8.	29.65	9.65	Hungry
9.	30.94	4.46	Hungry

Table 3: Pitch of different Baby Cries

SL.NO	SIGNAL TYPE	USING STAC (Hz)	USING CEPSTRAL DOMAIN (Hz)
1.	Tired	470.6	425.9
2.	Tired	432.4	421.1
3.	Pain	444.4	408.5
4.	Pain	421.1	416.2
5.	Hungry	432.4	457.1
6.	Hungry	410.3	396.4

Table 4: Energy Comparison

SL. NO.	SIGNAL TYPE	ENERGY
1.	Tired	273.59
2.	Hungry	642.2
3.	Pain	375.36
4.	Noise	150.27

5 Conclusion

Signal Processing is very important in our day-to-day lives. It helps us develop numerous applications and one such application is the Baby Monitoring system. In a world where every parent is busy trying to make the ends meet, this application will prove to be very useful. Although this application cannot replace a caretaker, they at least do not have to constantly keep an eye on the baby. Feature extraction methods like Magnitude Sum function and Pitch Extraction proved to be very useful by giving us 100% accuracy. The use of traditional feature extraction methods like Mel Frequency Cepstrum Coefficients (MFCC), although providing us with a system that closely resembles the human speech and auditory system, does not give a good accuracy. This is mainly because of the various disadvantages associated with it like degradation even in the slightest presence of noise. Our project used a combination of three feature extraction methods namely: MSF, Pitch and Energy. Individually these techniques do not provide a very good accuracy but a combination of the three will yield accurate results. A major advantage of using this is that the computation is very minimal. Thereby, effectively reducing the computational time and if by using a DSP then the load on it is drastically reduced. The application presented is prototype. It can be modified to make it more sophisticated. We could have a Baby Monitoring System in the maternity ward of hospitals wherein multiple babies are monitored using just one system. In such cases, the use of DSP is imperative for faster and easier processing. Multiple inputs can be fed to the DSP by connecting the inputs to a Multiplexer whose output is connected to the DSP. The select lines of the Multiplexer should be changing periodically and if the microphone picks up any signal it should pass this signal to the DSP where the processing can be completed in a few milliseconds and the process continues. Another application of our project is to create a robot which responds to the needs of the baby according to the output. So if the baby is hungry, it could feed it with a bottle of milk. Similarly, it could change diapers but the risks involved is high and this could be implemented only way in the future where our resources permits us.

Acknowledgment

We sincerely wish to thank one and all who have helped us in this venture, as this project is the result of a co-operated team work whereby we have been accompanied and supported by many people. The satisfaction that accompanies the successful completion of any task would be incomplete without the mention of people who made it possible whose constant guidance and encouragement crowned our effort with success. Firstly, we would like to thank the Head of Department of Electronics and Communication, Amrita School of Engineering, Bangalore for her constant support and exposing us to the many application of signal processing. She inspired us to dwell into the subject further and learn more. We would like to take this opportunity to thank Mr Narendranath, Lab Assistant, Electronics and Communication, for assisting us through very practical aspects of our project and his willingness to help us even at the most inconvenient of times. Last, but not the least, we are grateful to Mr Gajendra Kumar and Mr Nitin Awasti of TATA Power SED for their efforts to convert our vision of the project into a working model. And also, we would like to take this opportunity to thank the ICTS department of Amrita School of Engineering for their co-operation and helping us procure and install software.

REFERENCES

[1] Manikandan, B.Venkataramani, M.Bhaskar, K.Ashish, R.Raghul and V.Mathangi, “Implementation of a Novel Phoneme Recognition System using TMS320C6713 DSP”, VLSI Design, 2010. VLSID '10. 23rd International Conference on, Bangalore, India, 3-7 Jan. 2010, pp-27-32

[2] S K Hasnain, Aresh Dinshaw Daruwalla and Ameer Ishrat Saleem, “A unified approach in Audio Signal Processing using the TMS320C6713 and Simulink Block sets”, Computer, Control and Communication, 2009. IC4 2009. 2nd International Conference on, Karachi, Pakistan, 17-18 Feb. 2009, pp 1-5

[3] J.Manikandan, B.Venkataramani, K.Girish, H.Karthic and V.Siddharth, “Hardware Implementation of Real-Time Speech Recognition System using TMS320C6713 DSP”, VLSI Design (VLSI Design), 2011 24th International Conference on, Chennai, 2-7 Jan 2011, pp 250-255

[4] Li Lin1, Dejun Guan1, Dongrui Zhang , Jinhan Feng, Lisheng Xu, “Refined Analysis of Heart Sound Based on Hilbert-Huang Transform”, Information and Automation (ICIA), 2012 International Conference on, Shenyang, 6-8 June-2012, pp 100-105

[5] D. Barda, P. Runeforsb and E. Arnbj¨ornssonc, “Newborn pain cry analysis based on pitch frequency tracking”, University Hospital Lund, 7-Jan-2010, pp 50-60

[6] Dror Lederman, “Automatic classification of infant’s cry”, Master’s Thesis, Ben- Gurion University of the Negev, Faculty of engineering sciences, Department of Electrical and Computer engineering.

[7] Shota Yamamoto, Yasunari Yoshitomi, Masayoshi Tabuse, Kou Kushida and Taro Asada, “ Recognition of a Baby’s Emotional Cry Towards Robotics Baby Caregiver”, International Journal of Advanced Robotic Systems, 5-Dec-2012, pp 20-25

[8] J.T.Tou And R.C.Gonzalez, *Pattern Recognition Principles*, Addison-Wesley, 1981

[9] Lawrence Rabiner and Biing-Hwang Juang, Fundamentals of Speech Recognition, Prentice Hall Signal Processing Series, 1993.