THE EXTRACTION OF ACOUSTIC FEATURES OF INFANT CRY FOR EMOTION DETECTION BASED ON PITCH AND FORMANTS

Rahmat Hidayati[1,2], I Ketut Eddy Purnama[1], Mauridhi Hery Purnomo[1]
[1]Department of Electrical Engineering, ITS, Surabaya, Indonesia
[2]Department of Electronic Engineering, EEPIS-ITS, Surabaya, Indonesia
e-mail : yati@elect-eng.its.ac.id, {ketut,hery}@ee.its.ac.id

Abstract – In this paper, we present the development of a system for translating the normal infant cries, which come from pain, sadness, hunger, fear and anger cry sounds, of ages from one day up to nine months old. The aim of this research is to analyse the sound of the crying infant, and to derive the reason why the infant is crying. In this experiment we used acoustic features characteristic determined by pitch and formants. The acoustic feature vectors are then clustered using K-means algorithm to determine the class or the reason of the cry. The proposed system perform well with the maximum accuracy of 90%.

***Keyword*:** Acoustic Feature Extraction**,** Emotion Detection, Infant Cry, Clustering, K-Means

I. INTRODUCTION

Crying of an infant is a way to express hungry, pain, discomfort (e.g. need to be changed) or unhappy. It is also used to let others know about their needs or problems. The cry is unique, however, mothers (and experienced nurses) recognize the sounds of their own infants.

Cry is the infant's most powerful, multimodal, dynamic behavior, and sometimes the only means of communication and sign of life a birth. It involves characteristic vocalization, facial expressions and limb movements, all of which change over time. It is in the most sensitive range of the human auditory sensation area. Several models of cry sound have been proposed. The theory that underlies most acoustic analyses of cry sounds is the sound-filter theory [9]. In this theory, cry is expressed by a waveform that impinges upon the listener's ear is a function of the characteristics of the source *(i.e.* the vibrating vocal cords) and its filters *(i.e.* the resonances of the supraglottal vocal tract and the radiation characteristics from the lips). Fig. 1 shows a universal schema of the voice-production by Gordos *et al.* [3].

Fig. 1A shows a schematized model of the human phonation system, while Fig. 1B illustrates the way of an acoustic signal from the vocal cords until the radiation.

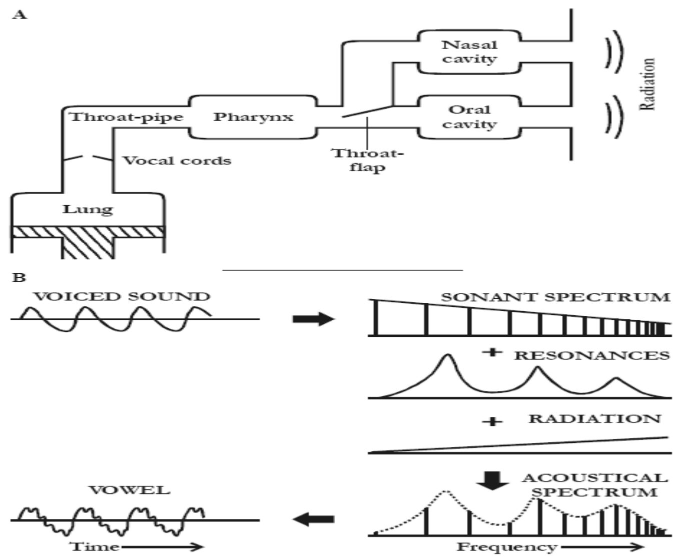


Fig. 1. A. universal schema for the voice-production by Gordos et al. [3]. B. Formation of a vowel in the human phonation system.

Golub's physioacoustic model of crying assumes three levels of central processing of the muscles contributing to the source and filters of crying [4]. These three levels are identified as the upper, middle and lower processors. The upper processor determine the state of the infant *(e.g.* fussiness), and the middle processor is involved with the infant's vegetative states, such as swallowing, coughing, digestion and crying. While the lower processor involves control of many muscle groups, including subglottal, supraglottal, glottal and facial muscles. These muscle groups are coordinated in the act of crying.

According to the statements mentioned above, crying is an important source of information about the infant needs or conditions, and it is benefecial to have such an automatic system to determine the cause of the infant's cry. In this paper, we propose an automatic system to determine whether the cry is caused by pain, hunger, anger, sadness or fear. Infant's cries are analyzed to obtain such features. Then by applying clustering method, the classes of features can be obtained. A class represents a cause of the cry. The proposed system is illustrated in Fig. 2.

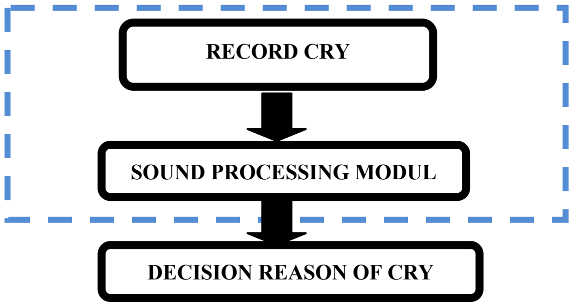


Fig. 2 Block diagram of the infant cry interpretation system

This paper is organized as follows. In *Method,* subjects and data acquisition are shown. Crying parameters and methods are described in *Sound Processing Module.* A new system for emotion detection is introduced in *Results.* Finally, concluding remarks and future aspects are given in *Conclusions and Future Work.*

II. MATERIALS AND METHOD

The cries of about 50 infants were recorded. The age of the infants was 1 day up to 9 months on the average. The recording process was done at the Dr. Soetomo Hospital, Surabaya and at a child clinic for the sick infants. The sound data was obtained from a public site [10]. Voice Pen IC Recorder (SONY ICD-TP620) was used to recognize the infant cry and the circumstances of the cry. A sampling frequency of 10 KHz for audio recording was chosen. The sampling type was PCM 16 bits mono, and there were 10 bits assigned to each sample. No sound processing method was applied. During the recording procedure, the infant was sitting on his or her mother's lip. The microphone was held approximately 6-10cm away from the infant's mouth, so as to avoid any clipping of the data. The time of a recording was 30 seconds on the average.

To cluster the features of the cries, K-means algorithm with Euclidian Distance was applied.

Infant cries and important information about the babies *(e.g.* name, age, predetermined auditory diagnosis, etc.) were stored in a special database. A special Matlab software package was developed to record and to analyze the cries.

Sound Processing Module

The sound processing module (Fig. 2) contains processes that can be illustrated in Fig. 3 [5]. The acoustical features that are important in this research were extracted from the cry sound by means of a sequence of sampling, pre-emphasizing, and windowing frames of the sound. Both static and dynamic features were determined such as LPC, Cepstrum, and Del Cepstrum.

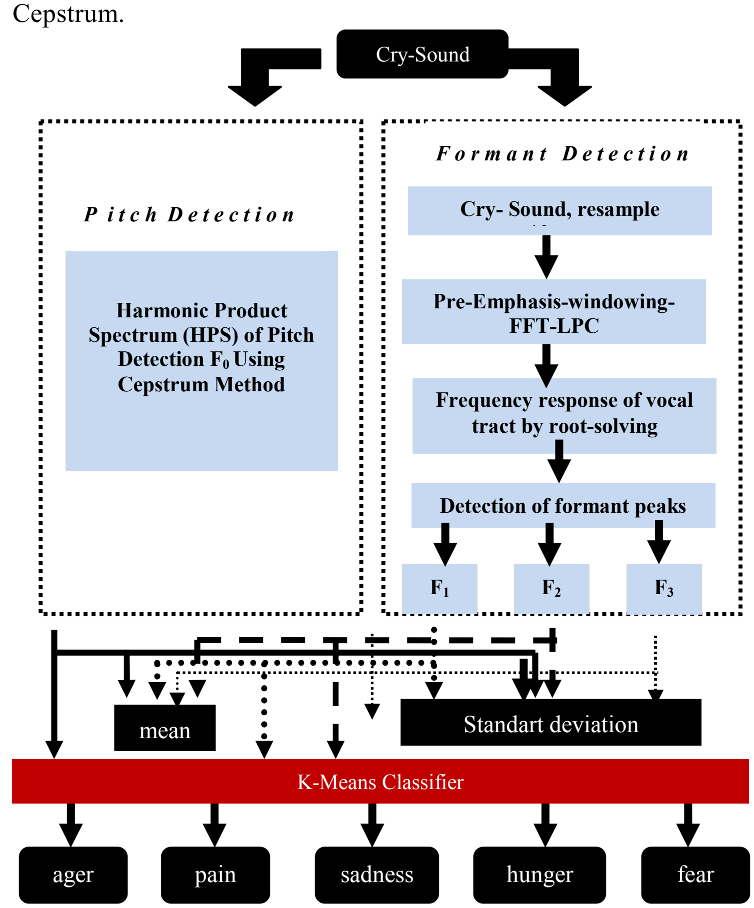


Fig. 3. Block diagram of the sound module

The fundamental frequency (F_0) and the first three formants F_1, F_2 and F_3 of infant vocalization [6][7] were also determined. These parameters contain important information regarding the emotional state of infant. Since F_0 for infant cries varies widely and rapidly, the Harmonic Product Spectrum (HPS) using cepstrum method was chosed to obtain F_0.

Decision of the Reason of Cry

The sound data is one of attributable to the same cause due to discrepancies in data acquisition and performance unreliability of the module. Therefore, the performance of a system that will decide the reason of why an infant cry should be reliable. In this research we performed an approach [6] that is illustrated in Fig. 6. The matrices that depicting the sound module performances are obtained by supplying individual classifiers with a number of cry data files. The performance of the module on the basis of hits and misses was noted. For the classifier, its truth has uncertainty. From the knowledge of its confusion matrix (3rd block of Fig. 4), such an uncertainty can be described by the conditional probabilities which can be formulated by Equation 1:

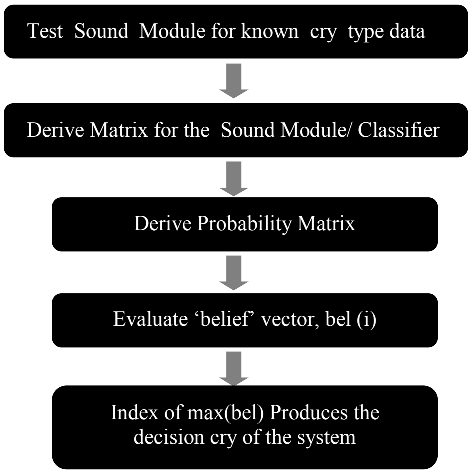


Fig.4. Block diagram for the decision cry sound process [8].

$$P(x \in C_i / e_k(x) = j) = \frac{n_{ij}^{(k)}}{\sum_{i=1}^{M} n_{ij}^{(k)}}, i = 1, \ldots, M \ldots\ldots\ldots\ldots\ldots(1)$$

Here, 'i' corresponds to known cry-types and 'j' corresponds to classified cry-types by individual classifier where M=5 (for the 5 different types of cries). Once individual conditional probability matrices for the two classifiers are obtained, a 'belief' vector can be derived as (the 4th block of Figure 4):

$$bel(i) = \eta \prod_{k=1}^{K} P(x \in C_i / e_k(x) = j_k) \ldots\ldots\ldots\ldots\ldots(2)$$

The index of the maximum value in the belief (column) vector, corresponds to decision of the system. For example, if the maximum value index is 1, then the fused decision is 'Pain', and if the maximum value index is 2, then the fused decision is 'hungry'. The maximum value index of 2,3,4, and 5, respectively correspond to the fused decision of ’hungry’, ‘fear’, ‘sad’, and ‘angry’.

III. RESULTS

Infant cry shows significant differences between several types of cry which can be perceptually distinguished by a trained person. The general acoustical features of several cry types in this experiment show raising-falling pitch pattern, ascending and descending melody with high intensity as shown in Fig.5. and Fig.6.

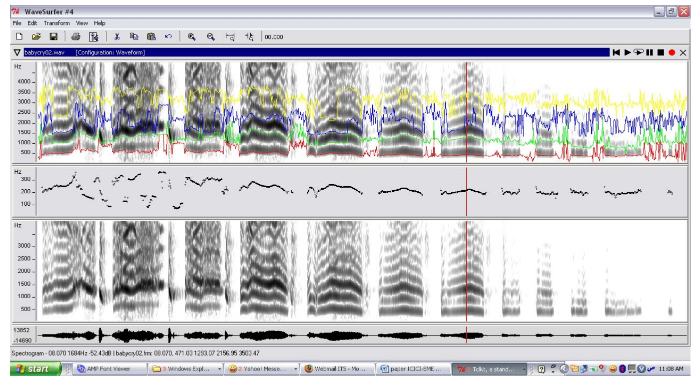


Fig.5. waveform, spectrogram and pitch pattern of sad crying

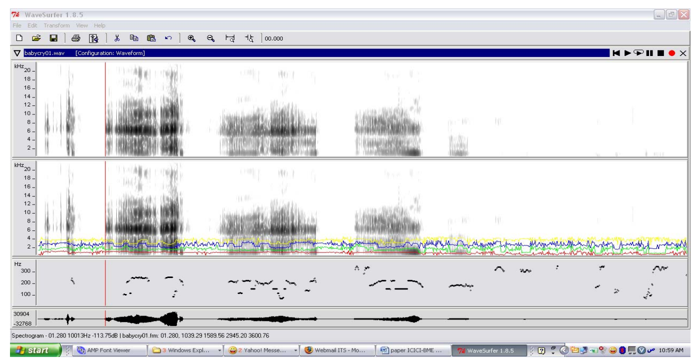


Fig.6. waveform, spectrogram and pitch pattern of hunger crying

The sample outputs of the sound processing module are shown in Fig. 7 until Fig. 11. Fig. 7 shows the sound of the original infant cry. Its spectrum is displayed in Fig. 8, the obtained fundamental frequency with cepstrum method, F0=258,06Hz is shown in Fig. 9. Formants spectrum of the sound of Fig. 7 is displayed in Fig. 10, while the output LPF infant cry sound is illustrated in Fig. 11.

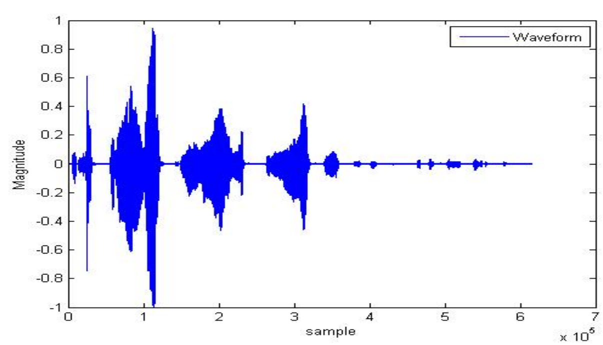


Fig.7. Original infant cry’babycry01.wav

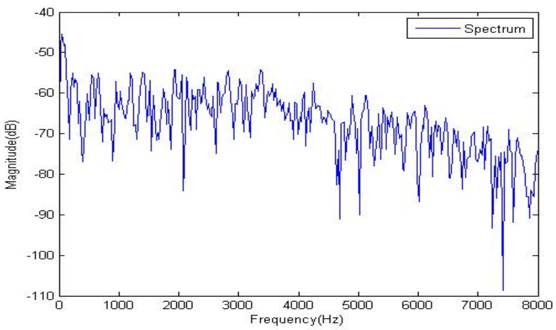


Fig.8. Spectrum of infant cry sound

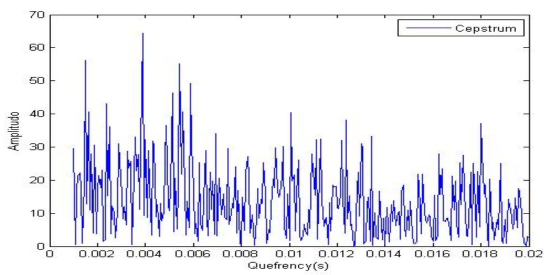


Fig.9. Fundamental frequency with cepstrum method, F_0=258,06Hz

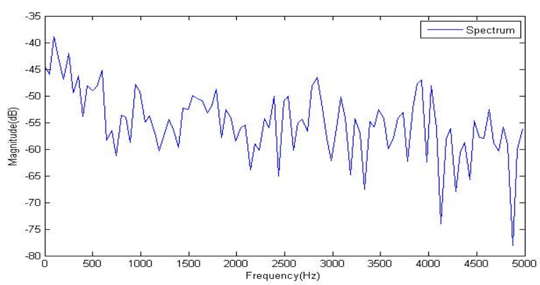


Fig.10. Formants spectrum of infant cry sound

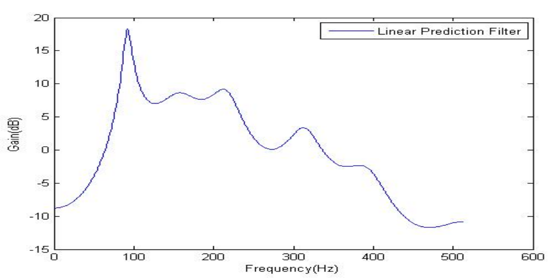
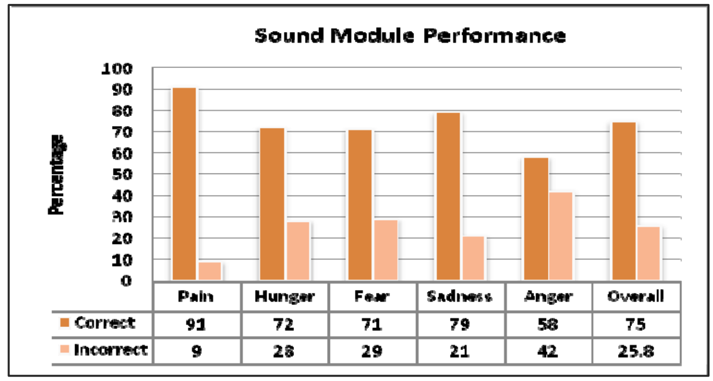


Fig.11. Output LPF infant cry sound

Here is Formants Frequencies output :

```
>> fformant
Formant 1 Frequency886.6
Formant 2 Frequency1513.1
Formant 3 Frequency2102.1
```

In Figure 12, the results of sound processing module show high accuracy for the 'pain' type of cries. This can be explained by the fact that this type of cry has a distinctive, very high fundamental frequency which is accurately detected by the pitch detection algorithm. The other types of cries have lesser distinct pitch and formants which lead to the formation of wider and more overlapping clusters, resulting in more errors in the overall detection process.



	Pain	Hunger	Fear	Sadness	Anger	Overall
Correct	91	72	71	79	58	75
Incorrect	9	28	29	21	42	25.8

Fig.12. Performance of Sound Module result show a very high accuracy for the pain type of cries 90%

IV. CONCLUSION AND FUTURE WORK

From the results it is concluded that the combination of F_0, F1 ,F_2 and F_3 produce the best clustering, according to the different cry types. Therefore this combination has been used in the sound processing module to distinguish the cry type of the infant cry sounds.

Based on experimental results, hence it can be concluded that

a. The result shows that the cry type of 'pain' has high accuracy 90% error 10%.
b. To get the accurate value of cries types we look from high or lower of fundamental frequency.
c. The decision of the reason of cry for each type of cry performs satisfactorily in the detection process.

The future work, experiments will be done more into multi feature for every type of cry and find new connections between illnesses and the infant cry. Then, classification into two different kinds of cries, normal and abnormal infant cry, will be determined using support vector machine. Further studies will develop a device capable to analyze crying and giving a help in diagnostics.

REFERENCES

[1] M. S. El-Nasr, T. R. Ioerger, J. Yen, D. H. House, and F. I. Parke, "Emotionally Expressive Agents", Proceedings of Computer Animation Conference 1999, Geneva, Switzerland; pp-48-57, 1999.

[2] S. D. Pollak, and P. Sinha, "Effects of Early Experience on Children's Recognition of Facial Displays of Emotion", Developmental Psychology 2002, Vol. 38, No. 5, 784-791

[3] G. Gordos, & Gy. Takács, “Digitális beszédfeldolgozás” (Műszaki Kiadó, Budapest, 1983).

[4] H.L. Golub,” A physioacoustic model of the infant cry and its use for medical diagnosis and prognosis” (Ph.D.Thesis, Massachusetts Institute of Technology, MA, USA,1980).

[5] H.E. Baeck, and M.N. Souza; "Study of Acoustic Features of Newborn Cries that Correlate with the Context", Engineering in Medicine and Biology Society, 2001. Proceedings of the 23rd Annual International Conference of the IEEE Volume 3, 25-28 Oct. 2001 Page(s):2174 - 2177 vol.3.

[6] M. Petroni, A.S. Malowany, C.C. Johnston, and B.J. Stevens, "A New, Robust Vocal Fundamental Frequency (F0) Determination Method for the Analysis of Infant Cries", Computer-Based Medical Systems, 1994., Proceedings 1994 IEEE Seventh Symposium on 10-12 June 1994 Page(s):223 - 228.

[7] M. Petroni, A.S. Malowany, C.C. Johnston, and B.J. Stevens, "A Cross-Correlation-Based Method for Improved Visualization of Infant Cry Vocalizations", Electrical and Computer Engineering, 1994. Conference Proceedings. 1994 Canadian Conference on 25-28 Sept. 1994 Page(s):453 - 456 vol.2.

[8] Pritam Pal, Ananth N, Iyer and Robert E, Yantoro, “Emotion Detection From Infant Facial and Expressions and cries”

[9] S. Möller, & R. Schönweiler, “Analysis of infant cries for the early detection of hearing impairment”, SpeechCommun 28, 1999, 175-193.

[10] http://fresound.org