A Database of Infant Cry Sounds
to Study the Likely Cause of Cry

Shivam Sharma[1], Shubham Asthana[2], V. K. Mittal[3]
[1,2,3]Indian Institute of Information Technology, Chittoor District, India
[1]shivam.sharma@iiits.in, [2]asthana.shubham@gmail.com and [3]vkmittal@iiits.in

Abstract

Infant cry is a mode of communication, for interacting and drawing attention. The infants cry due to physiological, emotional or some ailment reasons. Cry involves high pitch changes in the signal. In this paper we describe an 'Infant Cry Sounds Database' (ICSD), collected especially for the study of likely cause of an infant's cry. The database consists of infant cry sounds due to six causes: pain, discomfort, emotional need, ailment, environmental factors and hunger/thirst. The ground truth cause of cry is established with the help of two medical experts and parents of the infants. Preliminary analysis is carried out using the sound production features, the instantaneous fundamental frequency and frame energy derived from the cry acoustic signal, using auto correlation and linear prediction (LP) analysis. Spectrograms give the base reference. The infant cry sounds due to *pain* and *discomfort* are distinguished. The database should be helpful towards automated diagnosis of the causes of infant cry.

1 Introduction

Infant cry is an acoustic manipulation, that consists of different forms of vocalization, constrictive silence, coughing, choking and breaks (Neustein, 2010). Its analysis provides the information regarding health, disease, gender and emotions. In this way, infants communicate with their environment about what they feel. Infant cry falls in the most sensitive range of human auditory sensation (Jr et al., 2005).

First cry of an infant gives significant information to determine the APGAR count, a tool for the categorization of the new born baby as healthy, unhealthy or weak (Neustein, 2010). The time varying characteristics, limb movement and vocalization give insight to the transitional characteristics. Studies of physiological variables such as facial expression, muscular tonus, sleep and suction abilities were conducted to analyse the needs of an infant (Skogsdal et al., 1997). The study of infant cry has gained significance over the years, because of diverse applications.

Fundamental frequency and the first three formants of cry signals were studied earlier (Baeck and Souza, 2001); (Daga and Panditrao, 2011). Attempts were made to classify the cry signals on the basis of pain, sadness, fear and hunger etc. (Abdulaziz and Ahmad, 2010); (Mima and Arakawa, 2006). The pitch characteristics, aimed to classify the cry signals into categories: urgent, arousing, sick, distressing etc. (Zeskind and Marshall, 1988). A pitch detection algorithm was used to compute the instantaneous fundamental frequency (F_0) (Neustein, 2010). First three formants along with F_0 were used for the cry analysis (Hidayati et al., 2009). Cepstrum analysis was carried out for determining the F_0 and first three formants for cry (Chandralingam et al., 2012). A time domain cross-correlation pitch determination method gave details of every pitch epoch (Petroni et al., 1994).

Spectrographic analysis was carried out for characterizing pitch and harmonics (Neustein, 2010). Short-time energy, zero-crossing rate and LP coefficients were used for the analysis of the cry signals (Kuo, 2010). A cry is described by the duration and shape of the F_0 contour (Varallyay-Jr, 2007). Hyper-phonation was examined by

mean, standard deviation and peak of the fundamental frequency (Zeskind and Marshall, 1988). In another study, segment density, segment length and pause length were used to examine the relationship between these parameters and gender of the baby (Varallyay, 2006). The frame-wise comparison of the mean F_0 gave insight into the analysis of cries of infants with different heart disorders (Chandralingam et al., 2012). This study also correlated disphonation to muscle *pain* or *discomfort*.

A different method to extract F_0 was proposed (Petroni et al., 1994). The F_0 for cry was observed to be between 200-500 Hz, with average F_0 320 Hz for male infant and 400 Hz for female infant (Daga and Panditrao, 2011). Typical characteristics were observed in the shape of power spectra of cry signals for hunger, sleepiness and discomfort (Mima and Arakawa, 2006). Cries of infants were divided into *normal* and *disorders* like Tetralogy of Fallot (TOF), Ventricular Septal Defect (VSD), Atrial Septal Defect (ASD), and Patent Ductus Arteriosus (PDA) (Chandralingam et al., 2012). Specific segments of cry signals showed similarity amongst infants with hearing disorders and normal infants. However, detailed study of an infant cry, in conjunction with likely causes as per expert medical opinion, is still required to be carried out in detail.

In this paper, an 'Infant Cry Sounds Database' *(IIIT-S ICSD)* is described in detail. The database consists of different categories of cries due to different reasons. The data of infant cry signals was recorded in a doctors cabin, where the infants were brought-in for routine check-up, vaccination trips or due to any ailment. Spectrographic study was carried for each case. Changes in the F_0 contour and harmonics were observed for cry causes. An effort is made for the classification of cry signals, based upon the spectrographic analysis and using signal processing methods. The analysed causes of each cry are compared with the ground truth determined as per doctors or parents.

This paper is organized as follows. Section II discusses details of the *IIIT-S ICSD*. Section III discusses signal processing methods for analysing the infant cry signals. Features of the Infant cry signals are discussed in Section IV. In Section V, the observations are made from the results of infant cry analysis. The paper is summarised in Section VI, along with the scope of future work. 113

Table 1: Template for naming files in IIIT-S ICSD

SPKR01_M_S1a_CRY07	
(a) Symbols	*(b) Interpretation*
SPKR#	The infant number (Ex: 01)
M/F	Sex of the infant (Ex: Male)
S#a	Session number and session subpart (Ex: 1a)
CRY#	Number of cries in the session being considered (Ex: 07)

2 The IIIT-S ICSD Cry

2.1 Data Collection

The data was collected from Pranaam hospital, Madinaguda, Hyderabad, under the supervision of Dr. Manish Gour (MBBS, DCH) and Dr Nizam (MBBS). The age group of infants was restricted between 3 months and 2 years. The cry signals of infants were collected during the regular check-up visits, the vaccination trips or any emotional need of attention. People present in the room were requested to maintain silence, so as to record the cry sample. Also the parents were advised not to comfort the baby for brief duration, to ensure the uninterrupted data collection. Along with data, the personal details noted include: infant name, parent name, parent profession, sex and age of the infant and predictive causes of the cry.

For the recording purpose, Roland R-09 Wave/MP3 recorder was used and was placed at 10-20 cm from the infant's mouth. Precautions were taken to avoid any unwanted noise or crosstalk. Sampling rate of 48 KHz, with 24 bit coding rate, was used for recording in stereo mode. There were no interruptions from the social environment during the data recording. The only unwanted noise that could overlap the cry sound may be from fan and air-conditioner. The ambient temperature during the recordings was 38°C, which was regulated by the air conditioner at 25°C.

2.2 Organization of the IIIT-S ICSD

The terminologies used in naming of the database files, given in *Table 1*, are described below:

- *Session*: The acoustic signal, right from the time an infant starts crying (including all inhalations and exhalations), until the infant becomes quiet, is a *session*.
- *Session subpart*: Each session consists of *subparts*, characterised by the contiguous set

Table 2: *Causes of Infant cry (in IIIT-S ICSD)*

(a) Cry Causes	(b) Description
1. Pain	Cry due to pain (caused by vaccination, physical hurt or internal pain)
2. Discomfort	Cry due to irritation caused by the external environment (e.g., the doctor opening baby's mouth to pour-in drops, or the vaccination)
3. Emotional Need	Cry when the baby wants to go back to parents arms
4. Ailment	Cries due to any ailments like cold, cough, fever
5. Environmental factors	Cry due to fear of the surroundings or change in environmental conditions.
6. Hunger/Thirst	Cry when the baby is hungry or thirsty

Table 3: *Summary of contents in IIIT-S ICSD*

Attributes	Values
Total # of files	76
Total # of speakers	33
Total # of session	76
Total # of cries in sessions	693
Average # of sessions per speaker	2.3
Average # of cries in each session	9.1
Total duration of all sessions	670.1 s
Average duration of each session	8.817 s

of signals, separated by some noise
- *Cry*: Each session subpart comprises of a number of cries, separated by some noise.

A two stage process was followed for data collection in the study. The first stage involved *raw data collection* at the hospital, and the second stage included *pre-processing*. The unwanted noise, in the raw data was removed using 'WaveSurfer' tool, to render it cross-talk free. The cries were categorized as per the ground reality, i.e., the actual cause as per the doctor or parent. The main causes of cry that we came across, are described in *Table 2*, in columns (a) and (b).

The cry as combination of two or more causes is retained in a separate category. There was also a special case, when an infant cried by listening to another infant cry. Not many samples could be obtained for this cry due to *Domino effect*, but this is retained as another special category, for future study. The database consists of total 76 cry sound files, which can be categorised in 6 classes. The database summary is given in *Table 3*.

3 Signal Processing Methods

3.1 Short Time Fourier Transform

Short time Fourier Transform (STFT) (Oppenheim et al., 1989) is used to process the segments of the cry signals in the frequency domain. The data is divided into overlapping frames. The Fourier

transform for each frame is given by:

$$X(\tau, \omega) = \sum_{n=-\infty}^{n=\infty} x[n]w[n-m]e^{-jwn} \quad (1)$$

Where *x[n]* is the signal and *w[n]* is the window function. Here *m* is discrete and ω is continuous (Oppenheim et al., 1989). Magnitude of the STFT gives the *spectrogram*, i.e., $|X(\tau, \omega)|^2$. Spectrographic analysis basically represents the 3D spectral information obtained from the magnitude spectrum, for the short-time overlapped window segments. The *X-axis* represents time, *Y-axis* represents frequency, and the third dimension represents the log magnitude of the sinusoidal frequency components, which is converted to the proportional intensity.

3.2 Auto Correlation Analysis

The *auto correlation function* (Rabiner and Juang, 1993) reflects the similarity between a random sequence and the time-delayed same sequence. For the speech signal *x(n)*, the *correlation* function is defined as

$$r_x(m) = E[x(n)x(n+m)]$$
$$= \lim_{N \to \infty} \frac{1}{2N+1} \sum_{n=-N}^{n=N} x(n)x(n+m)$$
$$(2)$$

Where, *E[.]* represents statistical expectations (Rabiner and Juang, 1993). Since speech production system can be regarded as stationary within short-time frames, the cry signal is divided into frames. The auto-correlation function is:

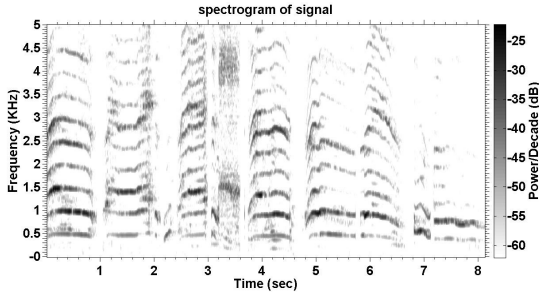$$r'_x(m) = \frac{1}{2N+1} \sum_{n=-N}^{n=N} x(n)x(n+m) \quad (3)$$

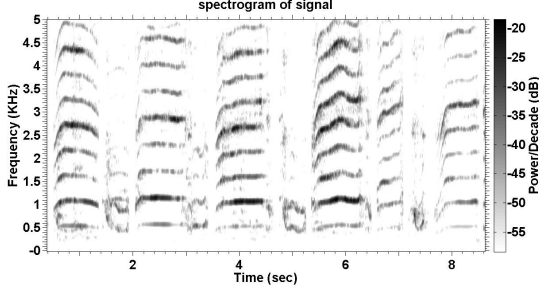Figure 1: Spectrogram of *Pain Cry* of Infant #1
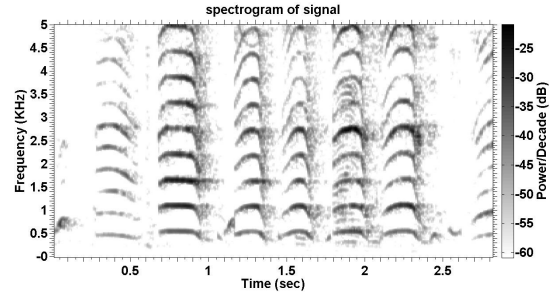

Figure 3: Spectrogram of *Pain Cry* of Infant #2

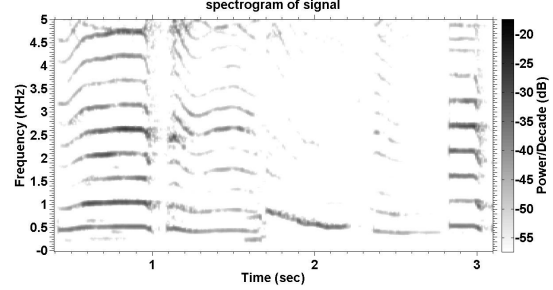Figure 2: Spectrogram of *Discomfort Cry*, Infant #1

Figure 4: Spectrogram of *Discomfort Cry*, Infant #2

When *m* = 0, then (3) represents *short-term energy* of the signal (Shuyin et al., 2009). The pitch period information is more pronounced in autocorrelation, than in speech signal.

3.3 Linear prediction (LP) analysis

Speech signal is produced by the convolution of the excitation source and time-varying vocal tract system components. These components can be separated, using *LP analysis* (Makhoul, 1975). The prediction of current sample as a linear combination of past p samples forms the basis of LP analysis, where p is the order of prediction. The predicted sample is $\hat{s} = -\sum_{k=1}^{p} a_k s(n-k)$, where $a_k s$ are the *LP coefficients*. Here $s(n)$ is the *windowed speech* sequence, obtained by multiplying short-time speech frame with a Hamming window, i.e., $s(n) = x(n)w(n)$, where $w(n)$ is the windowing function. *The prediction error e(n)* can be computed by the difference between the actual sample $s(n)$ and the *predicted* samples $\hat{s}(n)$ (Rabiner and Juang, 1993). This is given by $e(n) = s(n) - \hat{s}(n)$, i.e., $e(n) = s(n) + \sum_{k=1}^{p} a_k s(n-k)$.

The LP coefficients are used to minimize the *prediction error e(n)*. The coefficients are computed using the *least squares* method, that minimizes the LP residual or *total prediction error* (Makhoul, 1975). In frequency domain, it can be

represented as $E(z) = S(z) + \sum_{k=1}^{p} a_k S(z)z^{-k}$.

$$A(z) = \frac{E(z)}{S(z)} = 1 + \sum_{k=1}^{p} a_k z^{-k} \qquad (4)$$

The LP Spectrum *H(z)* can be obtained as,

$$H(z) = \frac{1}{1 + \sum_{k=1}^{p} a_k z^{-k}} = \frac{1}{A(z)} \qquad (5)$$

As *A(z)* is the reciprocal of *H(z)*, the LP residual is obtained by inverse filtering of the speech. The auto-correlation of LP residual can also give information about the pitch period.

4 Features explored

In this study, correlation between infant cry and the likely causes is examined using mainly the magnitude of short-time spectrogram and F_0. Spectrograms are used for obtaining the F_0 and harmonics as *Ground Truth*.

In this paper, F_0 is obtained by using two different methods to validate the results and the modified signal processing methods used. Autocorrelation and Linear Prediction analysis methods are used after modification, to detect peaks in the auto-correlation or LP residual. In both methods, the signal was divided into frames of 10 ms, with a shift of 6ms. The F_0 contour was smoothed further, using a binary mask derived using the signal energy for each frame. The results from auto-correlation and LP analysis are compared with the
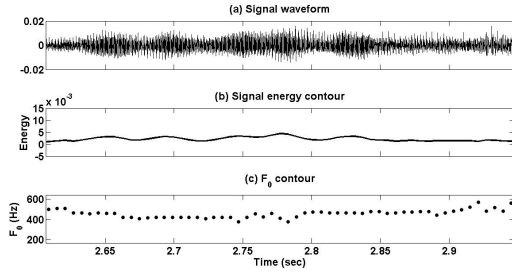
Figure 5: Illustration of (a) signal waveform, (b) signal energy contour and (c) F_0 contour of *Pain Cry* of infant #1 (using *auto-correlation*)
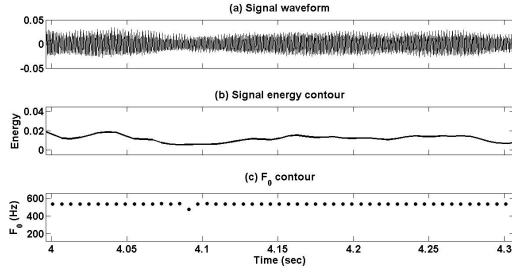

Figure 7: Illustration of (a) signal waveform, (b) signal energy contour, (c) LP residual and (d) F_0 contour of *Pain Cry* of infant #1 (using *LP analysis*)

Figure 6: Illustration of (a) signal waveform, (b) signal energy contour and (c) F_0 contour of *Discomfort Cry* of infant #1 (using *auto-correlation*)

Figure 8: Illustration of (a) signal waveform, (b) signal energy contour, (c) LP residual and (d) F0 contour of *Discomfort Cry*, infant #1 *(LP analysis)*

ground truth, which is obtained using the spectrograms. Similar patterns of changes in the F_0 contours are observed for the cries due to similar causes, in particular for *pain cry* and *discomfort cry*.

5 Preliminary Analysis

The changes in F_0 contour and harmonics for infant cry signals were observed using the spectrograms. The spectrograms were plotted for cry signal frames of 30 ms with shift of 9ms, for the same infant, crying for different causes. The F_0 contours and harmonics of a *pain cry* (Fig. 1 and Fig. 3) indicate *cyclic changes* with larger fluctuations. These could be due to psychological conditions during *pain*. The shape of F_0 contour in the case of cries due to *discomfort* (Fig. 2 and Fig. 4) have relatively *flat* nature, with changes in F_0 at larger periods and having lesser fluctuations.

The F_0 contours derived using auto-correlation and LP analysis were observed in similar way. Similar to the observations from spectrograms, these F_0 contours show *cyclic* changes with larger fluctuations for cries due to *pain*, in Fig. 5 and Fig. 7. In the case of cries due to *discomfort*, the
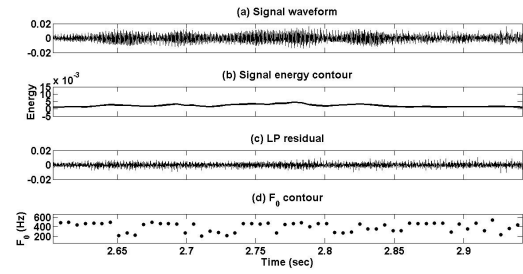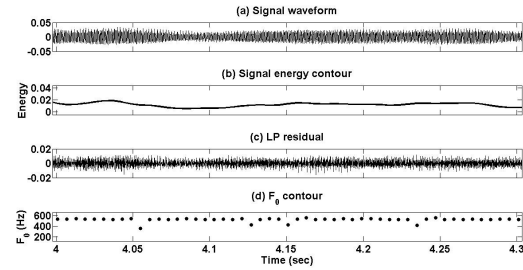
F_0 contours are relatively *flat* in Fig. 6 and Fig. 8, with lesser fluctuations.

6 Summary and Conclusions

In the study of infant cry signals the database collection and collation plays a vital role, as the entire analysis depends upon it. Hence, in this paper emphasis is laid upon collection and organization of infant cry data, for the classification of infant cries according to different causes. The predictive reasons for infant cries were noted as per the inference by doctor and parents. Details of the IIIT-S ICSD database are also elaborated.

In this paper, the spectrographic analysis of the infant cry is carried out, by observing changes in the F_0 contour and harmonics, to determine the likely cause of the cry. Signal processing methods such as auto-correlation and linear prediction analysis are used after modification, in order to observe some particular patterns of infant cry. Possible leads, for association with any kind of ailment the infant is suffering from, are explored. The observations from spectrograms are similar to the changes in F_0 contours, derived using auto correlation and LP analysis, for cries due to same cause. The F_0 contours have *cyclic changes* with

larger fluctuations for *pain cry*, and are relatively *flat* with lesser fluctuations for *discomfort cry*. The observations are consistent across different infant cries, and signal processing methods.

Further, we intend to utilize this database to examine the cry characteristics using other signal processing methods and carry out quantitative analysis in detail. This study would be helpful towards enabling the early diagnostics and medical care, if the reason of the infant cry due to a particular ailment is established early, especially where the reaction time could be critically important.

Acknowledgement

References

Y. Abdulaziz and S.M.S. Ahmad. 2010. Infant cry recognition system: A comparison of system performance based on mel frequency and linear prediction cepstral coefficients. In *Int. Conf. on Information Retrieval Knowledge Management, (CAMP)*, pages 260–263, March.

H.E. Baeck and M.N. Souza. 2001. Study of acoustic features of newborn cries that correlate with the context. In *Engineering in Medicine and Biology Society, 2001. Annual IEEE Int. Conf. of the 23rd Proceedings*, volume 3, pages 2174–2177.

S. Chandralingam, T. Anjaneyulu, and K. Satyanarayana. 2012. Estimation of fundamental and formant frequencies of infants cries; a study of infants with congenital heart disorder. In *Indian Journal of Computer Science and Engineering*, volume 3(4), pages 574–582.

Raina P. Daga and Anagha M. Panditrao. 2011. Article: Acoustical analysis of pain cries in neonates: Fundamental frequency. *IJCA Special Issue on Electronics, Information and Communication Engineering*, ICEICE(3):18–21, December. Full text available.

R. Hidayati, I.K.E. Purnama, and M.H. Purnomo. 2009. The extraction of acoustic features of infant cry for emotion detection based on pitch and formants. In *Int. Conf. onInstrumentation, Communications, Information Technology, and Biomedical Engineering (ICICI-BME)*, pages 1–5, Nov.

G. Varallyay Jr, Z. Benyo A. Illenyi, Z. Farkas, and G. Katona. 2005. The speech-chorus method at the analysis of the infant cry. In *Acoustic Review*, volume 6(2), page 915.

K. Kuo. 2010. Feature extraction and recognition of infant cries. In *IEEE Int. Conf. on Electro/Information Technology (EIT)*, pages 1–5, May.

J. Makhoul. 1975. Linear prediction: A tutorial review. *Proceedings of the IEEE*, 63(4):561–580, April.

Y. Mima and K. Arakawa. 2006. Cause estimation of younger babies' cries from the frequency analyses of the voice - classification of hunger, sleepiness, and discomfort. In *International Symposium on Intelligent Signal Processing and Communications, IS-PACS '06*, pages 29–32, Dec.

Amy Neustein, editor. 2010. *Advances in Speech Recognition: Mobile Environments, Call Centers and Clinics*. Springer, New York.

Alan V Oppenheim, Ronald W Schafer, John R Buck, and et al. 1989. *Discretetime signal processing*, volume 2. Prentice-hall Englewood Cliffs.

M. Petroni, M.E. Malowany, C.C. Johnston, and B.J. Stevens. 1994. A new, robust vocal fundamental frequency (f_0) determination method for the analysis of infant cries. In *Computer-Based Medical Systems, Proceedings IEEE Seventh Symposium*, pages 223–228, June.

Lawrence Rabiner and Biing-Hwang Juang. 1993. *Fundamentals of Speech Recognition*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA.

Zhang Shuyin, Guo Ying, and Wang Buhong. 2009. Auto-correlation property of speech and its application in voice activity detection. In *First Int. Workshop on Education Technology and Computer Science, ETCS '09*, volume 3, pages 265–268, March.

Y Skogsdal, M Eriksson, and J Schollin. 1997. Analgesia in newborns given oral glucose. *Acta Pdiatrica*, 86(2):217–220.

Gyorgy Varallyay-Jr. 2007. The melody of crying. *Int. Journal of Pediatric Otorhinolaryngology*, pages 1699–1708.

Gyorgy Varallyay. 2006. Future prospects of the application of the infant cry in the medicine. In *Electrical Engineering*, volume 50(1-2), pages 47–62.

Philip Sanford Zeskind and Timothy R. Marshall. 1988. The relation between variations in pitch and maternal perceptions of infant crying. *Child Development*, 59(1):193–196.