SLPAT 2016 Workshop on Speech and Language Processing for Assistive Technologies
13 September 2016, San Francisco, USA

Discriminating the Infant Cry Sounds Due to Pain vs. Discomfort Towards Assisted Clinical Diagnosis

Vinay Kumar Mittal

Indian Institute of Information Technology Chittoor, Sri City, India
vkmittal@iiits.in

Abstract

Cry is a means of communication for an infant. Infant cry signal is usually perceived as a high-pitched sound. Intuitively, significant changes seem to occur in the production source characteristics of cry sounds. Since the instantaneous fundamental frequency (F_0) of infant cry is much higher than for adults and changes rapidly, the signal processing methods that work well for adults may fail in analyzing these signals. Hence, in this paper, we derive the excitation source features F_0 and strength of excitation (SoE) using a recently proposed *modified zero-frequency filtering method*. Changes in the production characteristics of acoustic signals of infant cries due to *pain* and *discomfort* are examined using the features F_0, SoE and signal energy. These changes are validated by visually comparing their spectrograms with the spectrograms of the acoustic signals. Effectiveness of these discriminating features is examined for different pain/discomfort cry sounds pairs in an 'Infant Cry Signals Database (IIIT-S ICSD)', especially collected for this study. Fluctuations in the features F_0, SoE and energy are observed to be larger in the case of infant cry due to *pain*, than for *discomfort*. These features can help in developing further the clinical assistive technologies for discriminating different infant cry types and initiating the remedial measures automatically.

Index Terms: Infant cry signals, modified zero-frequency filtering, pain cry, discomfort cry, excitation source characteristics

1. Introduction

Whenever an infant cries, his/her mother invariably knows the reason why her baby is crying. Infants *cry* to communicate either their some need or condition that could be physiological, pathological or environmental. In today's fast paced life, there are situations routinely, when the infants are under the physical/medical care of people other than the parents. For such conditions, if the assistive technologies could be developed that can help the parents and other care-taking people know the cause of an infant's crying, several diverse applications can then be evolved. It could also assist clinical diagnosis of any medical condition that an infant may be suffering from. But, this necessitates characterising the changes in the acoustic signal of infant cry, and also possibly understand the differences in the features of infant cry signal due to different causes. This paper aims at exploring few such discriminating features of infant cry signal.

Infant cry is a combination of vocalization, constrictive silence, coughing, choking and interruptions [1]. It provides information about the health, gender, disease and emotions etc. of the infant. The first cry of an infant is an important parameter in determining the Apgar count, which can be used to classify neonates into healthy and unhealthy (or weak) [1]. This is the first tool of communication and the sign of life at birth. Various characteristics of the first cry are vocalizations, facial expressions and limb movements, all of which change over time. Infants cry to let others know about their problems or needs, just like adults do by talking. Thus infant cry falls in the most sensitive range of the human auditory perception [2]. Physiological variables such as, facial expressions, muscular tonus, sleep and suction abilities have been studied as parameters to estimate the needs of an infant [3]. The study of infant cry has gained significance over the years, for diverse applications including early detection of the cause of cry and the possible ailment.

Infant cry signal, produced in response to a stimuli, involves a rhythmic pattern of cry sounds and inhalation. An important feature used for the analysis of infant cries in most studies is the instantaneous fundamental frequency (F_0). The fundamental frequency and its first three harmonics were studied [4, 5]. Infant cries were attempted to be classified on the basis of pain, sadness, hunger, fear and other few causes [6, 7, 8, 9]. Pitch characteristics of infant cries were categorized into urgent, arousing, distressing or sick, using linear predictive (LP) coding [10]. Pitch measurements at every epoch were taken, using a time-domain based cross-correlation [11]. The start and end time of each cry segment were detected, using short-time energy function and zero-crossing rate [8, 12].

Spectrographic analysis of cry signal was carried out to characterize pitch and its harmonics [1]. First three formants along with fundamental frequency were used for the analysis in [7, 13]. Cepstrum analysis was used for extracting the fundamental frequency, along with LP analysis for extracting the first three formants [13]. Features such as short-time energy, zero-crossing rate and linear prediction cepstral coefficients (LPCCs) were used for the analysis of cry signals [8]. Parameters such as mean, standard deviation and peak value of the fundamental frequency were used to examine hyper-phonation [10]. Parameters such as segment density, segment length and pause length were used to study their relation with the gender of the baby [12]. Analyses of cries of infants with different heart disorders were carried out by comparing frame-wise mean F_0, and minimum and maximum F_0 values [13]. The parameters such as duration, fundamental frequency and the shape of F_0 contour were also explored to describe a cry [14].

In infant cry, the harmonic structure was observed to be related to abnormalities in larynx, and dysphonation due to muscle pain or discomfort [1]. The LPCC magnitudes for cries due to similar causes were observed to be similar, and different for cries due to different causes [8]. Typical characteristics from the shape of power spectra of signals were obtained for cries due to hunger, sleepiness and discomfort, with classification accuracy of 85% [9]. Segment length analysis of cry signals had shown similarity among normal infants as compared to those with hearing disorder [12]. Cries of infants were divided into normal,

10.21437/SLPAT.2016-7

and due to disorders such as Tetralogy of Fallot, Ventricular Septal Defect, Atrial Septal Defect, and Patent Ductus Arteriosus [13]. However, the production characteristics of the acoustic signals of infant cries due to different causes have not been studied much. Our preliminary study of infant cry signals [15] also indicated the need for examining in detail their production characteristics, where intuitively larger changes seem to occur. However, evolving the signal processing methods for reliable F_0 extraction from the infant cry signal remains a challenge.

This paper focuses on examining changes in the acoustic signals of infant cries due to different causes, from their production point of view. Especially, the excitation source characteristics derived from the acoustic signals of infant cries due to *pain* and *discomfort* are analysed. An *Infant Cry Signals Database (IIIT-S ICSD)* [16] is used. The database consists of infant cries due to six different causes. But, due to better availability of multiple pain/discomfort cry sound pairs produced by same infant, the discriminating features are examined for infant cry sounds due to *pain* and *discomfort* categories of cry-causes. Changes in the signals are examined using three production features, namely, F_0, strength of excitation (SoE) and signal energy. The excitation source features F_0 and SoE are extracted using the *modified zero-frequency filtering (modZFF)* method [17, 18, 19]. Effectiveness of the discriminating features derived using the *modZFF method* is validated by visually comparing the spectrograms of an excitation source feature, the SoE impulse sequence, with that of the acoustic signal. Results indicate larger changes in the production features of infant cry due to *pain*, than for *discomfort*. These discriminating features can be used in developing further the assistive systems for clinical diagnosis of infant cry sounds, with wide ranging applications.

This paper is organized as follows. In Section 2, the details of the data collected are discussed. The *modZFF method* used for extracting the excitation source features from the infant cry signals, is described briefly in Section 3. Section 4 analyses the production characteristics of acoustic signals of infant cries due to pain and discomfort. Results are discussed in Section 5. Section 6 gives a summary, along with scope of further work.

2. Data for the study

The data of acoustic signals of infant cries due to different causes was especially collected for this study. This database is named as the *Infant Cry Signals Database (IIIT-S ICSD)* [16]. The infant cry signals in the ICSD were recorded in a private paediatric hospital, under the supervision of two medical experts (paediatricians). The infant cry data was recorded for infants needing routine check-up, vaccination or cure to some ailment. The recordings were made in multiple sessions in the doctor's room whenever an infant was brought-in for regular check-up, vaccination or for cure of some ailment, and the infant cried because of pain, ailment, discomfort, emotional need, change of environment or hunger/thirst etc. Infant cries were categorized as per the doctors and parents, into six classes of cause factors [16], as elaborated in Table 1. The cry data was recorded for infants in the age group of 3 months to 2 years.

The acoustic signals' data in the ICSD was recorded using a Roland Edirol R-09 Wave/MP3 recorder, placed at around 10-20 cm from infant's mouth. The data was recorded in stereo mode, at a sampling rate of 48 kHz and 24 bit/sample coding. People in the doctor's room were requested not to speak (for few secs) during recording of the cry data. Parents were requested not to make any efforts to calm the baby for a short period of time, so as to record the clean signals. Data was further preprocessed manually by listening carefully, and using the software tools such as Wavesurfer, Audacity and MATLAB to make it free from any noise or any overlapping speech sounds. The *IIIT-S ICSD* consists of 693 infant cry samples of 33 speakers (infants), recorded for total about 670 sec, that are stored in 76 files [16]. This *IIIT-S ICSD* database collected for the purpose of research towards evolving the speech sounds based assistive technologies, can be made available on request.

Table 1: *Categories of Possible Causes of Infant Cry*

(a) *Causes*	(b) *Description of Causes of Infant Cry*
1. Pain	Cry due to internal pain, or external pain caused by vaccination or any physical hurt on the body
2. Ailment	Cries due to any ailment such as cold, cough or fever etc.
3. Discomfort	Cry due to irritation caused by the external factors, e.g., the doctor opening baby's mouth (investigation) or nurse holding the baby (vaccination)
4. Emotional need for attention	Cry when the baby has emotional need to go back into parents arms and feel their touch, or need cuddling
5. Environmental factors	Cry due to fear of the surroundings or need a change in the environment (e.g., need for changing the diapers)
6. Hunger/thirst	Cries due to hunger or thirst

The study of relative changes in the production characteristics of infant cry sounds due to different causes requires availability of acoustic signals data of cry sounds due to different causes, produced by the same speaker, i.e., an infant. Since, multiple pairs of acoustic signals of pain/discomfort cry sounds produced by same infant are better available in the ICSD, the discriminating features are examined in this paper for *pain* and *discomfort* categories of infant cry sounds. Though, the *IIIT-S ICSD* consists of acoustic signals of infant cries due to six categories of different causes, as diagnosed by medical experts and the parents (see Table 1). The data-pairs of *pain* and *discomfort* cry signals both produced by same infant are chosen from the *IIIT-S ICSD* [16]. These are examined using a recently proposed signal processing method, the *modZFF method*, discussed in the next section.

3. Deriving the excitation source features using Modified Zero-Frequency Filtering

Changes in the acoustic signals of infant cry sounds are examined in this study using three production features, namely, the instantaneous fundamental frequency (F_0), the strength of excitation (SoE) and frame-wise signal energy (E). Production of acoustic signal of infant cry apparently involves significant changes in the glottal excitation source characteristics [15]. Hence, the excitation source features F_0 and SoE are focused in this study. Though the feature F_0 can be derived from the acoustic signal using the autocorrelation or linear prediction residual methods, but those methods give only the indicative results, good for preliminary investigation [15]. In this study, the excitation source characteristics F_0 and SoE are extracted from the acoustic signal of infant cry using the *modified zero-frequency filtering (modZFF)* method [17, 18, 19]. Differences

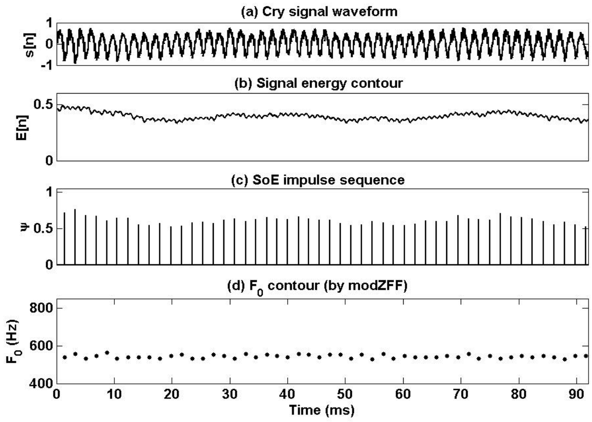

Figure 1: Illustration of features for *discomfort cry signal*: (a) cry signal waveform, (b) signal energy contour, and (c) SoE impulse sequence and (d) $F0$ contour derived using the *modZFF* method, for (discomfort) cry signal of infant #S02 (male).

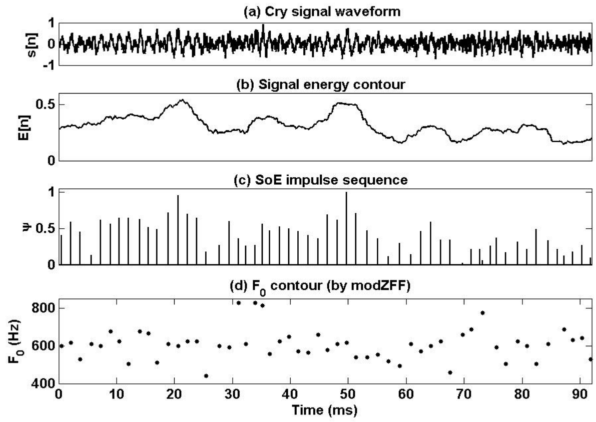

Figure 2: Illustration of features for *pain cry signal*: (a) cry signal waveform, (b) signal energy contour, and (c) SoE impulse sequence and (d) $F0$ contour derived using the *modZFF* method, for (pain) cry signal of infant #S02 (male).

of changes in these features are examined using their mean and standard deviation. The infant cry cause assessed by the doctor and parents is used as base reference. Effectiveness of the discriminating features, mainly the excitation source features, is validated by visually comparing the spectrograms of the SoE impulse sequence and the acoustic signal.

The excitation source characteristics can be derived from the normal speech signals, using the zero-frequency filtering (ZFF) method [17, 18]. But use of the ZFF method to derive the excitation source features from the acoustic signal of *nonverbal sounds* such as *infant cry*, that have significant changes in their source characteristics, pose two limitations [19]. (i) Shorter window length would be required for trend removal operation. (ii) Impulse sequence for aperiodic signals may be affected by the choice of shorter window length. Both of these limitations are addressed in the *modified zero-frequency filtering (modZFF)* method [19, 20], that uses gradually reducing window lengths instead of a fixed window length, for the trend removal operation. Key steps involved in the *modZFF* method are as follows:

1. Pre-process the input cry signal ($s[n]$) by downsampling to 8 kHz, smoothen it over m sample points to obtain an equivalent effect of low-pass filtering, and then upsample it back to the sampling frequency of the original signal [19]. The resultant pre-processed signal is ($s_p[n]$).
2. Get the differenced signal ($\tilde{x}[n]$) from the pre-processed signal ($s_p[n]$), using:

$$\tilde{x}[n] = s_p[n] - c\, s_p[n-1] \tag{1}$$

   where c is a constant (usually value of c = 0.9 to 1.0 is chosen) and $n = 1, 2, 3, ...$. The differenced signal ($\tilde{x}[n]$) gives a zero-mean signal ($\hat{x}[n]$).
3. Pass the zero-mean signal $\hat{x}[n]$ through a cascade of two zero-frequency resonators (ZFRs), i.e., two ideal digital resonators at 0 Hz, to get the ZFR output signal $\tilde{y_1}[n]$.

$$\tilde{y_1}[n] = \sum_{k=1}^{4} a_k \tilde{y_1}[n-k] + \hat{x}[n], \tag{2}$$

   where, $a_1 = +4$, $a_2 = -6$, $a_3 = +4$, $a_4 = -1$.
4. Remove the trend built-up in the cascaded ZFRs' output ($\tilde{y_1}[n]$) by successive integration operations, i.e., by subtracting the local mean computed over a window. Gradually reducing window lengths of 20 ms, 10 ms, 5 ms, 3 ms, 2 ms and 1 ms are used in the successive trend removal stages, to highlight the excitation source information better. Output of each trend removal stage ($\hat{y_2}[n]$) is given by below equations:

$$\hat{y_2}[n] = \tilde{y_1}[n] - \bar{y}[n] \tag{3}$$

$$\bar{y}[n] = \frac{1}{2N+1} \sum_{n=-N}^{N} \tilde{y_1}[n] \tag{4}$$

   where, $2N + 1$ is window length in number of samples. Here, $\bar{y}[n]$ represents the local mean computed over each successive window. The final trend removed output ($\hat{y_2}[n]$) is called the *modified zero-frequency filtered (modZFF)* signal, i.e. $z_m[n]$ [19].
5. The positive to negative going zero-crossings of the *modZFF* signal ($z_m[n]$) give locations of impulses.
6. The slope of the *modZFF* signal ($z_m[n]$) around each impulse location indicates the relative *strength of excitation (SoE)* there. The *SoE* is denoted as ψ in this paper.

An illustration of SoE impulse sequence and F_0 contour derived from the acoustic signal of infant cry, using the *modZFF method* is shown in Fig. 1(c) and Fig. 1(d), respectively. The *modZFF method* helps deriving an SoE impulse sequence (as the excitation source characteristics) from the acoustic signal of infant cry. The *SoE*, i.e., the amplitudes of impulses, indicate the strength of excitation around the respective impulse locations. Using these excitation source features (F_0 and SoE), derived using the *modZFF method* larger changes are observed in the production features of infant cry due to *pain*, than for *discomfort*. Analysis details are discussed in the next section.

4. Discriminative analysis of Pain vs. Discomfort infant cry sounds

In this paper, the production characteristics of acoustic signals of infants cries data in the *IIIT-S ICSD* are examined under two

Table 2: *Changes in F_0 for pain vs. discomfort cry signals:* (a) speaker #, the (b) mean (μ_{F0_D}) and (c) std. dev. (σ_{F0_D}) for *discomfort* cry, the (d) mean (μ_{F0_P}) and (e) std. dev. (σ_{F0_P}) for *pain* cry, and the changes (Δ %) in (f) mean (μ_{F0}), (g) std. dev. (σ_{F0}) and (h) normalized std. dev. ($\sigma_{N_{F0}}$) in F_0 for *pain* cry from *discomfort* cry signals. Note: M indicates Male and F indicates Female infant.

(a) Speaker (Infant) #	(b) μ_{F0_D} (Hz)	(c) σ_{F0_D} (Hz)	(d) μ_{F0_P} (Hz)	(e) σ_{F0_P} (Hz)	(f) $\Delta\mu_{F_0}$ (%)	(g) $\Delta\sigma_{F_0}$ (%)	(h) $\Delta\sigma_{N_{F_0}}$ (%)
S02 (M)	667.9	184.6	737.9	272.5	10.48	47.60	33.60
S03 (M)	705.9	228.5	620.1	268.3	-12.16	17.42	33.68
S04 (F)	676.9	311.8	754.3	380.0	11.43	21.88	9.38
S05 (M)	654.9	259.9	707.8	306.3	8.10	17.83	9.00
S06 (M)	734.7	232.7	647.9	277.5	-11.81	19.21	35.18
S07 (F)	581.8	219.7	619.8	252.9	6.53	15.12	8.07
S09 (M)	620.7	200.4	664.4	308.8	7.04	54.06	43.94
S10 (F)	667.8	239.8	658.6	283.2	-1.37	18.05	19.69
S11 (F)	579.3	152.2	664.0	211.1	14.63	38.70	20.99
S12 (M)	684.1	224.9	685.4	294.8	0.19	31.09	30.85
S13 (F)	523.6	132.9	530.1	150.4	1.24	13.18	11.79
Average	*645.2*	*217.1*	*662.8*	*273.3*	*3.12*	*27.74*	*23.29*

Table 3: *Changes in SoE (ψ) for pain vs. discomfort cry signals:* (a) speaker #, the (b) mean (μ_{ψ_D}) and (c) std. dev. (σ_{ψ_D}) for *discomfort* cry, the (d) mean (μ_{ψ_P}) and (e) std. dev. (σ_{ψ_P}) for *pain* cry, and the changes (Δ %) in (f) mean (μ_ψ), (g) std. dev. (σ_ψ) and (h) normalized std. dev. (σ_{N_ψ}) in SoE (ψ) for *pain* cry from *discomfort* cry signals. Note: M/F indicates Male/Female infant.

(a) Speaker (Infant)#	(b) μ_{ψ_D}	(c) σ_{ψ_D}	(d) μ_{ψ_P}	(e) σ_{ψ_P}	(f) $\Delta\mu_\psi$ (%)	(g) $\Delta\sigma_\psi$ (%)	(h) $\Delta\sigma_{N_\psi}$ (%)
S02 (M)	.3253	.2132	.2056	.1374	-36.80	-35.55	1.97
S03 (M)	.1613	.1247	.2640	.1776	63.37	42.42	-12.98
S04 (F)	.2454	.1880	.2185	.1556	-10.96	-17.23	-7.04
S05 (M)	.2986	.1860	.2149	.1815	-28.02	-2.42	35.56
S06 (M)	.2782	.1950	.2098	.1674	-24.59	-14.15	13.83
S07 (F)	.2321	.1770	.2955	.2292	27.32	29.49	1.71
S09 (M)	.2305	.1830	.1713	.1528	-25.68	-16.50	12.35
S10 (F)	.1582	.1305	.2262	.1637	42.98	25.44	-12.27
S11 (F)	.2745	.1880	.2212	.1676	-19.42	-10.85	10.63
S12 (M)	.1447	.1307	.1453	.1433	0.41	9.64	9.19
S13 (F)	.3542	.2220	.3127	.2407	11.72	8.42	22.81
Average	*.2457*	*.1762*	*.2259*	*.1743*	*0.03*	*1.70*	*6.89*

categories, namely, *pain* and *discomfort*. Since the data available in other categories is less for same speaker (i.e., same infant), these cry categories may be analysed in future after extending the database. The excitation source features F_0 and SoE are derived using the modZFF method, for each infant cry signal. The signal energy E, a production feature, represents the combined effect of the excitation source and the vocal tract filter. It is computed for frame size of 5 ms at each time-instant.

Relative changes are examined in the production features (F_0, SoE and E) of the acoustic signals of infant cries due to *pain* and *discomfort*. An illustration of energy (E) contour, the SoE impulse sequence and the F_0 contour for *discomfort cry* is shown in Fig. 1(b), (c) and (d), respectively. Changes in these features for *pain cry* are illustrated in Fig. 2, in a similar way. Some patterns can be observed. In Fig. 2, for infant cry due to *pain*, the F_0 contour has near *cyclic* changes with larger fluctuations, that could be due to physiological conditions during pain. In Fig. 1, for infant cry due to *discomfort*, the F_0 contour is relatively *flat*, with changes at larger intervals and shorter fluctuations. Similar changes are observed in the SoE and E also.

Quantitative analysis is carried out by measuring the statistical parameters *mean* (μ), *standard deviation* (σ or std dev) and *normalized standard deviation* ($\sigma_N = \sigma/\mu$) in the production features F_0, SoE and E. Changes in these parameters for *pain* cry vs. *discomfort* cry are compared using the percentage difference, e.g., $\Delta\mu_{F_0} = (\mu_{F_{0_P}} - \mu_{F_{0_D}})/\mu_{F_{0_D}} \times 100(\%)$. Likewise, changes in the features SoE and E are compared using the $\Delta\mu_\psi(\%)$ and $\Delta\mu_E(\%)$, respectively. Changes in fluctuations in the features F_0, SoE and E are given in Table 2, Table 3 and Table 4, respectively, for *pain* vs. *discomfort* cry signals of first 11 infants (6 males, 5 females). Data of two speakers (S01 and S08) is discarded due to overlapping speech present in the signal. Similar changes in the production features are observed for pain/discomfort across speakers in the database.

In Table 2, the *fluctuations* in F_0 ($\Delta\sigma(\%)$ in column (g)) are larger for *pain* cry in comparison to *discomfort* cry. *Normalized fluctuations* ($\Delta\sigma_N(\%)$ in column (h)) are also larger for *pain* cry. In general, the average F_0 values increase for *pain* cry w.r.t. *discomfort* cry ($\Delta\mu_{F_0}(\%)$ in column (f)). But, in few cases (S03, S06 and S10), the F_0 for *discomfort* cry (if high) may reduce for *pain* cry.

In Table 3, the average SoE values ($\Delta\mu_\psi(\%)$ in column (f)) reduce in general for *pain* cry w.r.t. *discomfort* cry. This reduction in the SoE with increase in F_0, is in line with earlier observation of relative changes (in opposite direction) in the F_0 and SoE, for normal and shouted speech of adults [21, 22]. Interestingly, in the cases (S03, S10) where the mean F_0 (μ_{F_0}) decreases for *pain* cry w.r.t. *discomfort* cry, the mean SoE

Table 4: *Changes in signal energy (E) for pain vs. discomfort cry signals:* (a) speaker #, the (b) mean (μ_{E_D}) and (c) std. dev. (σ_{E_D}) for *discomfort* cry, the (d) mean (μ_{E_P}) and (e) std. dev. (σ_{E_P}) for *pain* cry, and the changes (Δ %) in (f) std. dev. (σ_E) and (g) normalized std. dev. (σ_{N_E}) in E for *pain* cry from *discomfort* cry signals. Note: M indicates Male and F indicates Female infant.

(a) Speaker (Infant)#	(b) μ_{E_D}	(c) σ_{E_D}	(d) μ_{E_P}	(e) σ_{E_P}	(f) $\Delta\sigma_E$ (%)	(g) $\Delta\sigma_{N_E}$ (%)
S02 (M)	.2351	.1361	.1904	.1042	-23.44	-5.46
S03 (M)	.1370	.1025	.1625	.0905	-11.71	-25.56
S04 (F)	.1474	.1070	.1525	.0893	-16.54	-19.34
S05 (M)	.1818	.1080	.1386	.0882	-18.33	7.12
S06 (M)	.1756	.1170	.1561	.1037	-11.37	-0.28
S07 (F)	.1530	.0960	.1917	.1049	9.27	-12.77
S09 (M)	.1617	.1120	.1303	.0874	-21.96	-3.14
S10 (F)	.1444	.0930	.1346	.0739	-20.54	-14.75
S11 (F)	.1602	.1120	.1502	.1082	-3.39	3.04
S12 (M)	.0829	.0746	.1073	.0981	31.50	1.60
S13 (F)	.1922	.1020	.1759	.1243	21.86	33.16
Average	*.1610*	*.1055*	*.1536*	*.0975*	*-5.88*	*-3.31*

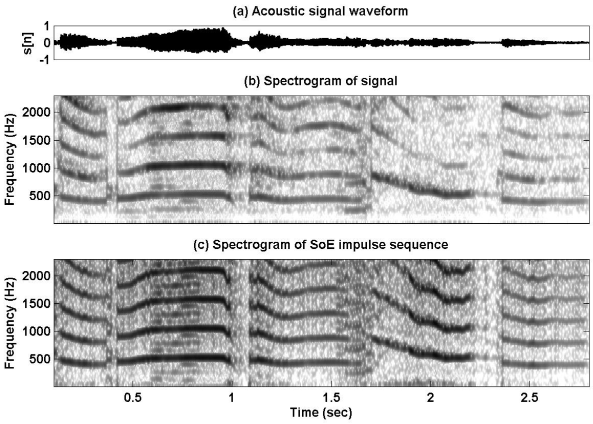

Figure 3: Validation of the excitatation source characteristics of acoustic signal of *discomfort* cry, using spectrograms: (a) acoustic signal, and spectrograms of (b) the acoustic signal and (c) the SoE impulse sequence derived using the modZFF method, for *discomfort* cry signal of infant #S13 (female).

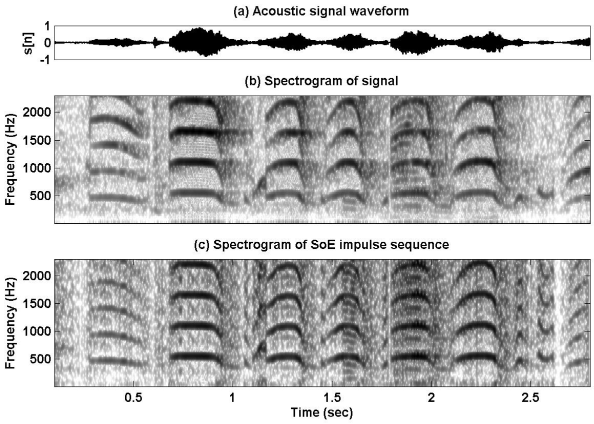

Figure 4: Validation of the excitatation source characteristics of acoustic signal of *pain* cry, using spectrograms: (a) acoustic signal, and spectrograms of (b) the acoustic signal and (c) the SoE impulse sequence derived using the modZFF method, for *pain* cry signal of infant #S13 (female).

(μ_ψ) increases. The fluctuations in the SoE ($|\Delta\sigma_\psi|$(%) in column (g)) and normalized fluctuations ($|\Delta\sigma_{N_\psi}|$(%) in column (h)) are larger for *pain* cry w.r.t. *discomfort* cry of most infants. Similar trend of changes in the fluctuations in signal energy (E) are observed in Table 4, in columns (f) and (g).

5. Discussion on results

Changes in the excitation source feature SoE (in Table 3) appear to be more prominent than changes in the signal energy E (in Table 4), as also observed in Fig. 2. Prominence of changes in the source characteristics in comparison to acoustic signal is validated by visual inspection of the spectrograms ($|X(\tau,\omega)|^2$), used as the ground truth. These spectrograms are obtained using the *short-time Fourier transform* $X(\tau,\omega) = \sum_{n=-\infty}^{n=\infty} x[n]\, w[n-m]e^{-jwn}$ [23, 24], for signal frames of size 20 ms, with a frame shift of 1 ms. Spectrograms for the SoE impulse sequences (Fig. 3(c) and Fig. 4(c)), reveal the excitation source characteristics better than the spectrograms of the acoustic signals (Fig. 3(b) and Fig. 4(b)). These spectrograms also indicate the nature of inter-cry changes in the excitation source features, that are similar to those illustrated in Fig. 1(d) and Fig. 2(d).

For *pain* cry signal, the contours of F_0 and harmonics have *cyclic* changes with larger fluctuations (region 1.2 sec to 2.4 sec in Fig. 4). But, for cries due to *discomfort* these contours are relatively *flat* with fewer fluctuations (region 1.1 sec to 2.3 sec in Fig. 3). The cyclic changes with larger fluctuations for *pain* cry are possibly due to significant changes in the excitation source characteristics during the production of cry signal in shorter and louder bursts. Whereas, relatively flat contours with lesser fluctuations for *discomfort* cry may be due to slow moaning, with related smaller changes in the excitation source characteristics.

6. Summary and conclusion

In this paper, the production characteristics are examined for acoustic signals of the infant cries due to *pain* and *discomfort*. Aim is to characterise the infant cry signal and identify features that help distinguishing the causes of infant cries. Acoustic signals of infant cries due to *pain* or *discomfort* are examined using three production features F_0, SoE and signal energy. The excitation source features F_0 and SoE are derived using the *modified zero-frequency filtering* method. The F_0 contour for *pain* cry has *cyclic* changes with larger fluctuations, but it is relatively *flat* with lesser fluctuations for *discomfort* cry. This observation is validated quantitatively by comparing the relative fluctuations and normalized fluctuations in F_0 and SoE for *pain vs. discomfort* cry. Significance of changes in the source characteristics is validated using spectrograms of the SoE impulse sequence and the acoustic signal. The results are consistent across cry signals of speakers (infants) in the database.

In future, changes may be examined in the production characteristics of acoustic signals of infant cries due to remaining categories other than pain and discomfort. This author is working towards developing the systems where automated detection of cause of infant cry may help the parents as well the doctors towards assisted clinical diagnosis of an infant's ailment. Details of these attempts may be expected to appear in future publications of the author.

However, this study highlights the importance of examining changes in the excitation source characteristics of acoustic signals of the paralinguistic sounds such as infant cry. The study should also be helpful towards developing the assistive technologies and systems that may help early diagnosis of the ailment and medical care to an infant by identifying the cause of cry, from the acoustic signal. It could be immensely useful in the cases of ailments where the reaction-time of ailment detection and remedial measures could be of critical importance for an infant's life.

7. Acknowledgements

The author would like to thank paediatricians Dr. Manish Gour (M.B.B.S, DCH) and Dr. Nizam (M.B.B.S) from Pranaam Hospital, Hyderabad for their help in collecting the infant cry data. The author is also grateful to all parents of the infants, for willingly giving the details about themselves and their infants, and cooperate in the data collection. The author would also like to thank the anonymous reviewers for their insightful suggestions.

8. References

[1] A. Neustein, *Advances in speech recognition: mobile environments, call centers and clinics*. Springer, 2010, ch. fourteenth, pp. 324–345.

[2] G. Varallyay Jr, Z. Benyo, A. Illenyi, Z. Farkas, and G. Katona, "The speech-chorus method at the analysis of the infant cry," *Acoustic Review*, vol. 6, no. 2, pp. 9–15, 2005.

[3] Y. Skogsdal, M. Eriksson, and J. Schollin, "Analgesia in newborns given oral glucose," *ACTA Paediatrica*, vol. 86, no. 2, pp. 217–220, 1997.

[4] H. E. Baeck and M. N. Souza, "Study of acoustic features of newborn cries that correlate with the context," in *Proc. 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 3, 2001, pp. 2174–2177.

[5] R. P. Daga and A. M. Panditrao, "Acoustical Analysis of Pain Cries in Neonates: Fundamental Frequency," *IJCA Special Issue on Electronics, Information and Communication Engineering (ICEICE)*, vol. 3, pp. 18–21, Dec. 2011.

[6] Y. Abdulaziz and S. M. Syed Ahmad, "Infant cry recognition system: A comparison of system performance based on mel frequency and linear prediction cepstral coefficients," in *Proc. IEEE International Conference on Information Retrieval & Knowledge Management (CAMP)*, 2010, pp. 260–263.

[7] R. Hidayati, I. K. E. Purnama, and M. H. Purnomo, "The extraction of acoustic features of infant cry for emotion detection based on pitch and formants," in *Proc. IEEE International Conference on Instrumentation, Communications, Information Technology, and Biomedical Engineering (ICICI-BME)*, 2009, pp. 1–5.

[8] K. Kuo, "Feature extraction and recognition of infant cries," in *Proc. IEEE International Conference on Electro/Information Technology (EIT 2010)*, 2010, pp. 1–5.

[9] Y. Mima and K. Arakawa, "Cause Estimation of Younger Babies' Cries from the Frequency Analyses of the Voice-Classification of Hunger, Sleepiness, and Discomfort," in *Proc. IEEE International Symposium on Intelligent Signal Processing and Communications (ISPACS'06)*, 2006, pp. 29–32.

[10] P. S. Zeskind and T. R. Marshall, "The relation between variations in pitch and maternal perceptions of infant crying," *Child Development*, pp. 193–196, 1988.

[11] M. Petroni, M. E. Malowany, C. C. Johnston, and B. J. Stevens, "A new, robust vocal fundamental frequency (F_0) determination method for the analysis of infant cries," in *Proc. IEEE Seventh Symposium on Computer-Based Medical Systems*, 1994, pp. 223–228.

[12] G. Várallyay, "Future prospects of the application of the infant cry in the medicine," *Electrical Engineering*, vol. 50, no. 1-2, pp. 47–62, 2006.

[13] S. Chandralingam, T. Anjaneyulu, and K. Satyanarayana, "Estimation of Fundamental and Formant frequencies of infants cries; a study of Infants with congenital Heart disorder," *Indian Journal of Computer Science and Engineering*, vol. 3, no. 4, pp. 574–582, 2012.

[14] G. Várallyay Jr., "The melody of crying," *Elsevier International Journal of Pediatric Otorhinolaryngology*, vol. 71, no. 11, pp. 1699–1708, 2007.

[15] S. Asthana, N. Varma, and V. K. Mittal, "Preliminary Analysis of Causes of Infant Cry," in *Proc. IEEE 14th International Symposium on Signal Processing and Information Technology (ISSPIT 2014)*, Dec. 15-17 2014.

[16] S. Sharma, S. Asthana, and V. K. Mittal, "A Database of Infant Cry Sounds to Study the Likely Cause of Cry," in *Proc. ACL Anthology 12th International Conference on Natural Language Processing (ICON-2015)*, IIITM-K, Trivendram, Dec. 11-14 2015.

[17] K. S. R. Murty and B. Yegnanarayana, "Epoch extraction from speech signals," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, no. 8, pp. 1602–1613, 2008.

[18] B. Yegnanarayana and K. S. R. Murty, "Event-based instantaneous fundamental frequency estimation from speech signals," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 17, no. 4, pp. 614–624, 2009.

[19] V. K. Mittal and B. Yegnanarayana, "Study of characteristics of aperiodicity in Noh voices," *The Journal of the Acoustical Society of America*, vol. 137, no. 6, pp. 3411–3421, 2015.

[20] V. K.Mittal and B. Yegnanarayana, "Analysis of production characteristics of laughter," *Elsevier Computer Speech & Language*, vol. 30, no. 1, pp. 99–115, 2015.

[21] V. K. Mittal and B. Yegnanarayana, "Production features for detection of shouted speech," in *Proc. 10th Annual IEEE Consumer Communications and Networking Conference, 2013 (CCNC'13)*, Jan. 11-14, 2013, pp. 106–111.

[22] V. K. Mittal and B. Yegnanarayana, "An automatic shout detection system using speech production features," in *Proc. Second International Workshop on Multimodal Analyses enabling Artificial Agents in Human-Machine Interaction (MA3HMI 2014), Held in Conjunction with INTERSPEECH 2014*, Singapore, Sep. 2014, pp. 88–98, published as Lecture Notes in Artificial Intelligence (LNAI 8757) by Springer.

[23] A. V. Oppenheim, R. W. Schafer, J. R. Buck *et al.*, *Discrete-time signal processing*. Prentice-hall Englewood Cliffs, 1989, vol. 2, ch. third.

[24] L. Rabiner, B. H. Juang, and B. Yegnanarayana, *Fundamentals of Speech Recognition*, Indian subcontinent adaptation, first ed. New Delhi, India: Pearson Education Inc., 2009, ch. third, pp. 88–113.