Infant Cry Language Analysis and Recognition: An Experimental Approach

Lichuan Liu, *Senior Member, IEEE*, Wei Li, *Senior Member, IEEE*, Xianwen Wu, *Member, IEEE*, and Benjamin X. Zhou

***Abstract*—Recently, lots of research has been directed towards natural language processing. However, the baby's cry, which serves as the primary means of communication for infants, has not yet been extensively explored, because it is not a language that can be easily understood. Since cry signals carry information about a babies' wellbeing and can be understood by experienced parents and experts to an extent, recognition and analysis of an infant's cry is not only possible, but also has profound medical and societal applications. In this paper, we obtain and analyze audio features of infant cry signals in time and frequency domains. Based on the related features, we can classify given cry signals to specific cry meanings for cry language recognition. Features extracted from audio feature space include linear predictive coding (LPC), linear predictive cepstral coefficients (LPCC), Bark frequency cepstral coefficients (BFCC), and Mel frequency cepstral coefficients (MFCC). Compressed sensing technique was used for classification and practical data were used to design and verify the proposed approaches. Experiments show that the proposed infant cry recognition approaches offer accurate and promising results.**

***Index Terms*—Infant cry signal, feature extraction, language recognition, compressed sensing.**

I. Introduction

CRYING is the primary means of communication for infants. Experts, including experienced parents, pediatricians and child care specialists, can often distinguish infant cries though training and experience [1]. However, it is difficult for new parents and inexperienced pediatricians and caregivers to interpret infant cries. Hence, differentiating cries with various meanings based on related cry audio features is of great importance [1], [2].

Prior works on infant cry analysis have either investigated the difference between normal and pathological (deaf or hearing disabled infants) cries, or they have attempted to differentiate conditional cries such as pain from immunization shots, fear from jack-in-the box toys, or frustration from head restraints [3]. Previously, in [4], [5], we proposed a preliminary approach which can recognize cry signals of a specific infant. However, only limited normal cry signals such as hunger, a wet diaper and attention have been studied, and the algorithms work only for specific infants in the study in a controlled lab environment. Nevertheless, an abnormal cry can be associated with severe or chronic illness, so the detection and recognition of abnormal cry signals are of great importance. Compared with normal cry signals, abnormal cry signals are more intense, requiring further evaluation [6]. An abnormal cry is often related to medical problems, such as: infection, abnormal central nervous system, pneumonia, sepsis, laryngitis, pain, hypothyroidism, trauma to the hypopharynx, vocal cord paralysis, etc. Therefore, approaches which can identify and recognize both normal and abnormal cry signals in practical scenarios is of extreme importance. In this paper, we propose a novel cry language recognition algorithm which can distinguish the meanings of both normal and abnormal cry signals in a noisy environment. Additionally, the proposed algorithm is individual crier independent. Hence, this algorithm can be widely used in practical scenarios to recognize and classify various cry features.

The proposed algorithm can be used to interpret a babies' needs, providing parents with an appropriate way to sooth infants. Furthermore, it can help parents or infant caregivers avoid misunderstanding of their babies' cries thereby reducing their own stress. It also helps prevent child abuse and neglect. Moreover, analyzing infant cries provides a non-invasive diagnostic of the condition of the infant without using invasive tests [7]. Using an infant's cry as a diagnostic tool plays an important role in various situations: tackling medical problems in which there is currently no diagnostic tool available (e.g. sudden infant death syndrome (SIDS), problems in developmental outcome and colic), tackling medical problems in which early detection is possible only by invasive procedures (e.g. chromosomal abnormalities), and finally tackling medical problems which may be readily identified but would benefit from an improved ability to define prognosis, (e.g. prognosis of long term developmental outcome in cases of prematurity and drug exposure [8]).

In our model, cry signals are output signals from the vocal tract system, which is also called the linear system. The stimuli signal, which excites the linear system, is the airflow from an infants' lungs [9]. Similar to digital speech signal processing, we use a time-varying Fourier transform to study the spectral properties of cry signals. Therefore, we can identify

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. This work was supported by the Gerber Foundation and the Northern Illinois University Research Foundation. Recommended by Associate Editor Mengchu Zhou. *(Corresponding author: Lichuan Liu.)*

Citation: L. C. Liu, W. Li, X. W. Wu, X. Zhou, "Infant cry language analysis and recognition: an experimental approach," *IEEE/CAA J. Autom. Sinica*, pp. 1−11, 2019. DOI: 10.1109/JAS.2019.1911435

L. Liu, W. Li and X. Wu are with Department of Electrical Engineering, Northern Illinois University, DeKalb, IL 60115 USA (e-mails: {liu, weili, z1648342}@niu.edu).

Zhou is with the Department of Biology, The College of New Jersey, Ewing, NJ 08618 USA (email: zhoub1@tcnj.edu)

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JAS.2019.1911435

the difference between vocal tract systems and input signals, which are related with different cry reasons. In this paper, short-time Fourier transform (STFT) is used to analyze the cry signals. Recently, speech recognition and acoustic signal classification techniques have been widely used in many areas such as manufacturing, communication, consumer electronic products and medical care [10]−[12]. Speech recognition is a signal processing procedure that transfers speech signal waveforms in a spatial domain into a series of coefficients, called a feature, which can be recognized by the computer [10]−[13]. Since infant cry signals are time-varying non-stationary random signals which are similar to speech signals. The stimuli for infant cry signal is the same as the stimuli for voiced speech signal. In this paper, we use techniques originally designed and used in automatic speech recognition to detect and recognize the features for infant cry signals, and use compressed sensing to analyze and classify those signals. Fig. 1 shows the procedures of cry signal recognition which consists of the following steps:

Step 1. Cry unit detection

Step 2. Feature extraction

Step 3. Analysis and classification

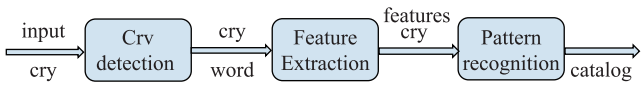


Fig. 1. Block Diagram for Infant cry Recognition.

This paper is organized as follows. Section II introduces the anatomy of infant-related cries and the physiology of cry signals. In section III, short time Fourier analysis is proposed and cry detection techniques are presented. Section IV presents feature pattern extraction algorithms, and proposes a compressed sensing model to recognize and classify infant cry signals. Experimental results are presented in section V. Finally, in section VI, we conclude the paper.

II. Infant Cry Modelling and Categorization

A. Physiology of Infant Cry

From a physiological point of view, increasing alertness and decreasing crying, as part of the sleep/wakefulness cycle, suggests that there may be a balanced exchange between crying and attention. The change from sleep/cry to sleep/alert/cry necessitates the development of control mechanisms to modulate arousal. An infant has to increase arousal gradually to maintain states of attention for longer periods.

The infant cry is the result of complex interactions between anatomic structures and physiologic mechanisms. These interactions involve the central nervous system, the respiratory system, the peripheral nervous system, and a variety of muscles [14].

Newborns differ from one another in their response to different stimuli. There are two main physiological states which infants can switch between: a sleep state and an awake state. Within the sleep state, infants fall under two categories-either the quiet sleep or the active sleep category. On the other hand, the awake state is characterized by four main behaviors: drowsy, quiet alert, active alert, and crying. Physiological changes can easily affect an infant's cry behavior directly. In the first few weeks after birth, crying has a reflexive-like quality and is most likely tied to the regulation of physiological homeostasis as the neonate is balancing internal demands with external demands [15].

As physiological processes stabilize, periods of alertness and attention increase, which place additional demands on regulatory functions. Crying can occur when the system becomes overloaded due to external stimulation. Crying is also considered as a mechanism for discharging energy or tension. The need for tension reduction is especially acute at times of major developmental upheavals and shifts. Unexplained fussiness and sudden increases in crying occur between 3 and 12 weeks of age due to maturational changes in brain structure and shifts in the organization of the central nervous system. Physiological and anatomical changes that occur around 1 to 2 months result in more control over vocalization, thus crying becomes more differentiated. At the age of 7−9 months there is a second bio-behavioral shift characterized by major cognitive and affective changes that are also thought to reflect central nervous system reorganization. Crying now occurs for additional reasons, such as fear and frustration [15].

B. Catalog of Cry Signal

The cry production mechanism in infants resembles the speech production process in adults [9]. First, external or internal stimuli will stimulate the infant's brain. Then, the nervous system will transmit the brain's commands to speech and respiratory muscles which control the ejection of air from the lungs to the vocal tract, changing the vocal tract status [14]. As a result, a different acoustic sound is uttered. The vibration of vocal cords and muscle movements results in a change in air pressure. The cord vibration fundamental frequency is called the pitch.

Similar to speech signals, infant sounds can also be defined as voiced or unvoiced excitations based on different utterance mechanisms. Voiced excitations occur in the larynx and involve vocal cord vibration while unvoiced excitations involve air turbulence of occlusion caused by the soft palate, tongue, teeth, or lips.

Crying serves several useful purposes for infants. Crying is a way for infants to communicate when they are hungry or uncomfortable [15]. Crying helps them shut out intensive stimuli, such as: sights, sounds, and other sensations. Additionally, it helps infants release tension. Sometimes, crying even helps babies get rid of excess energy. Normal cries can be due to hunger, a need for a diaper change or a need to be held. However, there are also cries associated with something more severe (abnormal cries), such as a hair tourniquet (a piece of hair wrapped very tightly around a finger or toe), an obstruction in the intestine, or pain and sickness. Understanding and identifying the different reasons for various infant cries, especially the abnormal cries can help parents or caregivers choose the proper healthcare service and reduce the risk of health impairment for infants. The following are some common reasons for infant crying [16], hunger, stomach

problems, needing to sleep, a dirty diaper, wanting to be hold and etc.

III. Cry Signal Time Frequency Analysis and Detection

After obtaining cry signals, we analyze the recorded signals by using waveform and time frequency analysis. Then we conduct signal detection and segmentation for later pattern extraction. Signal detection processes instances of voiced activity instead of spending computational time during silent periods. To accurately detect potential periods of voiced activity, two short term signal detection techniques are used.

A. Short-Time Fourier Analysis

In this section, we use time frequency analysis to analyze the infant cry signals. It is well known that Discrete Fourier Transform (DFT) of a long sequence is an estimate of the power spectrum density (PSD), called a periodogram [11]. Different cry signals from different infants would produce similar gross PSD. Therefore, we use STFT to obtain the time-varying properties of cry signals. STFT is defined as:

$$X_n(e^{j\omega}) = \sum_{m=-\infty}^{\infty} x(m)w(n-m)e^{-j\omega n} \tag{1}$$

where $w(n-m)$ is a real window sequence to determine the portion of the signal $x(n)$ that receives emphasis at a particular time index, n. STFT is a time dependent complex function of time index n and frequency ω.

We can observe STFT as the discrete time Fourier transform (DTFT) of the sequence $x(m)w(n-m)$. An alternative interpretation of STFT is to consider $X_n(e^{j\omega})$ as a function of n with a given frequency. Then it becomes a discrete-time convolution and can be considered as linear filtering.

The shape of the window sequence has an important effect on this time-dependent FT. The STFT of a given signal is

$$X_n(e^{j\omega}) = \frac{1}{2\pi}\int_{-\pi}^{\pi} W(e^{-j\omega})e^{-j\omega n}X(e^{j(\omega-\widetilde{\omega})}d\widetilde{\omega} \tag{2}$$

Fourier transform (FT) of the sequence of input signal is convolved with the FT of the shifted window. To represent $X(e^{j\omega})$ by using STFT $X_n(e^{j\omega})$, we choose a window function with a spectral highly concentrated around the origin. In this paper, Hamming window is used to conduct STFT.

B. Short-Time Energy

Short-time energy (STE) is defined as the average of the square of the sample values in a suitable time window [10]:

$$E(n) = \frac{1}{N}\sum_{m=0}^{N-1}[w(m)x(n-m)]^2, \tag{3}$$

where $w(m)$ are the coefficients of the window function of length N, m stands for window index, and n stands for index of sample. The Hamming window was defined as a time window which minimizes the maximum side lobe in frequency domain.

Short-time processing of crying should take place during segments between 10-30ms in length [9]. For signals of 8kHz sampling frequency, a time window of 128 samples (16ms) was used. As shown in Fig. 7, STE estimation performs well as a cry detector because there is a noticeable difference of average energy between voiced and unvoiced cry signals, and between crying and silence. This technique is usually paired with short-time zero crossing for a robust detection scheme.

C. Short-Time Zero Crossing

Short-time zero crossing (STZC) is defined as the rate of signal sign change [11]:

$$Z(n) = \frac{1}{N}\sum_{m=0}^{N-1}|sing(x(n-m)) - sign(x(n-m-1))|, \tag{4}$$

where $\text{sign}(x(m)) = \begin{cases} 1 & x(m) \geq 0 \\ -1 & x(m) < 0 \end{cases}$.

STZC estimation works well with the crying detector because there are noticeably fewer zero crossings in voiced crying as compared with unvoiced crying. It is obvious that short-time zero crossing can predict the start and endpoints of cry signals, as shown in Fig. 8. STZC approach can effectively obtain the envelope of a non-silent signal, and combined with short-time energy, STZC can effectively track instances of potentially voiced signals that are the signals of interest for analysis.

Not all signals bounded by the STZC boundary contain cries. Large STZC envelopes with low energy tended to contain cry precursors such as whimpers and breathing events. Not all signals with non-negligible STE contained cries as well. Infant coughing could also have similar STZC envelopes and contain noticeable STE values. In this research, crying is defined as a high energy segment of sufficiently long duration. In this research, we use both STE and STZC to detect cry units. As the normal infant cry duration is around 1.6sec, the two quantifiable threshold conditions to constitute a desired voiced cry are [14]:

1) Normalized energy $>$ 0.05 (to eliminate non-voiced artifacts such as breathing/whimpering and to supersede cry precursors);

2) Signal envelope period $>$ 0.1 seconds (to eliminate impulsive voiced artifacts such as coughing).

IV. Feature Extraction and Recognition

A. Audio Features

Through the sense of hearing, people can distinguish similar sounds of different types. This is done through the human perception of qualitative audio features. There are four primary auditory qualities associated with sound: loudness, pitch, timbre, and the source of the sound [11]. Loudness is a quantitative measure of the amplitude of the sound compared to a reference level and can be qualitatively described from being quiet to loud. Pitch is a quantitative measure of the actual fundamental frequency of a signal and can be qualitatively described from low to high. Timbre is a qualitative measure

of a sound that can be used to help differentiate between two sounds of equal loudness and pitch through the tonal quality of the sound. Essentially whenever a sound is heard, the human brain will actively process those analog auditory qualities and make decisions regarding the sound.

Audio feature extractions hinge upon digital signal processing of audio signals to quantize acoustic information in a manner that makes classification practical and tractable. The comparison of time domain waveforms can be used as a measure for signal classification. On the other hand, time-domain signals can also be segmented and processed in smaller time windows to generate frequency domain snapshots of the segments through Fourier transform. The frequency domain analysis of signals yields information closely tied with timbre and pitch. In this paper, we leverage both time and frequency domain analysis to cover all four primary auditory qualities.

In this section, Linear predictive coding (LPC), linear predictive code cepstral (LPCC), Mel-frequency cepstral coefficients (MFCC), and Bark-frequency cepstral coefficients (BFCC) are extracted from cry signals as features. Additionally, Compressed Sensing (CS) is used for cry feature recognition in this paper.

B. Linear Predictive Coding

The waveforms of two similar sounds will also be similar. If two infant cries have very similar waveforms, it indicates that they should possess the same impetus. However, it is impractical to conduct a full sample by sample comparison between cry signals due to the complexity of the sampled audio signals. For better performance of the time domain comparison of infant cry signals, linear predictive coding (LPC) is applied.

There are two acoustic sources associated with voiced and unvoiced speech. Voiced crying is produced by the vibration of the vocal cords caused by the airflow from the lungs and this vibration is periodic in nature; unvoiced crying is produced by constrictions in the air tract resulting in random airflow [12]. The basis of the source-filter model of speech is that crying can be synthesized by generating an acoustic source and passing it through an all-pole filter.

LPC produces a vector of coefficients that represent a spectral shaping filter [11]. The input signal to this filter is either a pitch train for voiced sounds, or white noise for unvoiced sounds. This shaping filter is an all-pole filter represented as [11]:

$$H(z) = \frac{1}{1 - \sum_{i=1}^{M} a_i z^{-i}} \tag{5}$$

where a_i are the linear prediction coefficients and M is the number of poles. The present sample of the cry signal could then be described as a linear combination of the past M samples of the cry signals.

The coefficients $\{a_i\}$ can then be estimated by either autocorrelation or covariance methods [10]. Effectively, the purpose of LPC is to take a large size waveform and then compress it into coefficients, a more manageable form. Because similar waveforms will also result in similar acoustic output, LPC serves as a time domain measure of how close two different waveforms are.

Linear Predictive Cepstral Coefficients (LPCC) represents LPC coefficients in the cepstral domain [12]. This feature reflects the difference of the biological structure of the human vocal track [9]. LPCC derives from LPC recursively as [11]

$$\begin{cases} LPCC_1 = LPC_1 \\ LPCC_i = LPC_i + \sum_{k=1}^{i-1} \frac{k}{i} LPCC_{i-k} LPC_k, 1 < i \leq M \end{cases} \tag{6}$$

where M is LPCC coefficients order, $i = 2, \ldots, M$.

C. Mel Frequency Cepstral Coefficients

Mel frequency cepstral coefficients (MFCC) are coefficients that describe the mel frequency cepstrum [13], [18]. In sound processing, mel frequency cepstrum is a representation of the short-time power spectrum of a sound based on a linear cosine transform of a log spectrum on a non-linear mel scale of frequency. The mel frequency cepstrum is obtained with the following steps. The short-time Fourier transform of the signal is taken to obtain the quasi-stationary short-time power spectrum $F(f) = F\{f(t)\}$. The frequency portion is then mapped to the mel scale perceptual filter bank with 18 triangle band pass filters equally spaced on the mel range of frequency $F(m)$. These triangle band pass filters smooth the magnitude spectrum such that the harmonics are flattened to obtain the envelope of the spectrum The log of the filtered spectrum is obtained and then the Fourier transform of the log spectrum squared results in the power cepstrum of the signals.

$$Mel(f) = 2595 log_{10}(1 + \frac{f}{700}) \tag{7}$$

At this point, the discrete cosine transform (DCT) of the power cepstrum is taken to obtain the MFCC, a tool commonly used to measure audio signal similarity. The DCT coefficients are retained as they represent the power amplitudes of the mel frequency cepstrum.

D. Bark Frequency Cepstral Coefficients

Similar to MFCC, BFCC warps power cepstrum in such a way that it matches human perception of loudness. The method of obtaining BFCC is similar to that of MFCC [12]. In BFCC, frequencies are converted to the bark scale as following:

$$Bark(f) = 13\text{arctan}(0.00076f) + 3.5\text{arctan}((\frac{f}{7500})^2) \tag{8}$$

where $Bark$ denotes bark frequency and f is the frequency in Hertz. The mapped bark frequency is passed through 18 triangle band pass filters. The center frequencies of these triangular band pass filters correspond to the first 18 of the 24 critical frequency bands of hearing.

BFCC is obtained by applying DCT to the bark frequency cepstrum and the 10 DCT coefficients describe the amplitudes of the cepstrum. The power cepstrum also possesses the same sampling rate as the signal, so the BFCC is obtained by performing LPC algorithm on the power cepstrum in 128 sample frames. BFCC encodes the cepstrum waveform in a compact fashion that makes it suitable for classification schemes.

E. Classification with Compressed Sensing

In this section, we propose a solution to cry recognition by obtaining the sparse representation of the training cry signals through compressed sensing [19]. Cry signals could be represented by a vertical vector and we organize all the training cry data into matrix **A**. Test cry files can be contained as vector **y**. The test data sets can be linearly combined by the training data set. The solution vector of this problem will be sparse.

To obtain the sparse solution of the linear system

$$\mathbf{Ax} = \mathbf{y}, \tag{9}$$

where **A** is a $m \times n$ matrix constructed by the training data vectors, **y** is the test cry signal vector, and **x** is the solution of the linear system. The total number of classes will be less than n. If $m \gg n$, the system is called an overdetermined system. Typically, one can use the Randomized Kaczmarz algorithm for feature recognition [19] under noise free situation. The algorithm process is as follows:

Input: Standardized matrix **A**, and standardized vector **y**.

Output: Estimation vector **x** of the linear system **Ax=y**.

Set x_0 with random number, for k=0.

1. Randomly select $r \in \{1, 2, \ldots, m\}$, update the vector

$$\mathbf{x}_{k+1} = \mathbf{x}_k + (y_r - < \mathbf{A}_r, \mathbf{x}_k >)\mathbf{A}_r \tag{10}$$

Repeat the iteration until the solution is obtained.

Since the ambient noise level at a hospital is high, both training data and test data we obtained from the collaborating hospital were polluted by noise. Therefore, we used a modified Kaczmarz algorithm to improve the performance with the noisy background [20]. In the modified Kaczmarz algorithm, training cry signal matrix **A** was constructed with normalized columns. Matrix **A** is a $m \times n$ matrix. The input test cry feature vectors were also normalized. Training data with various iteration cycles were chosen to verify the performance of the modified Kaczmarz algorithm [20]. The procedure is as follows:

Input: Standardized matrix **A**, k is the index number and standardized vector **y**.

Output: Estimation vector **x** of the linear system **Ax=y**.

Set $\mathbf{x}_0$ with random number, for k=0.

1. Randomly select $r \in \{1, 2, \ldots, m\}$,

2. Set $\mu_k = \langle \mathbf{A}_r, \mathbf{A}_s \rangle$,

$$\mathbf{h_k} = \mathbf{x}_{k-1} + (y_r - \langle \mathbf{x}_{k-1}, \mathbf{A}_s \rangle)\mathbf{A}_s \tag{11}$$

$$v_k = \frac{(\mathbf{A}_r - \mu_k \mathbf{A}_s)}{\sqrt{1 - |\mu_k|^2}} \tag{12}$$

$$\beta_k = \frac{(y_r - \mu_k y_s)}{\sqrt{1 - |\mu_k|^2}} \tag{13}$$

3. update the vector $x_{k+1} = h_k + (\beta_k - \langle h_k, v_k \rangle)v_k$

Repeat the procedure $1-3$ until the solution is obtained. This algorithm convergence with expected exponential rate.

V. Experiments and Results

All the baby cry audio data studied in this paper were recorded in the neonatal intensive care unit (NICU) of a hospital. The probable reason for each cry signal file is given by experienced neonatal nurses as known facts. A Shure omnidirectional SM94 microphone was used to collect infant cry signals. When the baby was crying, we placed the microphone around 6-10 inches away from the infant's mouth to pick up the cry audio signal. A Sound Digital 722 digital audio recorder was used to record infant cry signals. The sampling frequency was 44.1 kHz with a resolution of 16 bits, and then down sampled to 7350Hz.

The probable reason for each cry signal file was given by experienced neonatal nurses, experienced nurses and caregivers who were able to identify the reason for a baby's cries after a bit of listening. For example, there are some observed types of newborn cries associated with different audio cues:

The "neh" sound is generally related to being "hungry". Typically, when a baby has the sucking reflex, and his/her tongue is pushed to the roof of the mouth, a "neh" sound is generated.

The "owh" sound is made in the reflex of a yawn which means "sleepy".

The "heh" sound means 'I need something', such as: being too cold, being itchy, needing a new diaper, or needing a new body position, etc.

The "eair" is a deeper sound which comes from the abdomen, so it means lower gas pain. It is usually accompanied by a newborn pulling his/her knees up or pushing down his/her legs.

The "eh" sound means that a baby needs to burp. Generally speaking, it happens after feeding.

Besides listening to cry signals, experienced personnel can confirm the reasons for different cries by considering other cues, such as gesture, facial expressions, and motion. For example, some hunger signs for newborns include fussing, lip smacking, rooting (a newborn reflex that makes babies turn their head toward your hand when you stroke their cheek), and putting their fingers to their mouth [21]. Crying caused by wet diapers can be distinguished by just checking infant's diaper. The signs for being "sleepy" are yawning, rubbing eyes and nodding. Attention crying can be easily soothed by holding infants or interacting with them. Discomfort crying, such as an injection or blood test, could be associated with a certain medical procedure.

All the babies have their own nursing logs containing information including: age, sex, temperature, blood pressure, feeding time, diaper change time, sleep time and so on. Nurses can use the data provided as well as deductive logic to then interpret an infant's cry. For instance, if a baby was fed a few minutes ago, then their crying is most likely not due to hunger and an infant who just woke up usually does not cry because they are sleepy.

There were 48 cries obtained from 26 infants, of which 11 were female and 15 were male. Among the samples, there were 25 Asian babies and 1 Caucasian baby. The age of these infants ranged from 3 days to 6 months. None of the

infant has hearing impairment. Cry signals were filed under five different causes: needing a diaper (6 observations), being hungry (16 observations), needing attention (8 observations), needing sleep (8 observations), and being in discomfort (10 observations), which included injection, sputum induction and blood tests. In the 20 40 second recording time, we assumed that an infant would not change his/her mood or desire within the recording period.

Based on the data obtained from the babies and known facts from the experts, we listed 'discomfort' cry in the abnormal category of crying and the other cries in the normal category. We used 'hungry', 'diaper', 'attention', 'sleepy' and 'discomfort' as pilot features, but more features can be easily added such as tired, cold or hot, need a burp etc..

Signal acquisition and numbering of cry audio files with the associated infant is shown in Table I. In the Age column of the table, d stands for day, w stands for week, and m stands for month.

TABLE I
CRY SIGNAL INFORMATION

	Cause	Sex	Age	Race	File
1	Diaper	F	2w	Asian	T07
2	Attention	F		Asian	T10A
3	Attention				T34
4	Attention				T105
5	Hungry	M	1w	Asian	T11
6	Attention				T33
7	Hungry				T35
8	Hungry	M	1w	Asian	T19
9	Sleepy	M	3m	Asian	T20
10	Disturbed				T32
11	Sleepy	F		Asian	T21
12	Sleepy				T23
13	Diaper	F	3d	Asian	T22
14	Inject	M	1w	Asian	T24
15	Sputum induction	M	2w		T110
16	Sleepy	M	1w	Asian	T25
17	Hungry	M	2w		T113
18	Hungry	F	3d	Asian	T26
19	Attention	F	1w		T104
20	Hungry	F	1w		T122
21	Attention	F	8d	Asian	T27
22	Uncomfortable	M	2w	Asian	T28
23	Blood test	M	3w	Asian	T109
24	Diaper	F	2w	Asian	T29
25	Attention	M	9d	Asian Asian	T30
26	Attention				T31
27	Diaper	M	2d	Asian	T36
28	Diaper	M	6d		T116
29	Diaper	F	9d	Asian	T37
30	Hungry	F	2w		T117
31	Other	M	1w	Asian	T106
32	Hungry	M	2w		T121
33	Hungry	M	2w	Asian	T107
34	Blood test	F	2w	Asian	T108
35	Uncomfortable	F	1m	Asian	T111
36	Hungry	F			T124
37	Hungry	M	5d	Asian	T112
38	Hungry	M	11d		T114
39	Hungry	M	2w	Asian	T115
40	Hungry	M	2w		T120
41	Sleepy	F	8d	Asian	T118
42	Sleepy	F			T119
43	Hungry	M	1w	Caucasian	T123
44	Uncomfortable	M	14w	Asian	T125
45	Sleepy	M			T126
46	Hungry	M			T127
47	Sleepy	M			T128
48	Uncomfortable	M			T129

We analyzed the different cry signals by using time-frequency analysis. A Hamming window with length 256 was used, the overlap was 128 and a 512 point fast Fourier transform (FFT) was used for calculating the STFT.

Figs. 2−6 show the waveforms and STFTs (spectrograms) of different cry signals. It is obvious that different catalogs of cry signals have different waveform and spectrum characteristics.

Diaper-related crying is considered a normal cry and has a pattern of crying and silence as shown in Fig. 2. This kind of crying starts with a cry coupled with a briefer silence, which is followed by a short high-pitched inspiratory whistle. Then, there is a brief silence followed by another cry. Fig. 3 shows attention-related crying, which is also a normal cry. This type of cry is characterized by a similar temporal sequence but can be distinguished by differences in the length of the various frequency components.

Hunger-related crying, which is also a normal cry, is the most general cry. The duration of crying is not only longer but it is also followed by a longer silence as shown in Fig. 4. Typically, this cry is louder and more abrupt compared with attention or diaper-related crying.

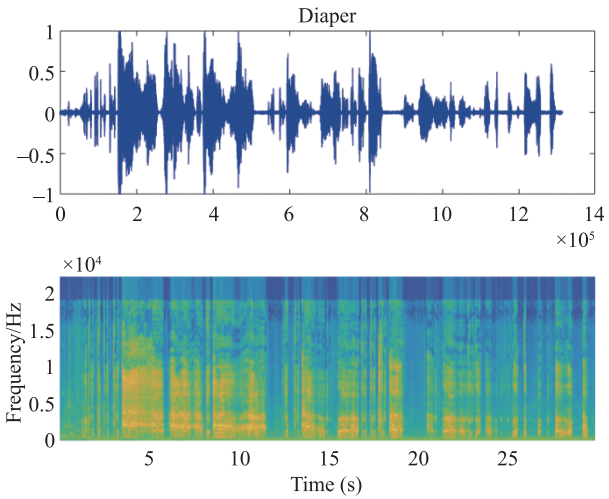


Fig. 2. Diaper signal wave form (upper) and spectrogram (lower).

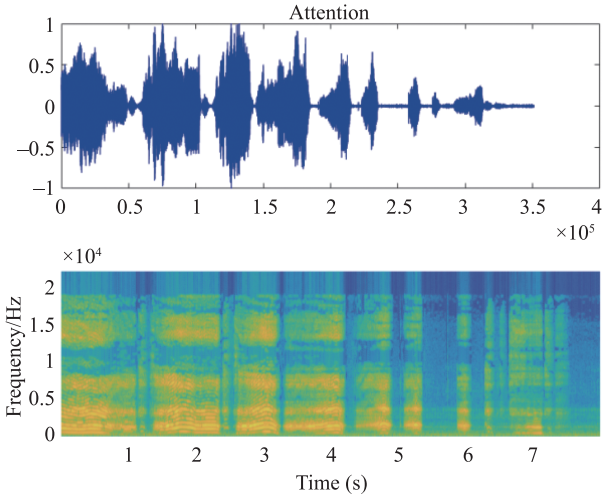


Fig. 3. Attention: signal waveform (upper), spectrogram (lower).

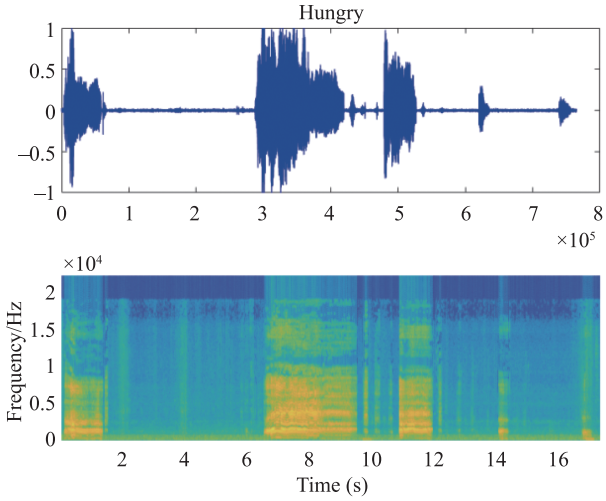


Fig. 4. Hungry: waveform (upper), spectrogram (lower).

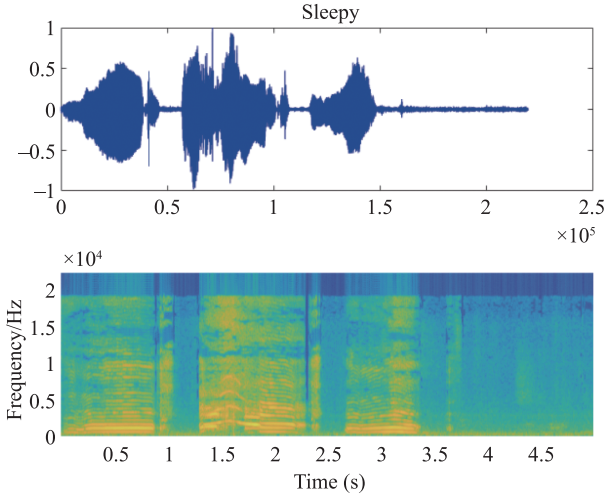


Fig. 5. Sleepy: waveform (upper) and spectrogram (lower).

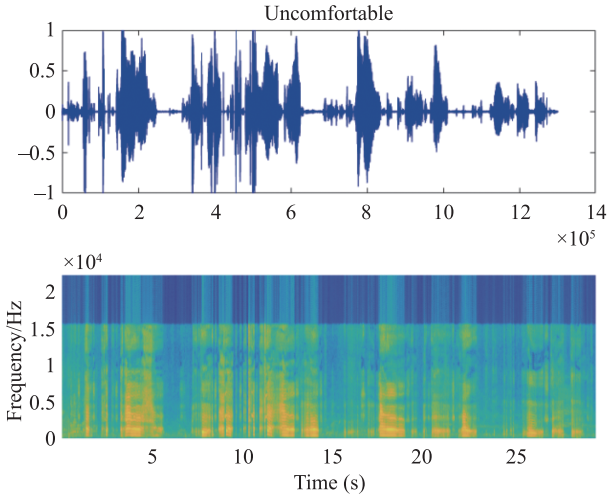


Fig. 6. Uncomfortable: waveform (upper), spectrogram (lower).

Fig. 5 shows sleep-related crying, which is also a normal cry. However, it is quite different from the previous normal cry signals. Its duration is longer and starting from a low amplitude, the cry gradually increases in loudness then drops slowly. The silent period is also a little longer compared with other cry signals.

Uncomfortable-related crying is shown in Fig. 6. Since it is a pain-related cry, unlike the previous normal cry signals, this cry has no preliminary moaning. The pain cry is a loud cry, followed by a period of breath holding [22].

In this study, 48 recording files were segmented by short time processing. Each wave file was processed to obtain "cry units". As an example, five "hungry baby cry units" were detected from file T19.wav. Fig. 7 shows the original cry signal, the short time energy analysis and the cry units detected based on STE. By using short time zero crossing, we see that higher zero crossing is associated with crying (Fig. 8). The detected cry units are the segments after the low power signal

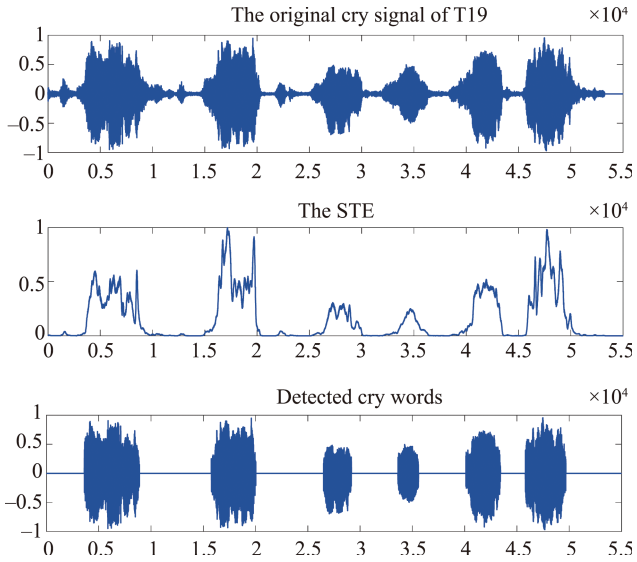


Fig. 7. Baby cry signal, short time energy and detected cry unit for cry file T19.wav.

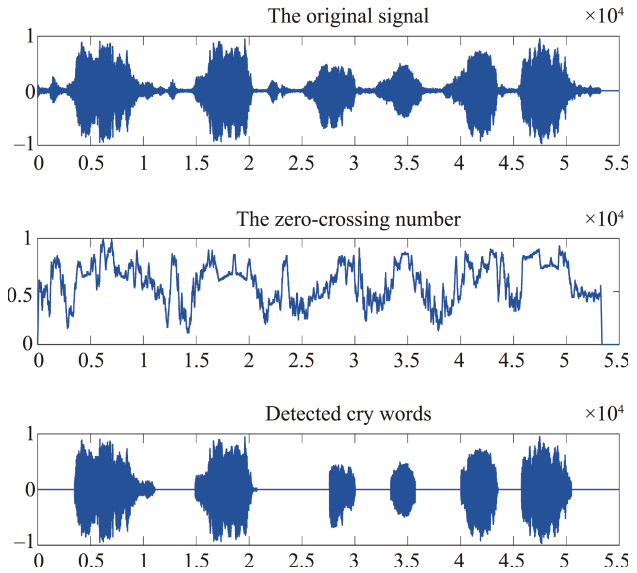


Fig. 8. Baby cry signal, short time zero-crossing and detected cry for T19.wav.

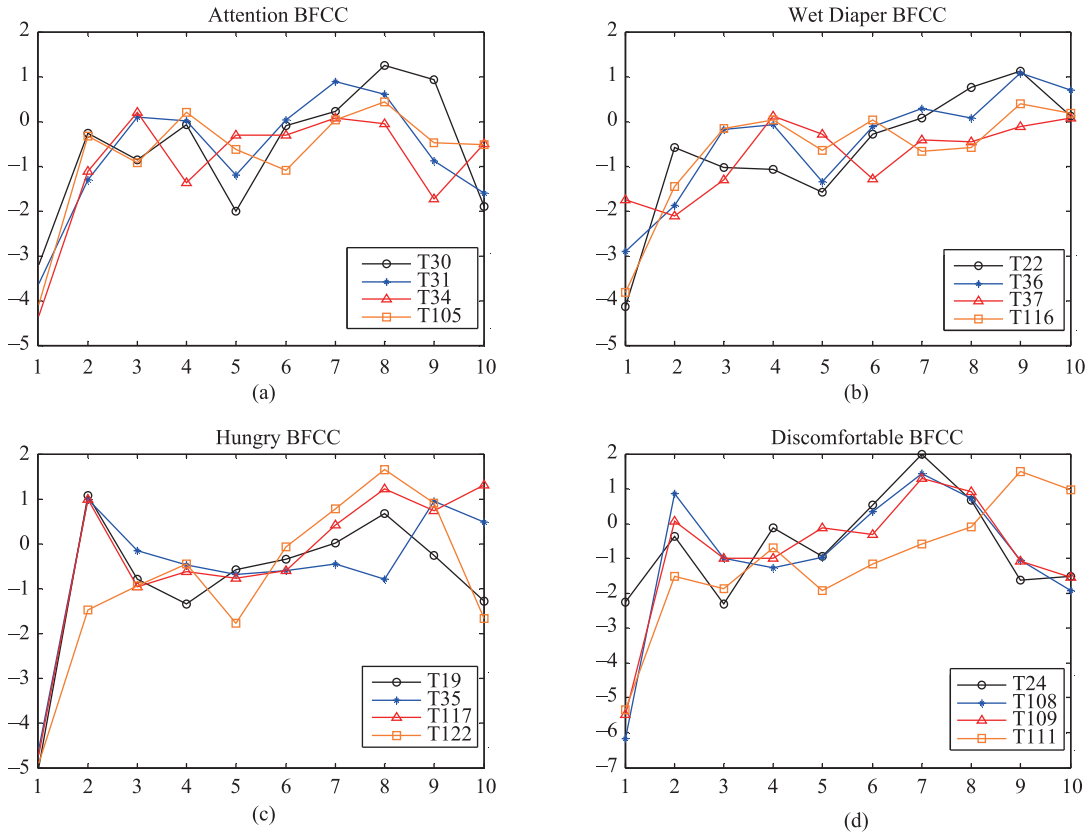


Fig. 9. BFCC features for attention, diaper, hungry and discomfort cry signals.

segments were removed from the cry signals. For all those (48) recording files, we got 151 "attention cry units", 137 "diaper change needed cry units", 422 "hungry cry units", 79 "sleepy" cry units and 182 "discomfort cry units".

Fig. 9 shows the BFCC features for different catalogs from different infants. BFCC features for attention from 4 different babies are shown in (a). Features from one infant are similar to other infants when they had a similar reason to cry. Subplot (b) shows "Diaper change needed cry units" BFCC features of 4 different cry files. Again, the results show similar features for needing a diaper change across different infants. Since attention-related crying and diaper-related crying both are characterized as normal crying, their intensity levels are similar but less than hunger-related crying.

Fig. 9 (c) shows "Hungry cry units" BFCC features of 4 different babies. Hunger-related crying had the highest intensity level in the normal cry catalog. BFCC features obtained from 'hungry' is quite different from those of 'attention' crying and 'diaper' crying. It is shown that the BFCC patterns changed from the low stress level cries to high stress level cries. There is an abrupt jump from coefficient 1 to coefficient 2 which is close to the trend of abnormal cry signals. Fig. 9 (d) shows discomfort-related crying from 4 files associated with 4 different babies. The BFCC features show a similar trend among those infants. They are quite different from normal cry signals, especially the low intensity level cries, such as diaper-related and attention-related crying. And even compared with hunger-related cry signals, the values of the coefficients were higher which means discomfort-related crying produced higher energy cry signals which matches the experts' experience.

Cry units from each class (100 "Draw attention cry units", 50 "Diaper change needed cry units", 120 discomfort, and 200 "Hungry cry units") were used as training signals, and the rest of the data (51 attention, 87 diaper and 222 hungry) were used for testing purposes. Figure 10 shows the cry units for 3 different cry signals by using LPC, LPCC, MFCC an BFCC features. It is obvious that different cry signals have different features.

Compressed sensing technique was used to conduct recognition and classification and classification rate was used to evaluate the performance and was defined as:

$$P_c = \frac{N_{right}}{N_{total}} \times 100\% \tag{14}$$

where N_{right} was the number of right classifications, and N_{total} was the total number of test cry units. We used sleep-related crying and hunger-related crying to test the performance of CS, 79 sleepy cry units and 200 hungry cry units have been used for training data and 100 sleepy cry units and 222 hungry cry units have been used as testing data.

Comparing Fig. 5 and Fig. 6, the differences between the waveforms of the two types of cries are more pronounced in the frequency domain than in the time domain. Since LPC features were obtained only in the time domain, CS cannot distinguish hungry and sleepy accurately based on LPC features, as shown in Table II. However, LPCC represents LPC coefficients in the cepstral domain to reflect the differences of the biological structure of the vocal track and the LPCC algorithm produces different features for hungry and sleepy

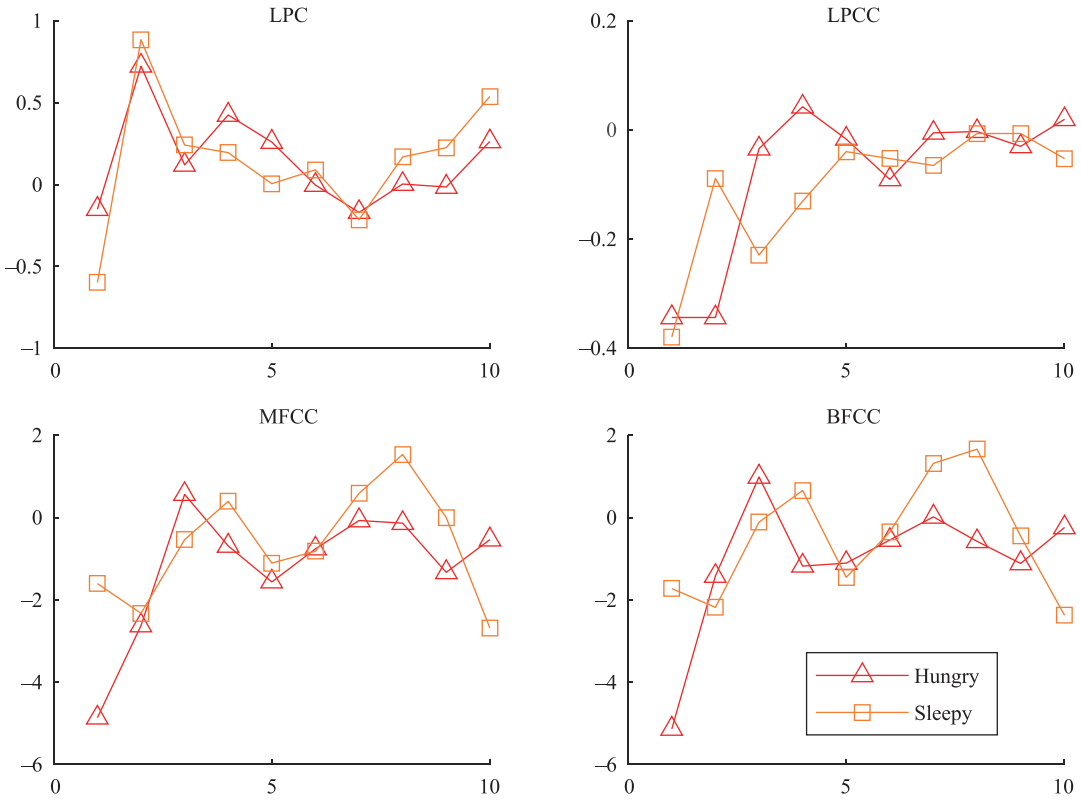


Fig. 10. Different features for hungry and sleepy cry.

cry. Nonetheless, since BFCC and MFCC algorithms capture both time and frequency domain information of cry signals, it is obvious that BFCC and MFCC can produce different features for hunger-related crying and sleepy crying, as shown in Fig. 10. As a result, BFCC and MFCC features outperform LPC and LPCC features with a classification rate around 70% (Table II).

Experimental results on hungry, discomfort, attention and diaper cry signals are shown in Fig. 11 and Table III. For the same reason mentioned above, BFCC and MFCC features outperform LPC and LPCC features.

We also investigate the performance in terms of the recognition rate for different features and different popular classification methods. We found that MFCC and BFCC outperform other features, and ANN and CS technique can provide higher recognition rate. Different combination can achieve different performance. For example, LPC can work with NN well, MFCC and LPCC combined with CS can achieve higher recognition rate than CS or NN. The highest recognition correct rate for infants cry application is achieved at 76.47% by using BFCC feature and ANN.

It is obvious that there are universal individual independent patterns for infant cry signals. Based on the time and frequency features, it is feasible to discern between different cry units. BFCC features and CS algorithms can provide reasonable and accurate recognition capabilities. Experimental results of the proposed approach match experts' knowledge and judgments very well.

TABLE II
INFANT CRY RECOGNITION CORRECT RATE WITH COMPRESSED SENSING TECHNIQUE AND DIFFERENT FEATURES FOR CRY SIGNALS (SLEEPY AND HUNGRY)

Features	The data ratio of constructing the matrix					
	0.4	0.5	0.6	0.7	0.8	0.9
BFCC	0.6991	0.6915	0.7067	0.6842	0.7105	0.6842
LPC	0.5133	0.4681	0.4933	0.4737	0.4211	0.5789
LPCC	0.6018	0.6064	0.6267	0.5965	0.5789	0.4737
MFCC	0.6814	0.6596	0.6767	0.7018	0.7105	0.6842

TABLE III
INFANT CRY RECOGNITION CORRECT RATE WITH COMPRESSED SENSING TECHNIQUE AND DIFFERENT FEATURES

Features	The data ratio of constructing the matrix					
	0.4	0.5	0.6	0.7	0.8	0.9
BFCC	0.5701	0.5393	0.5742	0.5754	0.5111	0.6842
LPC	0.6131	0.6225	0.5854	0.5688	0.5419	0.4667
LPCC	0.5009	0.4989	0.4986	0.4944	0.5028	0.4889
MFCC	0.5907	0.5910	0.5938	0.5502	0.5140	0.5333

VI. CONCLUSION

This paper presents a novel detection and recognition method for individual independent infant cries in a noisy environment. Audio features of infant cry signals were obtained in time and frequency domains, and were used to perform infant cry language recognition. Practical data from hospitals were

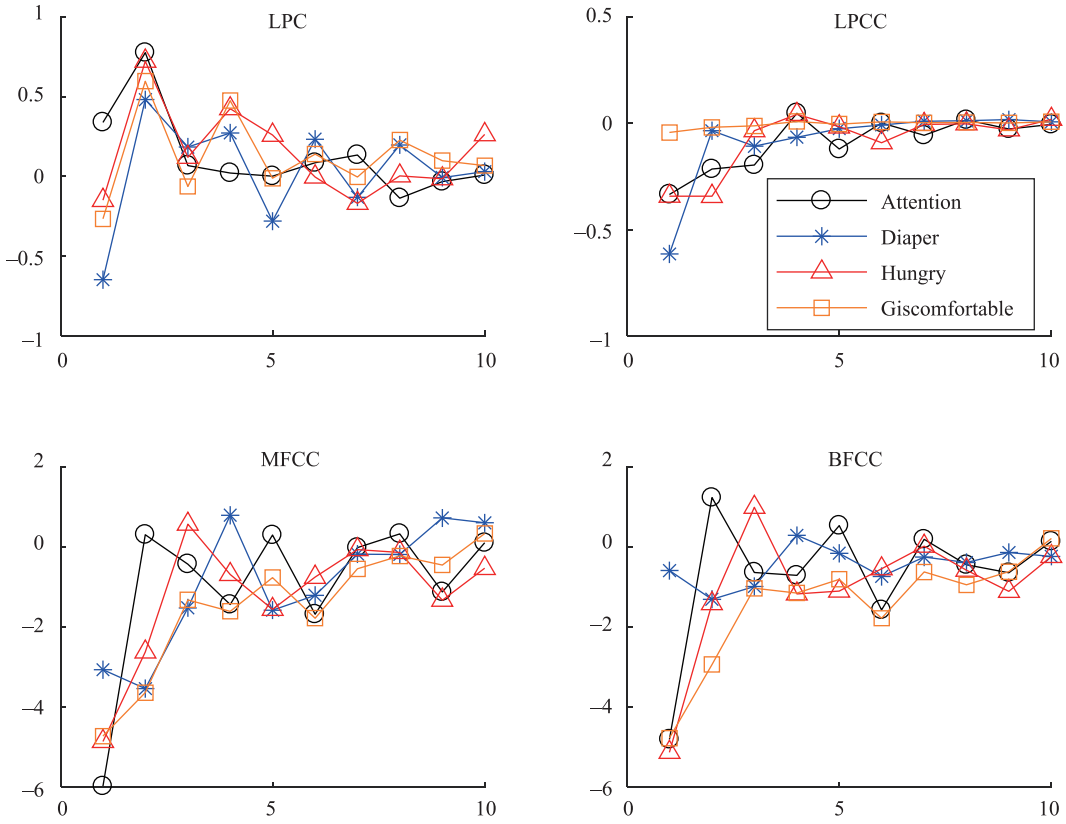


Fig. 11. Features for attention, diaper, hungry and discomfort cry.

TABLE IV
INFANT CRY RECOGNITION CORRECT RATE BY USING DIFFERENT FEATURES AND RECOGNITION TECHNIQUES

Features	LPC	LPCC	MFCC	BFCC
Nearest neighborhood (NN)	0.6384	0.4795	0.6389	0.6522
Artificial Neural Network (ANN)	0.5455	0.5188	0.6045	0.7647
Compressed sensing (CS)	0.5789	0.6267	0.7105	0.7064

used to design and verify the proposed approaches. Experiments proved that the proposed infant cry unit recognition models offer accurate and promising results with far-reaching applications medically and societally. Our future research includes: takes multiple features into consideration and reinforcement learning to improve the performance. We plan to collect more data and include more cry reasons as well.

REFERENCES

[1] H. Karp, The Happiest Baby on the Block; Fully Revised and Updated Second Edition: The New Way to Calm Crying, New York City, NY, USA, Bantam, 2015

[2] J. A. Green, P. G. Whitney and M. Potegalb, "Screaming, Yelling, Whining and Crying: Categorical and intensity differences in Vocal Expressions of Anger and Sadness in Children's Tantrums," *Emotion*. vol.5, no. 11, pp.1124–1133 Oct. 2011.

[3] Y. Kheddache, C. Tadj, "Acoustic measures of the cry characteristics of healthy newborns and newborns with pathologies," *Journal of Biomedical Science and Engineering*, vol.6, no.8, 9 pages, 2013.

[4] L. Liu, K. Kuo and Sen M. Kuo, "Infant Cry Classification Integrated ANC System for Infant Incubators," *Proc. IEEE International Conf. on Networking, Sensing and Control*, Paris, France, 2013, pp. 383-387.

[5] L. Liu, K. Kuo, "Active Noise Control Systems Integrated with Infant Cry Detection and Classification for Infant Incubators," *Proc. Acoustic 2012*, HongKong, pp.1-6. 2012, 2012.

[6] L. LaGasse, A. Neal and M. Lester, "Assessment of infant cry: acoustic cry analysis and parental perception," *Ment Retard Dev Disabil Res Rev*., vol.11, no.1, pp.83-93, 2005.

[7] Várallyay Jr., György, "Future Prospects of the Application of the Infant Cry in the Medicine," *Periodica Polytechnica Ser. El. Eng*. vol. 50, no. 1-2, pp. 47-62, 2006.

[8] G. Buonocore, and C.V. Bellieni, "Neonatal Pain, Suffering, Pain and Risk of Brain Damage in the Fetus and Newborn," Berlin, Germany, Springer, 2008.

[9] L. L. LaGasse, R. Neal and B. M. Lester. "Assessment of infant cry: acoustic cry analysis and parental perception," *Mental Retardation and Developmental Disabilities Research Reviews*, vol. 11, no. 1. pp.83–93, 2005.

[10] L. Tan, J. Jiang, Digital Signal Processing: Fundamentals and Applications, , 3rd edition. Cambridge, MA, USA, Academic Press, 2017.

[11] Z. Ren, K. Qian, Z. X. Zhang, V. Pandit, A. Baird, and B. Schuller, "Deep scalogram representations for acoustic scene classification," IEEE/CAA J. of Autom. Sinica, vol. 5, no. 3, pp. 662-669, May 2018.

[12] Dong Yu, Jinyu Li. "Recent progresses in deep learning based acoustic models," IEEE/CAA Journal of Automatica Sinica, vol. 4, no. 3, pp. 396-409, April 2017

[13] B. Goldand N. Morgan, Speech and Audio Signal Processing. New York, NY, USA, John Wiley & Sons, 2011.

[14] V. R. Fisichelli, S. Karelitz, C.F.Z. Boukydis, and B.M. Lester, "The cry attencies of normal infants and those with brain damage," *Infant Crying*, Plenum Press, 1985.

[15] C. F. Z. Boukydis and B. M. Lester, "Infant Crying: Theoretical and Research Perspectives," Berlin, Germany, *Springer Science and Bussiness Media*, 2012.

[16] S. Ludington-Hoe, X. Cong and F. Hashemi, "Infant crying: nature, physiologic consequences, and select interventions," *Neonatal Netw*. vol. 21, no. 2, pp.29-36. Mar. 2002.

[17] P. Dunstan, Calm the Crying: The Secret Baby Language That Reveals the Hidden Meaning Behind an Infant's Cry, New York City, NY, USA, Avery, 2012.

[18] M. Sahidullah, G. K. Saha, "Design analysis and experimental evaluation of block based transformation in MFCC computation for speaker recognition," *Speech Communication*, vol.54 Issue 4, pp.543-565, May 2012.

[19] F. Katzberg, R. Mazur, M. Maass, P. Koch and A. Mertins, "A compressed sensing framework for dynamic sound-field measurements," *IEEE/ACM Trans. on Audio, Speech, and Language Processing*, vol. 26 , Issue 11, pp. 1962 – 1975, Jun. 2018.

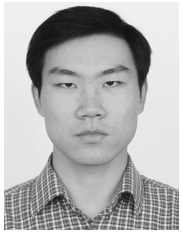
[20] D. Needell and R. Ward, "Two-subspace Projection Method for Coherent Overdetermined Systems," J*ournal of Fourier Analysis and Applications,* vol. 19, Issue 2, pp. 256–269, April 2013.

[21] C. Lau, "Development of suck and swallow mechanisms in infants," *Ann Nutr Metab*, vol. 7, no. 5, pp. 7−14, July, 2015.

[22] P. Runefors and E. Arnbjönsson, "A sound spectrogram analysis of children's crying after painful stimuli during the first year of life," *Folia honiatr Logop*, vol. 2, no.57, pp. 90−95, Mar-Apr., 2005.

Lichuan Liu received her B.S. and M.S. degree in Electrical Engineering in 1995 and 1998 respectively from University of Electronic Science and Technology of China, and Ph. D. degree in Electrical Engineering from New Jersey Institute of Technology, Newark, NJ in 2006. She joined Northern Illinois University in 2007 and is currently an Associate Professor of Electrical Engineering and the Director of Digital Signal Processing Laboratory. Her current research includes digital signal processing, real-time signal processing, wireless communication and networking. She has over 70 publications including 30 journal papers and one book chapter. She has three patents awarded. She has led and participated in many research grants, such as: NSF, NASA and NIH.

Wei Li (M'06-SM'11) received his Ph.D. degree in Electrical and Computer Engineering from the University of Victoria, Canada in 2004. He is currently an Assistant Professor at the Northern Illinois University, USA. His research interests are computer networks, smart grid, Internet of Things, applications of machine learning and artificial intelligence in e-Health, computer vision and natural language processing.



Xianwen Wu (M'06) received the B.S. degree in electrical engineering from North University of China in July 2005, the M.S. degree in biomedical engineering from Southeast University in June 2010, and the M.S. degree in electrical engineering from Northern Illinois University in August 2013. He received his Ph.D. degree in electrical engineering from the University of Arkansas in December 2016 and then joined Qualcomm Inc. as a system engineer. His research focuses on communication theory, wireless sensor networks, and signal processing.

Benjamin Zhou is currently pursuing a B.S. in Biology at The College of New Jersey as part of the 7-year B.S/M.D. program with NJMS. He currently researches at Perelman School of Medicine at the University of Pennsylvania. His current research interests include sepsis and its effects on the immune system using animal models.