Infant Cry Classification Integrated ANC System for Infant Incubators

Lichuan Liu, Kevin Kuo and Sen M. Kuo
Department of Electrical Engineering
Northern Illinois University
DeKalb, IL, USA
liu@niu.edu, akevinkuo@gmail.com, kuo@niu.edu

***Abstract*—The high noise level inside the infant incubator results in numerous adverse health effects for premature newborns and the active noise control (ANC) systems are developed to reduce the noise. This paper proposes an infant cry classification integrated ANC system for infant incubators. The developed system can dramatically reduce the harmful noise level, and the integrated infant cry detector and analyzer can monitor the infants' physical conditions. The infant cry signals are picked up and detected by the same microphones used by the ANC system, the cry signal's features are extracted and then recognized. The simulation and experiment results show that the cry recognition of specific infants yielded promising results.**

Keywords—Infant cry classification, active noise control, infant incubator.

I. INTRODUCTION

Every year, 20 million premature, low-birth-weight, very ill babies are born; a large number of these babies are saved by incubators. However, the high-level noise inside the incubator generated by medical equipment and activities of caregivers results in numerous adverse health effects [1,2]. Most attempts to improve the acoustic environment of the neonatal intensive care units (NICU) have focused on reducing staff activities [3], and/or incorporating sound containment and absorption strategies into the design of new NICUs [4]. Those methods block the view of incubator and are also ineffective for low-frequency NICU noises.

Unlike the conventional passive noise control methods such as sound-absorbing panels, the active noise control (ANC) system generates an 'anti-noise', which acoustically cancels the unwanted noise based on the principle of superposition. This research showed the ANC system can become a promising solution towards reducing the noise level inside infant incubators [5, 6].

It is well known that the infant's sound signal, for example, cry serves as the primary means of communication for infants. It is possible for experts (parents and child care specialists) to distinguish infant cries though training and experience. Therefore, it should be possible to extract audio features from the infant cry such that it can be uniquely differentiated from cries of different meaning. However, 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 [7] such as pain from immunization shots, fear from jack-in-the box toys, and frustration from head restraints. Therefore, it is important to detect and classify the infant audio signals in order to monitor infant's health features by using digital signal processing techniques. The proposed system will process the infant sound signal and alert and inform the parent or caretaker of the most likely reason behind the sound signals using the ANC system hardware.

II. INFANT CRY SIGNAL

A. Anatomy of Infant Vocal Tract

The airways of newborn infants are quite different from those of adults. The larynx in newborn infants is positioned close to the base of the skull. The high position of the larynx in the newborn is similar to its position in other animals and allows the newborn human to form a sealed airway from the nose to the lungs.

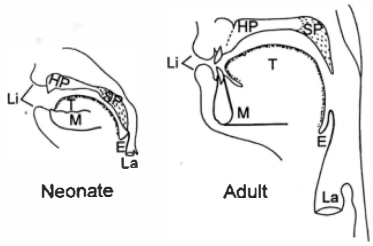


Fig. 1. Comparison of vocal tract between neonate (left) and adult (right) [8]

The soft palate and epiglottis work as a double seal, and liquid can flow around the relatively small larynx into the esophagus while air moves through the nose, through the larynx and trachea into the lungs. The anatomy of the upper airways in newborn infants is "matched" to a neural control

This research is supported by National Institution of Health (NIH) and Gerber Foundation.

system – newborn infants are obligated nose breathers. They normally will not breathe through their mouths even when their noses are blocked. The unique configuration of the vocal tract is the reason for the extremely nasalized cry of the infant.

B. *Physiology of Infant Cry*

Crying is a heightened (the highest) state of arousal produced by nervous system excitation triggered by some form of biological threat that may involve basic physiological processes, such as hunger, pain, sickness or insult, or individual differences in threshold for stimulation. Crying is modulated and developmental facilitated by control mechanisms that enable the infant to maintain non-crying states.

Newborns differ from one another in their response to the stimuli. They differ in the number of states available and in the way they switch between states, i.e. how rapidly and regularly. Distinguishing among these various types of infants will depend upon how the rest of the infant's behavior relates to the pattern of crying. Physiological changes directly affect cry behavior. In the first few weeks of life, crying has a reflexive-like quality and is most likely tied to the regulation of physiological homeostasis as the neonate is balancing internal with external demands.

III. System Model and ANC System for Infant Incubator

A. *System Model*

Figure 2 shows the system block diagram of the infant cry classification integrated ANC system. The integrated system includes the ANC system to reduce the harmful noise inside the incubator and the cry classification system. The same hardware is used for both subsystems. In cry signal classification part, there are cry detection, signal feature extraction and classification blocks.

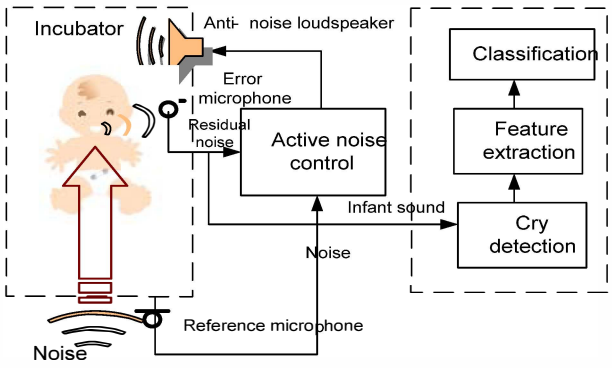


Fig. 2 Block diagram for Infant cry classification integrated ANC system

B. *A Brief Description for Infant Incubator ANC System*

A multiple-channel feedforward ANC system uses one reference microphone, two secondary speakers, and two error microphones independently, as shown in Fig. 3. Two error microphones inside the incubator obtain the error signals $e_1(n)$ and $e_2(n)$ at different positions, and the system is thus able to form a larger quiet zone centered at the error microphones. The ANC algorithm uses two adaptive filters $W_1(z)$ and $W_2(z)$ to generate anti-noise $y_1(n)$ and $y_2(n)$ to drive the two independent secondary loudspeakers. In Fig. 3, $\hat{S}_{11}(z)$, $\hat{S}_{12}(z)$, $\hat{S}_{21}(z)$ and $\hat{S}_{22}(z)$ are the estimates of the secondary path transfer functions.

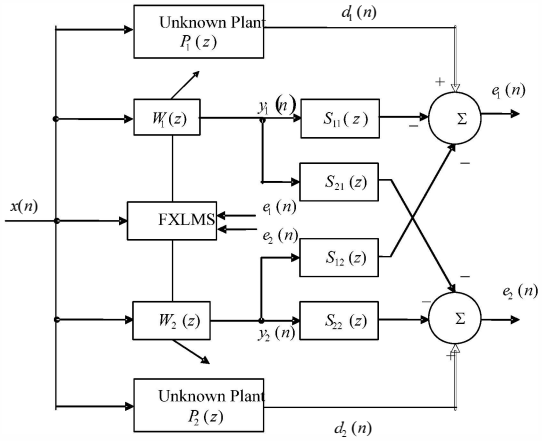


Fig. 3.The 1x2x2 FXLMS ANC algorithm

The 1x2x2 FXLMS ANC algorithm is summarized as follows [9]:

$$y_i(n) = \mathbf{w}_i^T(n)\mathbf{x}(n),\ i = 1,2 \quad (1)$$

where $w_1(n)$ and $w_2(n)$ are coefficient vectors and μ_1 and μ_2 are the step sizes for the adaptive filters $W_1(z)$ and $W_2(z)$, respectively, and $\hat{s}_{ij}(n)$ with $i = 1,2$ and $j = 1,2$ are the impulse responses of the secondary path estimates.

$$\begin{aligned}\mathbf{w}_i(n+1) &= \mathbf{w}_i(n) \\ &+ \mu_i\left[e_1(n)\mathbf{x}(n) * \hat{s}_{1i}(n) + e_2(n)\mathbf{x}(n) * \hat{s}_{2i}(n)\right],\quad i = 1,2\end{aligned} \quad (2)$$

IV. Cry Signal Detection

Cry signal detection is desirable to process 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.

Short-time energy (STE) is defined as the average of the square of the sample values in a suitable window. It can be mathematically described as follows [10]:

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

where $w(m)$ are coefficients of a suitable window function of length N. Short-time processing of speech should take place during segments between 10-30 ms in length. For our signals of 8 kHz sampling frequency, a window of 128 samples (~16 ms) was used. STE estimation is useful as a speech detector

because there is a noticeable difference between the average energy between voiced and unvoiced speech, and between speech and silence [10]. This technique is usually paired with short-time zero crossing for a robust detection scheme.

Short-time zero crossing (STZC) is defined as the rate at which the signal changes sign. It can be mathematically described as follows [10]:

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

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

STZC estimation is useful as a cry detector because there are noticeable fewer zero crossings in cry as compared with non-cry signals.

V. Feature Extraction

Even though cry is a non-stationary signal, over short time intervals, cry segments can be considered stationary. There are several techniques used to extract features from stationary signals: frequency extraction, homomorphic cepstral coefficients, Mel frequency cepstral coefficients (MFCC), linear predictive coding coefficients (LPCC), Bark frequency cepstral coefficients (BFCC) and perceptual linear prediction. These stationary feature extraction techniques can be classified into either cepstral based (taking the Fourier transform of the decibel spectrum) or linear predictor (determining the current speech sample based on a linear combination of prior samples) based algorithms. Furthermore, it is possible to apply filter banks of the Mel or Bark scale in order to perform pitch warping or subjective loudness warping, respectively, to the speech segments prior to analysis. Both time and frequency domain analysis techniques will be used to extract pitch, loudness, tonal and inflection characteristics from the cry signals. The techniques used in this paper are LPCC, MFCC and BFCC and their application will be described further on.

A. Linear predictive coding

There are two acoustic sources associated with voiced and unvoiced sound, respectively. Voiced sound is caused by the vibration of the vocal cords in response to airflow from the lung and this vibration is periodic in nature while unvoiced sound is caused by constrictions in the air tract resulting in random airflow [10]. The basis of the source-filter model of sound is that sound can be synthesized by generating an acoustic source and passing it through an all-pole filter. The present sample of the sound as a linear combination of the past M samples of the speech such that:

$$\hat{x}(n) = \sum_{i=1}^{M} a_i x(n-i) \quad (5)$$

where $\hat{x}(n)$ is the predicted value of $x(n)$, $\{a_i\}$ are the linear prediction coefficients and M is the number of poles (the roots of the denominators in the z transform) of the all-pole filter.

Then the coefficients $\{a_i\}$ can be estimated by either autocorrelation or covariance methods [11]. Effectively, the purpose of LPCC is to take a large size waveform and then compress it into a more manageable form. Because similar waveforms should also result in similar acoustic output, LPC serves as a time domain measure of how close two different waveforms are.

B. Mel Frequency Cepstral Coefficients

Mel frequency cepstral coefficients (MFCC) are coefficients that describe the mel frequency cepstrum [11]. In sound processing, the 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 through the following steps. The short-time Fourier transform of the signal is taken in order to obtain the quasi-stationary short-time power spectrum $F(f)=F\{f(t)\}$.The frequency portion of the spectrum is then mapped to the mel scale perceptual filter bank with the equation above using 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 in order to obtain the envelope of the spectrum with harmonics. The log of this filtered spectrum is taken and then the Fourier transform of the log spectrum squared results in the power cepstrum of the signal.

$$X_k = \sum_{n=0}^{N-1} x_n \cos\left[\frac{\pi}{N}(N+\frac{1}{2})k\right] \quad (6)$$

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.

C. Bark Frequency Cepstral Coefficients

Similar to the MFCC, the BFCC warps the power cepstrum such that it matches human perception of loudness. The methodology of obtaining the BFCC is similar to that of the MFCC except for two differences. The frequencies are converted to bark scale with the formula below:

$$\mathrm{b} = 13\tan^{-1}(.00076f) + 3.5\tan^{-1}\left[(\frac{f}{7500})^2\right], \quad (7)$$

where b denotes bark frequency and f is 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.

The BFCC is obtained by taking the DCT of 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 BFLPCC is obtained by performing the LPC algorithm on the

power cepstrum in 128 sample frames. The BFLPCC encodes the cepstrum waveform in a more compact fashion that makes it more suited for the classification scheme to be described later.

Audio feature extractions hinges upon using techniques in digital signal processing of audio signals to quantize acoustic information in a manner that makes classification tractable.

VI. Classification

Earlier, signal detection of cries was used to find the waveform boundaries of cries to be processed. Furthermore, these cry signals were further processed in smaller frames every 16 ms in time (or 128 samples) to accurate obtain feature parameters that could describe the frame in detail in a fashion such that it could be compared with parameters from different signals without losing the integrity of the comparison.

As the features of the infant cry signals were deconstructed into an array of LPCC, MFCC or BFCC. The format of LPCC is convenient in that the coefficients have a small dynamic range of value 2, as the coefficients have a value that range from -1 to 1. The MFCC and BFCC coefficients that are derived through DCT have a larger dynamic range between coefficients. This makes it unsuitable for a direct comparison between DCT and LPC derived coefficient values, but is still viable for comparison within its own family of results.

To generate static codebooks, codebooks whose size is independent of the signal duration, each cry signal was subdivided into 16 blocks. Within these 16 blocks, frame by frame analysis inside the blocks was performed by implementing the feature extraction algorithms over 50% overlapping frames of 16 ms that were windowed by a hamming window. The codebook for three different cry signals (1=hungry, 2=diaper and 3=attention)

$$\mathbf{C} = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1N} \\ c_{21} & c_{22} & \cdots & c_{2N} \\ c_{31} & c_{32} & \cdots & c_{3N} \end{bmatrix}$$

where N=16.

The method of choosing the best codebook matches was determined by developing a cost function, whose parameter was mean square error, where the lowest cost function values were designated as the best fit [11].

The test feature is $\mathbf{t} = \begin{bmatrix} t_1 & t_2 & \cdots & t_N \end{bmatrix}$, where N=16, and classification result is

$$\arg\min_i \sum_{j=1}^{N} (t_j - c_{i,j})^2$$

VII. Data Acquisition and Results

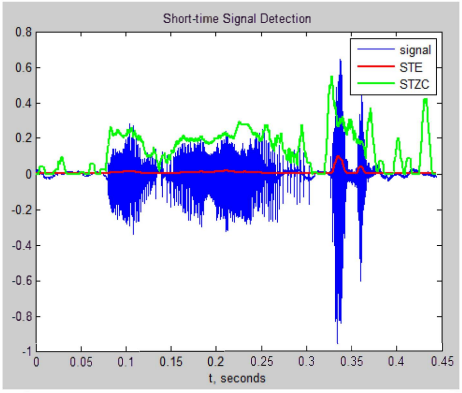
A. Data Acquisition

The process of isolating infant cries for signal processing was a multistage process. The first stage consisted of making audio recordings at a local daycare while obtaining the likely cry pathology information from the caregivers. As the recorded subjects in the infant room varied in age (0-1 years) and gender, further compartmentalization of the recordings could be performed for a more comprehensive statistical signal analysis.

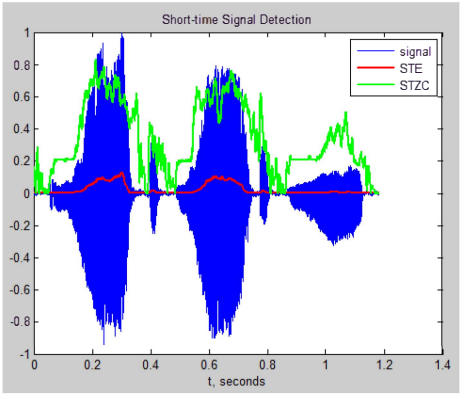
Each infant cry recording was 20 seconds in length with a sampling frequency of 48 kHz with a resolution of 16 bits. The cries were recorded using a microphone placed approximately 10 inches away from each infants face. These recordings were then classified to discrete causes by the caregivers. A total of 29 cries from 8 infants (4 male and 4 female) were obtained. At the time of recording, the infants ranged in age from 7 weeks up to 1 years old with a mean age of 7 months with a standard deviation of 4 months. The cries recorded by each infant are attached along with the diagnosis of the pathology of the infant cry, details concerning the recording (date, time, filename) as well as circumstances of the recording (background noises).

B. Results

There were 29 cries obtained from 8 infants of with cries designated to three basic causes: attention (or tired), hunger and wet diaper. Some of the causes for the cries identified by the caretakers overlapped these categories and such cries were taken note of.



(a)



(b)

Fig 4: Waveform of (a) two whimpers followed by a cough, (b) two cries and a whimper with impulsive cry artifacts.

In Fig. 4, each cry envelope is bounded by the STZC and the voiced portion of each cry is bounded by where the STE meets the $t = 0$ axis.

It was assumed that all of the cries recorded were indicative of the root cause of the cry. This assumption is based on the belief that an infant will not change their current mood or desires within the 20 second recording. From the recorded cries, the recordings that did not seem to be resultant of mixed causes were used for references while the less certain cries were kept for testing and comparison purposes.

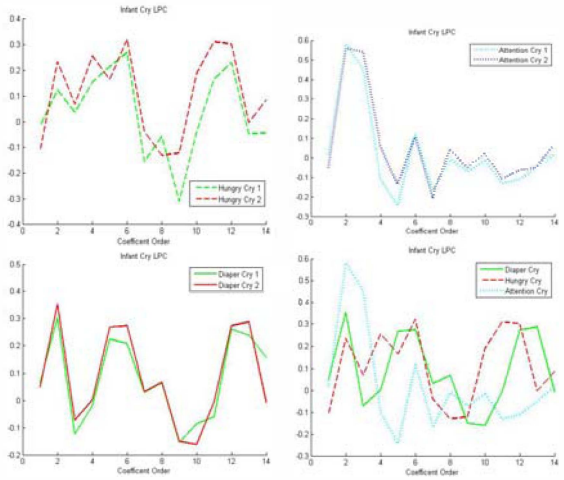


Fig. 5 (a) LPCC of adjacent attention cries, (b) LPCC of adjacent diaper cries, (c) LPCC of adjacent hungry cries, (d) LPCC of attention, diaper and hungry cries.

The reference codebooks as shown in Fig. 5 were generated from a sequence of cries in recording where we determined that the cry segments were not affected by background noises. The resultant codebook for each pure cry in the sequence would then be averaged and then compared against each other individual cry in the recording.

Cry samples	Number of cries	Hungry cry	Diaper cry	Attention cry
Hunger	28	25	3	10
Diaper	12	1	12	2
Attention	11	6	2	10

Table 1. Analysis of the three reference cries using the cry codebooks.

It is shown from Table 1, for an individual infant cry classification, the global classification is 60-80%. However, there is no substantial evidence that would suggest that there are universal patterns due to our sample size.

VIII. Conclusions

The infant cry classification integrated ANC system is presented in this paper. The hardware of the ANC system can be used to detect and classify infant audio signals such as the infant cry. It is believed that there are individual patterns that distinguish the meaning of infant cries, the recognition rate obtained for an individual infant cry is 60-80%. However, there is no substantial evidence that would suggest that there are universal patterns based on our current sample size. Future work can include determining the meanings behind infant cries based on not only audio cues but also upon observation and deductive logic as well.

Acknowledgment

This research is supported by National Institution of Health (NIH) and Gerber Foundation.

References

[1] S.E Jacobs, K. O'Brien, S. Inwood, E. N. Kelly and H. E. Whyte, H.E., "Outcome of infants 23-26 weeks' gestation pre and post surfactant". Acta Paediatr.Vol. 89: pp/ 959-965, 2000,

[2] J.M. Lorenz, "The outcome of extreme prematurity". Semin. Perinatol. Vol. 2, pp: 348-359, 2001

[3] A. Robertson, C. Cooper-Peel and P. Vos, "Sound transmission into incubators in the neonatal inten-sive care unit". J. Perinatol. Vol. 19, pp. 494-497, 1999

[4] M. K. Philbin and J. B. Evans, "Standards for the acoustic environment of the newborn ICU". J. Perinatol. Vol. 26, pp.27-30, 2006

[5] S. Kuo, L. Liu L and S. Gujjula "Development and Application of Audio-Integrated ANC System for Infant Incubators," Noise Control Engineering Journal, Vol. 58, pp.163-175, 2010

[6] 1. L. Liu, K. Beemanpally and S. Kuo, "Real-time experiments of ANC systems for infant incubator", Noise Control Engineering Journal, Vol. 60, No. 1, Jan-Feb, 2012, pp36-41.

[7] Petroni, M. Malowany, A.S. Johnston, C.C. and Stevens, B.J., "Classification of Infant Cry Vocalizations using Artificial Neural Networks (ANNs)," in Proc. of ICASSP-95, vol. 5. May 1995, pp. 3475-3478

[8] Lederman, D. "Automatic Classification of Infants' Cry". M.Sc dissertation, Ben-Gurion University, Beer-Sheva, Israel, 2002.

[9] Kuo, S. M. and Morgan, D. R., "Active noise control: A tutorial review," Proc. of the IEEE, Vol. 87, pp. 943-973, 1999

[10] J.R. Deller, J.G. Proakis, and J.H.L. Hansen, "Discrete-Time Processing of Speech Signals", Prentice Hall, NJ, USA, 1993.

[11] S. Theodoridis, K. Koutroumbas, "Pattern Recognition", Academic Press, San Diego, London, 1998