Infant Cry Analysis and Detection

Rami Cohen\textsuperscript{1} and Yizhar Lavner\textsuperscript{1,2}

\textsuperscript{1}Signal and Image Processing Lab
Department of Electrical Engineering
Technion - Israel Institute of Technology

\textsuperscript{2}Department of Computer Science, Tel-Hai College

15 November, 2012

2012 IEEE 27-th Convention of Electrical and Electronics Engineers in Eilat, Israel
1. Infant Cry
   - Former studies of Acoustical Characteristics of Infant Cry
   - Purpose of this Study
   - Acoustic Characteristics of cry signals
1 Infant Cry
   - Former studies of Acoustical Characteristics of Infant Cry
   - Purpose of this Study
   - Acoustic Characteristics of cry signals

2 Cry Detection Algorithm
   - Performance Evaluation
   - Results
Outline

1. Infant Cry
   - Former studies of Acoustical Characteristics of Infant Cry
   - Purpose of this Study
   - Acoustic Characteristics of cry signals

2. Cry Detection Algorithm
   - Performance Evaluation
   - Results

3. Conclusion
Infant Cry

- Biological alarm system
- First means of communication for newborns
- Signals distress or needs, calls for the attention of parents or caregivers
- Carries information about infants physical and physiological condition: health, weight, identity, gender and emotions
Former Studies of Infant Cry

- Recognition of potential neurological insults or the medical status of newborns
- Discrimination between normal and hearing impaired babies
- Enhancing social robots behaviour in childhood education settings
- Identification of babies solely by the sound of their cry


Detection of physical danger to infants

Detection

- Automatic detection of an infant cry
- Robustness in presence of noise

Alert

- Alerting parents in occasions where infants are left alone in closed apartments or vehicles
Acoustic Characteristics of cry signals

Main Features

- Periodic bursts: alternating cry utterances and inspirations
- Fundamental Frequency: 250Hz–600Hz
- Voiced Sound: Harmonics
In-depth Analysis

Short-time energy

\[ E_{[dB]} = 10 \times \log_{10} \left( \frac{1}{N} \sum_{n=N_0}^{N_0+(N-1)} x^2[n] \right) \]

Burst frequency

- Computed from the DFT of the short-time energy
In-depth Analysis

Fundamental frequency (Pitch)

- Cry bursts are produced by quasi-periodic excitations of the vocal tract
- The cry waveform is pseudo-periodic: at each point $t_0$ there exists such $T$, so that $x(t) \approx x(t + T)$
- Typical pitch periods are $1.5 - 4$ ms

- Pitch detection is based on integration of two methods:
  - Cepstrum method (Noll, 1967): Finding the peak in the cepstrum
  - Cross-correlation method: Finding the maximal correlation between adjacent sections of the signal
In-depth Analysis

Mel-Frequency Cepstrum Coefficients (MFCC)

- Filtering the signal using triangularly shaped filters, getting wider and sparser as the frequency rises
- Based on principles of human auditory perception

![Diagram of MFCC process]

In-depth Analysis

Harmonicity factor (HF)

• An estimate of the presence of harmonics in each analysis frame
  • Finding $N$ highest peaks in the DFT
  • Denoting their corresponding frequencies as $h_1, h_2, \ldots, h_N$
  • The *harmonicity factor* is defined as:

  $$HF = \sum_{i=1}^{N} h_i \mod f_0$$

• Tends to zero for harmonic signals
In-depth Analysis

Harmonic-to-Average Power Ratio (HAPR)

- HAPR is a spectral feature that determines the ratio of the harmonic component power and the average spectral power.

Computation steps:
1. Identifying the highest peaks around the harmonic frequencies in the DFT magnitude.
2. Computing the power component around the $m$th harmonic:
   $$|X(2\pi mf_0, t)|^2$$
3. Calculating the average spectral power:
   $$P_x(t) = \frac{1}{N} \sum_{k=0}^{N} |X(\omega_k, t)|^2$$
4. Finally:
   $$HAPR(t, M) [dB] = \frac{1}{M} \sum_{m=2}^{M} 10\log_{10} \frac{|X(2\pi mf_0, t)|^2}{|P_x(t)|^2}$$
Pitch, Harmonicity Factor and HAPR
Cry Detection Algorithm

Detection Algorithm Stages

Principles

- Various time-scales: frames (tens of ms), short segments (1 sec) and long segments (10 sec), are used
  - Obtaining high detection rate, with few false-positives

Stages

1. Dividing the signal into long segments (10 sec)
2. Applying Voice Activity Detector (VAD)
3. Dividing the segments into short ones (1 sec)
4. Computing pitch, MFCC, HF, HAPR for each frame (32 ms)
5. Classifying each frame as '1' (cry) or as '0' (not cry) using KNN, based on the MFCC
6. A segment is classified as cry if enough sections are classified as cry
Post processing

- Further validation in case of 'no cry' decision by the k-NN classifier, using:
  - Pitch
  - Harmonicity Factor (HF)
  - Harmonic-to-Average Power (HAPR)

- According to the values of the parameters above, the decision can be changed to 'cry'

```
PITCH DETECTION

HF COMPUTATION

HAPR COMPUTATION

At least 5 frames with $300 < f_0 < 600$?
At least 5 frames with $HF < 100$?
At least 50% of the frames with $300 < HAPR < 600$?
```
Performance evaluation

Database

- Database contained:
  - Baby cry samples
  - Car engines
  - Motorcycles
  - Car horns
  - Passer-by
  - Speech

Experimental setup

- The signal was contaminated with AWGN, using varying SNR
- Section (1 sec) and Segment (10 sec) detection rate was measured, using cross-validation

An efficient algorithm for detection of infant cry signals was presented. Three decision levels in different time-scales are used: a frame level, sections of few hundred msec and segments of several seconds. Keeping low rate of false negatives. Small number of features, relatively simple for implementation. The results demonstrate both high detection rate and robustness in presence of noise, even in low SNR values.
Thank You