Research Article

On the Robustness of Cry Detection Methods in Real Neonatal Intensive Care Units

Dang Manh Chinh[1]*, Antoni Martínez-Ballesté[2], Pham Ngoc Minh[1] and Dang Thanh Trung[3]

[1]*Institute of Information Technology, Hanoi, Vietnam*

[2]*Universitat Rovira i Virgili, Catalonia, Spain*

[3]*Electric Power University, Hanoi, Vietnam*

*Corresponding Author: Dang Manh Chinh, Institute of Information Technology, Hanoi, Vietnam.

Abstract

   The detection of cry is crucial in intelligent computerized systems that aim at assessing the wellbeing of neonates during their hospitalization periods. Moreover, a precise characterization of cry allows its classification (e.g. hunger, pain, tiredness…). Although several cry detection and characterization techniques can be found in the literature, there is no testing of such techniques in real-life environments such as hospital Intensive Care Units. In this article we first summarize the problem of background noise in Intensive Care Units, that may prevent the operation of cry detection algorithms from succeeding. Second, we implement a specific cry detection technique that is based on some of the relevant cry detection proposals that have been found in the literature. Finally, we test this method using audio samples recorded in a real neonatal intensive care unit.

   Keywords: Robustness; Cry Detection Methods; Neonatal Intensive Care Units

Introduction

   Some newborns have to spend a period of time in Neonatal Intensive Care Units (NICUs). Most of them are preterm newborns who suffer from some immaturity or disease. During hospitalization, newborns might suffer from pain (for instance, due to invasive procedures) and if this pain is not treated correctly, the child could suffer from some neurological disorders in the future. In order to detect whether neonatal patients are suffering from pain or discomfort, health professionals proceed by directly observing physiological parameters (i.e. heartbeat rate, temperature, etc.) and behavioral parameters (frowning, crying…) and using some assessment scales to figure out if some pain must be relieved. These procedures are done in an hourly basis, by a reduced number of people that also take care of the other typical duties in the NICUs.

   Moreover, current trends in nursing for the neonates focus on personalized observations of the newborn (for instance, the progressively adopted NIDCAP, Newborn Individualized Developmental Care and Assessment Program [1]).

   In order to improve the quality in this aspect of healthcare, some attempts to automatize the detection of neonate's cry have been published. Accurately characterizing the cry in real time could be used in both automatic pain assessment [2] and NIDCAP procedures. In fact, if some neonate in a large NICUs is crying, it could be useful to trigger alarms that request the attention of the nurses.

   Regarding methods and techniques to analyze baby's cries, the literature deals with "neat environment tests": only cry sounds are analyzed and, to the sake of our knowledge, proposals do not consider the background noise. This topic attracts the attention of some researches, merely on measuring typical sounds and noises in NICUs and how they affect the well-being of the neonates. In our work, we took a look at papers to get the brief knowledge of background noise in NICUs environments.

Background Noise in Neonatal Intensive Care Units

For cry analysis in NICUs, background noise is an important aspect which is taken into account. Background noise can come from several sources, namely equipment or machines in hospital rooms, ventilator systems, sounds generated when interacting with incubators, staff talking, phone ring, alarm noise, etc. There are some standards about noise level, for example the standard of American Academy of Pediatrics. If the noise level inside NICUs exceeds these standards for a long period, it will cause some problems on newborns, it can affect neurodevelopment or other important aspects.

In [3] authors describe these environments in terms of background noise, in a recent contribution. Measures were conducted in both an old and a new hospital. Their sound environment was surveyed for 24-hour periods in Melbourne and compared to Australian recommendation, which states maximum and minimum sound levels. For instance, the maximum sound level should be less than 65 dB and no more than 6 minutes in an hour. They also consider sound peaks (defined as jumps of more than 12 decibels). The granularity for noise measuring is 5 seconds, for which averages of decibels are computed. The results show that in general, that recommendation is not achieved in a half of observation periods and the results are not different from the old building to the new one.

Moreover, in [4] authors assess background noise according to American Academy of Pediatrics, which recommends that sound levels should not exceed the maximum acceptable level of 45 dB. Authors concluded that the sound environment in the NICUs is louder than most home or office environments and contains disturbing noises of short duration and at irregular intervals. Elevated levels of speech are needed to overcome the noisy environment in the NICU, thereby increasing the negative impacts on staff, newborns, and their families. High noise levels are associated with an increased rate of errors and accidents, leading to decreased performance among staff. The aim of interventions included in this review is to reduce sound levels to 45 dB or less. This can be achieved by lowering the sound levels in an entire unit, treating the infant in a section of a NICU (i.e. in a "private" room), or in incubators in which the sound levels are controlled, or reducing the sound levels that reaches the individual infant by using earmuffs or earplugs.

Finally, in [5] authors describe background noise both quantitatively (measuring sound) and qualitatively (by means of interviews). They compare data collected during day and night periods and conclude that there is no statistically significant difference.

Contribution and Plan of this Paper

In this paper we analyze the robustness of a cry detection method against real-life sounds that occur in a Neonatal Intensive Care Unit. To that end, we implement a method, described in Section 2, that is representative of the most outstanding ones found in the literature. In Section 3, we test this implementation using audio samples from a cry database, mixed with real audio samples of NICU's background noises. Finally, Section 4 concludes de paper.

A Cry Detection Technique

In this section the method implemented for cry detection is described. The goal of the method is to detect cry sounds in the waves obtained in the real neonatal intensive care unit (NICU). For the sake of simplicity, we have implemented a software that analyses the recorded waves. However, in a real deployment of the system, it would be able to analyze the sound over real-time acquired audio.

Key concepts

Cry analysis methods make use of several concepts or techniques that play a key role in the process. Sampling is the first step in analysis of audio from real environments. Sounds to be analyzed are in the form of samples of b bit resolution (usually, b = 16 for generic audio and b = 8 for voice-targeted applications). Samples are taken at a specific sampling rate (being 44.100 Hz a typical value for generic audio and 8.000 Hz the value for voice-targeted applications). Hence, sampling consists of the reduction of a continuous-time signal to a discrete-time signal.

Most audio analysis applications rely on frequency estimation. In this step, a window is applied to select a certain interval of data to analyze. Normally, a cry duration has the minimum length of 200 ms, so that the window size normally is 50 ms. After determining the window size, a set of data is collected consecutively from the sampling data set which is obtained from the sampling step. Then, a time to frequency transformation technique, such as Fast Fourier Transform, is applied.

Besides sampling and frequency estimation, the following concepts play a key role in the reviewed cry analysis techniques [6-8]:

- **Fundamental frequency:** Is the most significant frequency contribution of the analyzed audio sample. In terms of frequency analysis of periodic waves, it is the lowest frequency in the sum.

- **Short time energy (STE):** It is the average value of the samples, considering its absolute value, for a generally short number of samples. This granularity is related to the precision of the implementation.

- **Melodic shape:** It is the result of the observation of successive values of fundamental frequencies. Hence, the raising and falling of the cry sound can be characterized by analyzing the features of this melodic shape.

- **Glide:** It is a very rapid change of fundamental frequency.

Description of the Method

The method works with a WAVE audio file. We assume that it has been obtained using a high sensitivity microphone placed next to the incubator. The audio file is recorded at a sampling rate of r samples per second, b = 16 bit. The file is divided into a succession of N windows $\{w_1, w_2, ..., w_N\}$ of length S, being the latter a parameter of the system. Thus, w_i is a succession $\{s^i_1, s^i_2, ..., s^i_S\}$ of samples.

The first step consists of detecting a succession of non-silent windows. Note that several successions might be found in a file, but each succession will be analyzed individually. In order to classify a w_i into silent or non-silent, the STE is utilized. Then, if STE (w_i)>T, i.e. a specific threshold, wi is classified as non-silent, otherwise it is considered a silent window and therefore is discarder for further analysis.

The problem in the first step is related to detecting as non-silent window sound samples that belong to a near source of background noise. For instance, if a breathing machine is working next to the incubator, the microphone will record this noise but with a high STE. In order to avoid this problem and, assuming that cry frequencies are between 150 Hz and 900 Hz ([6]), we might proceed with a preprocessing. Preprocessing consists of eliminating from wi all the frequencies that according to [6] do not belong to cry. Naturally, higher frequencies pass the filtering and, as a result, this sound could be confused with cries. To that end, a second step analysis, that considers the entire succession of non-silent windows, is mandatory.

In the second step we analyze the entire succession of non-silent windows. This succession's length is L, and is variable in length. The goal of the second step is to compare the wave of the non-silent windows succession with the waves in a so-called Cry Dictionary.

The Cry Dictionary contains the wave features of several newborn cries, each with different durations, and in terms of fundamental frequencies.

Hence, in this second step, the fundamental frequencies of each window are obtained. As a result, a succession F of L fundamental frequencies $\{f_1, f_2, ..., f_L\}$ is obtained. In this step, we apply a filter algorithm to eliminate the fundamental frequency less than 150 Hz. Because the main purpose of our method is detecting cry unit in background noise environment, we only care about the fundamental frequency related to cry fundamental frequency. The fundamental frequency of background noise after filtering step can have values in the range of cry fundamental frequency, but it never has the shape of changing in frequency like cry.

Finally, the wave in the Cry Dictionary at a nearest distance to F is found. If this distance is below a specific threshold, the succession of L non-silent windows is considered a cry. For the sake of simplicity, our detection technique is based on distances.

Experimental Results

The aim of experimental results is to assess if the cry detection algorithm implemented detects sounds not related to cry (e.g. machinery, conversations…) as false cries. Hence we have tested our implementation with neonatal cries in the scenario of a real NICU.

Sound Samples for Testing

Sound samples used for testing have been generated using separate samples for background noise and newborn cries. On the one hand, it is not straightforward to obtain a representative sample of newborn cries. On the other, there is a database of neonate: The Baby Chillanto Data Base is a property of the Instituto Nacional de Astrofisica Optica y Electronica - CONACYT, Mexico and has been used in the literature to assess proposals on cry detection and characterization [9]. We have selected four samples from this database, namely "Normal_Cry_1", "Normal_Cry_2", "Pain_Cry" and "New_Cry". The three first samples have been used to build the Cry Dictionary.

Background noises have been recorded in a real NICU scenario: the neonatal unit in Hospital Universitari Joan XXIII. Samples, with a duration of 10 seconds each, were recorded during November 2016 using a high-sensitivity cardioid condenser microphone and an audio interface attached to a Mac Book Pro. Samples have been recorded by Audacity software.

We have selected the following background noise samples:

- "Plain NICU noise", background noise that includes humming of the machinery in the unit.
- "Alarm NICU noise", background noise that includes an alarm sound.
- "Conversation NICU noise", background noise that includes a conversation between adults.
- "Children NICU noise", background noise that includes some little girl speaking.
- "Telephone NICU noise", background noise that includes a telephone ringing sound and the voice of a nurse attending the call.

In order to generate the final samples that are used in the tests in this work, we have mixed cry sounds with the background noise at different proportions. Specifically, 80% cry plus 20% background and 50% cry plus 50% background. As a result, we test our algorithm with 5 x 3 x 2 = 30 samples with a duration of 10 seconds.

For each sample, we know the exact time in which cry starts and, hence, a false positive (i.e. false cry) will consist of detecting a cry outside the boundaries of the parts of the sample where we know the real cry is.

Robustness against NICU Noise

Tests have been conducted using a laptop equipped with a core i5-5200U processor and 4GB DDR3 RAM. Each sample took around 3-5 seconds to be evaluated. In this result, the accuracy (in time) is calculated by Total time of cry in detection/Total time of cry in one benchmark. Of course, we also consider these cry units are at the correct position. Tables 1, 2 and 3 show some of the results, for the sake of brevity. The list of the complete tests can be obtained from the document downloadable from the link http://smarthealthresearch.com/docs/cryanalysisreport.pdf.

For the "Normal_Cry_1" (Table 1) tests no false positives were found and, thus, our method is shown to be valid since it does not consider noises as cries. For the "Normal_Cry_2" (Table 2) the result of detecting in first situation of mixing with proportion 20 - 80% is quite good, there is no False positive in this situation. However, for the second case of mixing with proportion of 50 - 50%, there ex-

Sample	Cry Interval	Noise Interval	Non-silent Interval	Accuracy	False Positive
20_80_ Alarm_ Noise_ Normal_ Cry_1	4.0 - 4.4	1.0 - 1.2; 2.2 - 2.4	1.0 - 1.2; 2.2 - 2.4	91	No
	4.8 - 5.8	4.4 - 4.6; 5.8 - 6.0	4.0 - 4.6; 4.8 - 6.4		
	6.0 - 6.4	6.6 - 7.2; 7.4 - 7.6	6.6 - 7.2; 7.4 - 8.0		
	7.6 - 7.8	7.8 - 8.0; 8.4 - 8.8	8.4 - 8.8		
20_80_ Conver-sation_ Noise_ Normal_ Cry_1	4.0 - 4.4	1.2 - 1.4; 2.8 - 3.0	1.2 - 1.4; 2.8 - 3.0	82	No
	5.0 - 5.8	4.4 - 5.0; 5.8 - 6.0	4.0 - 6.4; 6.6 - 7.2		
	6.0 - 6.4	6.6 - 7.2; 7.4 - 7.6	7.4 - 8.0; 8.4 - 8.8		
	7.6 - 7.8	7.8 - 8.0; 8.4 - 8.8			
50_50_ Alarm_ Noise_ Normal _Cry_1	4.0 - 4.4	1.0 - 1.4	1.0- 1.4; 2.2 - 2.6	73	No
	4.8 - 5.8	2.2 - 2.6	4.0- 4.4; 4.8 - 5.8		
	6.2 - 6.4	7.0 - 7.8	6.2- 6.4; 7.0 - 7.8		
		9.4 - 9.6	9.4- 9.6		
50_50_ Tele-phone_ Noise_ Normal_ Cry_1	4.2 - 4.4	0.0 - 0.2; 0.8 - 1.0	0.0- 0.2; 0.8 - 1.0	54	No
	5.0 - 5.8	2.4 - 2.6; 4.0 - 4.2	2.4- 2.6; 4.0 - 4.4		
	6.2 - 6.4	4.8 - 5.0; 6.4 - 6.6	4.8- 5.8; 6.2 - 6.6		
		6.8 - 8.6; 9.2 - 9.4	6.8- 8.6; 9.2 - 9.4		

Table 1: Some representative results in the case of "Normal_Cry_1".

ists false positive. The minimum distance of frequency when false positive occur is 150 and this value could contribute to fine tune our detection algorithm using the Cry Dictionary and our current simple detection method.

Finally, in the case of "Pain_Cry" (Table 3) the results are optimal since no false positives have been detected.

Sample	Cry Interval	Noise Interval	Non-silent Interval	Accuracy	False Positive
20_80_ Alarm_ Noise_ Normal_ Cry_2	4.6 - 4.8	4.2 - 4.6	4.2 - 5.0	62	No
	5.2 - 5.6	4.8 - 5.0	5.2 - 6.8		
	5.8 - 6.2	5.6 - 5.8	7.0 - 7.2		
	6.4 - 6.8	6.2 - 6.4			
	7.0 - 7.2				
20_80_ Conversation_ Noise_ Normal_ Cry_2	4.2 - 4.4	0.8 - 1.0; 1.2 - 1.4	0.8 - 1.0	54	No
	5.8 - 6.8	2.8 - 3.0; 4.0 - 4.2	1.2 - 1.4		
	7.0 - 7.2	4.4 - 5.8; 6.8 - 7.0	2.8 - 3.0		
		7.2 - 7.4; 8.6 - 8.8	4.0 - 7.4		
			8.6 - 8.8		
50_50_ Alarm_ Noise_ Normal_ Cry_2	5.8 - 6.2	0.0 - 1.6	0.0 - 1.6	25	No
	7.0 - 7.2	1.8 - 5.8	1.8 - 8.2		
		6.2 - 7.0	8.4 - 10.0		
		7.2 - 8.2			
		8.4 - 10.0			
50_50_ Telephone_ Noise_ Normal _ Cry_2	5.4 - 5.6	0.2 - 0.6; 1.8 - 2.0	0.2 - 0.6; 1.8 - 2.0	35	Yes
	5.8 - 6.0	2.4 - 3.2; 3.4 - 4.0	2.4 - 3.2; 3.4 - 4.0		
	6.2 - 6.6	4.2 - 5.0; 5.2 - 5.4	4.2 - 5.0; 5.2 - 7.2		
	7.0 - 7.2	5.6 - 5.8; 6.0 - 6.2	7.4 - 7.6; 8.0 - 8.4		
	8.2 - 8.4	6.6 - 7.0; 7.4 - 7.6	8.6 - 9.2		
	8.6 - 8.8	8.0 - 8.2; 8.8 - 9.2			

Table 2: Some representative results in the case of "Normal_Cry_2".

Detection of a Cry not in the Dictionary

Until now, we have described the success of our method in terms of not detecting false positives. In addition, we consider assessing the validity and robustness of our method about detecting new cries that are not in the dictionary. To that end, we have tested our method using the fourth sample from the database, and creating mixes with background noise from the real NICU. Table 4 shows that out method gives does not detect false positives and detects the new cry as well.

Discussion

In most of the cases, our method can correctly detect cry regardless the background noise, even in case of mixing 50% of noise. False positive only exists in some tests of second cases with 50 - 50% mixing noise.

In case of "Normal_Cry_1" and "Pain_Cry", which we already know are the good cry recording, the quality of the method is correct. It gives very high accuracy in case of 20 - 80% of mixing noise. In "Normal_Cry_1" case, it always gives the accuracy more than 80%. In case of "Pain_Cry", the result is even better, around 100% of accuracy.

In case of "Normal_Cry_2", we knew that this file -i.e. the sample from the cry database-, it also contains a noticeable amount of background noise. Thus the detecting result is not good as the others. In situation of 20 - 80% mixing background noise, it only gives the result with accuracy around 60%, but in situation of 50 - 50% mixing background noise, the result is even worse and the false positive appears. It is not good enough to become the standard for detecting cry. However, the value of minimum distance of frequency obtained is very valuable for estimating the standard fundamental frequency threshold in the future.

Sample	Cry Interval	Noise Interval	Voiced Interval	Accuracy	False Positive
20_80_	4.0 - 5.2	1.0 - 1.4; 2.2 - 2.6	1.0 - 1.4	100	No
Alarm_	5.8 - 6.6	5.2 - 5.8; 6.6 - 7.0	2.2 - 2.6		
	7.0 - 7.8	7.8 - 8.0	4.0 - 8.0		
Noise_					
Normal_Cry_1					
20_80_	4.0 - 5.2	0.8 - 1.0; 1.2 - 1.4	0.8 - 1.0	100	No
Conversation_	5.8 - 6.6	2.8 - 3.0; 5.2 - 5.8	1.2 - 1.4		
	7.0 - 7.8	6.6 - 7.0; 7.8 - 8.0	2.8 - 3.0		
Noise_		8.6 - 8.8	4.0 - 8.0		
Normal_Cry_1			8.6 - 8.8		
50_50_	4.2 - 5.2	1.0 - 1.4; 2.2 - 2.6	1.0 - 1.4; 2.2 - 2.6	86	No
Alarm_	5.8 - 6.6	4.0 - 4.2; 5.4 - 5.6	4.0 - 5.2; 5.4 - 5.6		
	7.0 - 7.6	6.6 - 7.0; 7.6 - 7.8	5.8 - 7.8		
Noise_					
Normal_Cry_2					
50_50_	4.0 - 4.6	2.8 - 3.0	2.8 - 3.0	72	No
Conversation_	5.8 - 6.6	4.6 - 5.2	4.0 - 5.2		
	7.0 - 7.6	5.4 - 5.6	5.4 - 5.6		
Noise_		7.6 - 7.8	5.8 - 6.6		
Normal_Cry_2			7.0 - 7.8		

Table 3: Some representative results in the case of "Pain_Cry".

Sample	Cry Interval	Noise Interval	Non-silent Interval	False Positive
NewCry_	4.2 - 6.0	4.0 - 4.2; 6.0 - 6.4	4.0 - 7.4	No
Plain	6.4 - 7.2	7.2-7.4; 7.6-7.8	7.6 - 7.8	
_Noise		8.0-8.2	8.0 - 8.2	
NewCry_	4.0 - 4.6	2.8 - 3.0; 4.6 - 4.8	2.8-3.0; 4.0-5.8	No
Telephone	4.8 - 5.8	6.0 - 6.2; 7.2 - 7.4	6.0-7.4; 7.6-7.8	
_Noise	6.2 - 7.2	7.6 - 7.8; 8.0 - 8.2	8.0-8.2; 8.8-9.0	
		8.8 - 9.0		

Table 4: Results in the case of "New_Cry".

The result in the situation of mixing 20 - 80% mixing background noise is always better than the result in the situation of mixing 50 - 50%. In case of "Normal_Cry_1" and "Pain_Cry", situation of 20 - 80% gives result of over 80% and nearly 100% respectively. But the situation of 50 - 50% is only gives result of average 60% and 80% respectively

In case of "New Cry", our method gives a good result of detecting cry. There is no false positive in this case even the amplitude of noise is quite high. Although it still cannot detect all the cry duration in the audio file, it can detect the most important cry signals correctly.

Conclusions and Future Work

In this paper we have addresses the robustness of cry detection techniques against real-life neonatal Intensive Care Units. Our method has correctly detected cry regardless background noise. In our work, we determined that fundamental frequency is the most important feature of newborn infant cry. Thus, we focused on characterizing the feature of the changing and the shape in fundamental frequency. Comparing to the other method, it could reduce the computational cost but still gives us the good result.

Although we get good result in detecting cry unit in most of the case, but there still exists some false positives. The reason is insufficient quality of Cry Dictionary. The cry sample with the name "Normal Cry 2" containing noise inside, when creating standard cry unit from it, it reduces the quality of the Cry Dictionary. It lets to some false positives in detecting later. In future, we will investigate more about building the good and standard Cry Dictionary and make it to be a good benchmark for other researches.

In our method, the value of some importance parameters which are power threshold, minimum distance in fundamental frequency are still not fixed. Because we do not have the standard scenario for recording cry, the cry samples are mixed by other audio software. In near future, we will design a standard scenario, the device for recording cry of the newborn infant will be put near the incubator to warranty that the background should have the lower energy than the cry.

In this work, we just implemented the basic method to classifying cry unit based on calculating the similarity in the characteristic of fundamental frequency. It can be still improved because fundamental frequency is not the unique characteristic of cry unit. In the future, we will implement artificial intelligent method like Neural Network, Supported Vector Machine, etc. to classify cry unit based on the set of input attributes including fundamental frequency, some resonance frequencies and average power frequency ratio. The new method will take into account all the features of cry signal, thus, the quality of cry detection should be improved.

Bibliography

1. Nidcap.org, A. What is NIDCAP? – NIDCAP (2017).

2. Artinez-Allest A., *et al*. "An autonomous system to assess, display and communicate the pain level in newborns. In Medical Measurements and Applications (MeMeA)". 2014 IEEE International Symposium (2014): 1-5.

3. Shoemark H., *et al*. "Characterising the ambient sound environment for infants in intensive care wards". *Journal of Paediatrics and Child Health* 52.4 (2016): 436-440.

4. Almadhoob A and Ohlsson A. "Sound reduction management in the neonatal intensive care unit for preterm or very low birth weight infants". *Cochrane Database of Systematic Reviews* 1 (2015): CD010333.

5. Darcy AE., *et al*. "A descriptive study of noise in the neonatal intensive care unit ambient levels and perceptions of contributing factors". *Advances in Neonatal Care* 8.3 (2008): 165-175.

6. Manfredi C., *et al*. "High-resolution cry analysis in preterm newborn infants". *Medical Engineering and Physics* 31.5 (2009): 528-532.

7. Cohen R and Lavner YA. "Infant cry analysis and detection". In Electrical and Electronics Engineers in Israel (IEEEI), 2012 IEEE 27th Convention, IEEE (2012): 1-5.

8.   Díaz MAR., *et al*. "Automatic infant cry analysis for the identi-
      fication of qualitative features to help opportune diagnosis".
      Biomedical Signal Processing and Control 7.1 (2012): 43-49.

9.   Reyes-Galaviz OF., *et al*. "Evolutionary-neural system to clas-
      sify infant cry units for pathologies identification in recently
      born babies". In Artificial Intelligence, 2008. MICAI'08. Sev-
      enth Mexican International Conference. IEEE (2008): 330-
      335.

Volume 1 Issue 3 October 2018