ANALYSIS OF ACOUSTIC FEATURES OF INFANT CRY FOR CLASSIFICATION PURPOSES

Ali Messaoud , Chakib Tadj

École de Technologie Supérieure,
1100, Rue Notre-Dame Ouest, Montréal, Canada

ABSTRACT

The infant's cry is a vocal signal that was proven to be a way of communication that aims to get the attention of the listener to a given physical state like hunger or pain. This work deals with the analysis of the acoustic features of the infant cry offering a starting point for the development of an automatic classification system. Standard features such as pitch and intensity were analyzed. A sample-based and a sequence-based studies were carried out in order to discover special properties of the infant cry. The preliminary results delivered helpful information about acoustic statistics and correlation between cry features.

Index Terms— Infant cry, acoustic features, pitch detection, fundamental frequency.

1. INTRODUCTION

The analysis of biological signals has always been an essential mean for detecting pathologies and recognizing particular physiological and neurological states of the human body. A common biological signal is the infant cry, which, according to specialists, could contain valuable information about the state of the baby including hunger, pain [1], physiological abnormalities or special gestational conditions [2].

Experienced parents can achieve the perception of the infant cry acoustic characteristics. Nevertheless, it can be altered by individual or contextual factors [3]. Consequently, many researchers have been developing alternative solutions for infant cry perception. The main goal of this work is to set up a starting point for the future design of an automatic recognition system that would distinguish different kind of infant cries based on the extracted acoustic features. Thus, the first step in this project is the identification of these features.

This work deals mainly with the pitch i.e. the fundamental frequency as in almost all the studies that dealt with infant cry [4-5]. The Harmonic Product Spectrum algorithm was used for tracking the pitch of the cry signal. Besides, the intensity of the cries was used as an essential feature. In previous studies, features extracted from individual cry samples characterized the infant cry. Therefore, the time factor was totally ignored.

In order to overcome this problem, we explored a new approach involving the evolution of the "standard" features throughout a whole sequence of cries. Another important issue is the length of the segmented cry samples. In many studies, cry segments of fixed length were extracted from the cry records although the cry utterances are not equal in length [6-7]. This causes the samples that are longer than the chosen length to be truncated. This can be handled by performing a segmentation that takes into account the full length of the cry utterances thus producing complete cry samples with different lengths.

The results presented in this paper were obtained by the mean of two studies: first, a sample-based study provided statistics and correlation of acoustic cry features. Secondly, the evolution of the acoustic features in a cry sequence was investigated in a sequence-based study.

2. ANALYSIS METHODOLOGY

2.1. Data set

The cry samples used in this study were collected from 8 healthy babies aged less than 6 months regardless of gender, using the WS-310M Olympus digital voice recorder at a sampling frequency of 44100 Hz and a sample resolution of 16 bits. The context in which the babies were recorded was "hunger". Many recording sessions were necessary to compensate the low number of available subjects. A set of 20 recordings was obtained.

A noise reduction procedure was then applied to the raw cry records using Adobe Audition software to eliminate the background noise picked up by the recording device. Then, the manual segmentation of the cry records was achieved resulting in a data set formed by 1962 individual cry samples.

CCECE 2011 Niagara Falls, Canada
978-1-4244-9789-8/11/$26 ©2011 IEEE

2.2. Pitch tracking

In 2007, Varallyay *et al*. conducted a sudy about the melody of the infant's cry [8]. The study was based on the fundamental frequency which is an important feature in categorizing an infant cry. It can be extracted from the cry sample either as a statistical (average) or a dynamically varying value as illustrated by Fig. 1. The former does not consider the time factor whereas the latter is more appropriate because it offers the possibility to trace the changes of pitch over time and discover special time-related features in the obtained pitch curve.

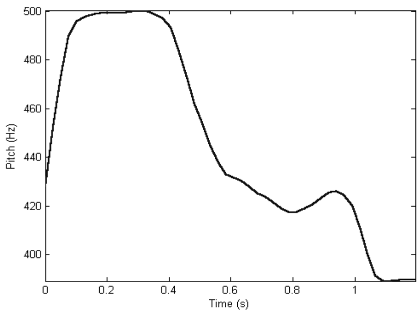

Figure 1. Example of a pitch curve.

2.2.1. Preprocessing

Before the determination of the pitch curve, each cry sample was preprocessed first, by removing the "silent" portions from the edges i.e. having a value of power below 1 % of the maximum power and then, by applying a 4th order low pass digital filter with a cutoff frequency of 3000 Hz.

2.2.2. Harmonic product spectrum

In order to determine the pitch curve i.e. the value of the fundamental frequency as a function of time, each cry sample was segmented in windows of 50 ms without overlapping [7]. The Hanning windowing function was used for this purpose. Then, the pitch of each window or frame was calculated using the Harmonic Product Spectrum (HPS) algorithm, which is based on the Fast Fourier Transform.
The HPS algorithm measures the coincidence for harmonics according to (1) for each spectral frame, X(ω).

$$Y(\omega) = \prod_{r=1}^{R} |X(\omega r)| \quad (1)$$

$$\hat{Y} = \max_{\omega_i}\{Y(\omega_i)\} \quad (2)$$

where R is the number of harmonics to be considered (for example R=3), and frequency ω_i is in the range of possible fundamental frequencies. The resulting periodic correlation array, Y(ω), is searched for a maximum value, $\hat{Y}$, as is shown in (2) [9].
Octave errors are a common problem in pitch detection from HPS. Almost always in these error cases, the pitch is detected one octave too high. To correct this error we proceed as follows: if there is another peak that corresponds approximately to half of the detected pitch and the ratio of amplitudes is above a threshold (e.g. 0.2), then select the lower octave peak as the pitch for the current frame [6]. To avoid low octave errors as well, a minimum frequency below which the spectrum is not scanned was considered.

2.2.3. Pitch curve correction

The resulting pitch curves of the cry samples are in their raw form and contain octave errors. In order to correct these errors, out of range points (beyond typical pitch values) of each curve were excluded before applying a nearest neighbor interpolation using the remaining points. The curves having more than 50 % of excluded points (frames) or reaching a maximum pitch that exceeds 600 Hz were rejected. Due to this selection technique, not all of the 1962 collected cry samples were used in the study. After applying this procedure, a set of 1449 samples was accepted. A second interpolation using a smoothing spline method was applied to make the pitch curves smoother.

2.3. Power tracking

The power tracking consists in representing the value of the power as a function of time. The extraction of power curves was done from the pre-processed cry samples. The obtained curves were smoothed by a smoothing spline interpolation. Figure 2 shows an example of a power curve.

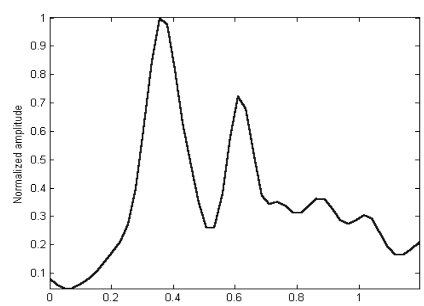

Figure 2. Example of a power curve.

The power of a signal x(k) is expressed by (3).

$$P = \frac{1}{n}\sum_{k=1}^{n} x^2(k) \quad (3)$$

3. EXPERIMENTAL RESULTS

The programming environment Matlab 6.5 including the Signal Processing Toolbox 6.0, the Curve Fitting Toolbox 1.1 and the Statistics Toolbox 4.0 was used for the implementation of the infant cry analysis functions and the visualization of the results. The experimental results were divided in two types of analysis approaches: a sample-based study and a sequence-based study.

3.1. Cry sample statistics

Table 1 presents the mean value and the standard deviation of the following cry features: length of the samples, pitch (F_0), minimum and maximum pitch in a pitch curve, pitch rise time (i.e. time to reach the maximum value of pitch) expressed in percentage of the sample length.

Table 1. Cry statistics.

Feature	Unit	Mean	Std. Dev.
Length	Sec	1.09	0.47
F_0	Hz	411.35	62.31
$F_{0\ min}$	Hz	363.23	58.25
$F_{0\ max}$	Hz	468.71	52.23
$F_{0\ rise\ time}$	%	50.56	24.97

3.2. Correlation between pitch and power curves

From the representation of normalized pitch and power curves of a given cry sample in the same graphic, a relation between them can be easily seen. Examples of curves from 2 cry samples are presented in Fig. 3.
The correlation coefficients of pitch and power vectors (curves) were calculated for each cry sample, then, a histogram is represented in order to inspect the characteristics of the distribution of the calculated correlation values as shown in Fig. 4.

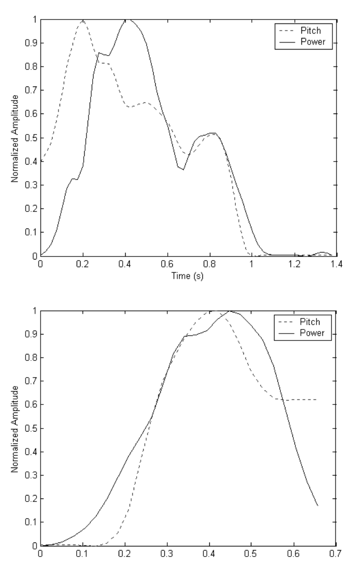

Figure 3. Pitch and power curves.

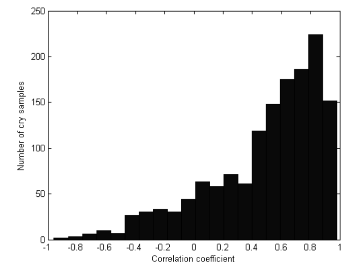

Figure 4. Distribution of correlation coefficients.

3.3. Acoustic features in a cry sequence

As the individual cry samples were originally part of cry sequences, they naturally should be studied as a whole unit in which cry features change from one sample to the next i.e. as a function of time. Figure 5 is an example that shows a bi-dimensional graphical representation of the pitch and power curves of the cry sequence N°2. Curves are aligned vertically forming neighboring slices. The color bar on the right side of the plots represents the pitch and power values. This representation shows the chronological evolution of time-related cry features such as the pitch rise time as well as the duration of the samples. The difference between the rise time of pitch and that of power (the unit is percent of the cry sample length) gives an idea about the "distance" between the maximum values of both features. The absolute mean values of rise time difference of the 20 cry sequences are gathered in Table 2.

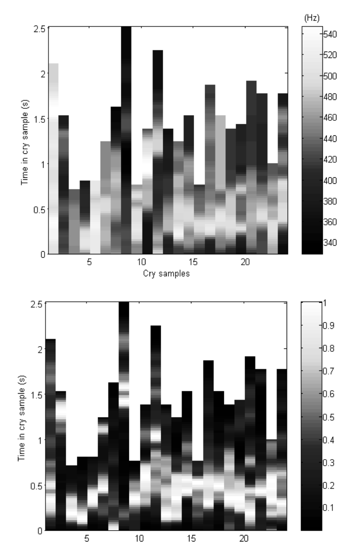

Figure 5. Pitch (*up*) and power (*bottom*) curves of sequence 2.

Table 2. Absolute mean rise time difference.

Seq.	1	2	3	4	5	6	7	8	9	10
Diff.	0.83	0.1	3.99	0.11	1.06	7.42	4.88	11	0.78	14.7
Seq.	11	12	13	14	15	16	17	18	19	20
Diff.	3.36	8.56	2.2	6.92	9.54	0.95	3.95	7	0.43	3.44

4. DISCUSSION

The collected cry database was partly used in this study because of the errors produced in the pitch detecting stage. The selection procedure excluded 26.1 % of the cry samples due to their erroneous pitch curves. Consequently, the pitch tracking algorithm and the pitch correction methods used in this study are rather efficient if we consider the rejection rate that we obtained.

As for the cry statistics, we can see in Table 1 that the pitch (F0) of a cry utterance varies from 363.23 Hz to 468.71 Hz with a mean value of 411.35 Hz. A value beyond that range could be considered as a clue for an abnormal type of crying. We can also deduct from the value of the pitch rise time that the absolute peak of the pitch curve occurs around the middle of a cry sample. Statistically, we found that 70 % of the cry samples had their pitch rise time in the range 25 % to 75 %.

In section 3.2, we studied the correlation between pitch and power. The distribution of the correlation coefficients presented in Fig. 4 demonstrates that the pitch and the power curves of a cry sample are highly correlated. Numerically, we found that in 60 % of the total number of samples, the correlation value exceeds 0.5.

In the sequence-based study, the example of Fig. 5 shows clearly that the high values of pitch and those of power (clear shades) are spread on almost the same regions. We can see in Fig. 6 that the difference between the rise time of pitch and power oscillates around a mean value of – 0.1 (sequence N° 2). The maximum difference is relative to the cry sequence N° 10 with a value of 14.7 %. Accordingly, based on the obtained delay values of the 20 sequences, we can say that pitch and power cry features are rather correlated.

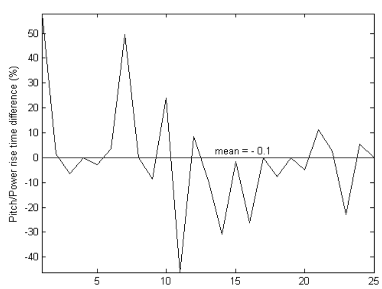

Figure 6. Pitch/Power rise time difference of sequence 2.

5. CONCLUSION

This work was devoted to the analysis of infant cry signals, which are thought to convey useful information about the physiological and neurological state of the baby. The study was mainly based on the analysis of standard acoustic features such as the pitch and the power of the cry signals and the relation between them.

The cry data set, collected from 8 healthy babies, underwent a selection process in order to reject the cry samples having erroneous pitch curves. In the sample-based study, cry statistics were gathered from a set of 1449 selected samples out of 1962. The obtained typical values can be very useful for future classification systems. Besides, we found that the pitch and the power of a cry utterance are highly correlated. This correlation was also demonstrated in the sequence-based study.

As a future work, we consider collecting more cry samples and implementing a more efficient pitch-detecting algorithm as well as undertaking experiments on other acoustic features using other techniques of signal processing.

Aknowlegements. We acknowledge the funding by the Natural Sciences and Engineering Research Council (NSERC) of Canada and the partial funding by Bill and Melinda Gates Foundation.

6. REFERENCES

[1] A. Branco, *et al.*, "The newborn pain cry: Descriptive acoustic spectrographic analysis," *International Journal of Pediatric Otorhinolaryngology,* vol. 71, pp. 539-546, 2007.

[2] Z. L. Quick, *et al.*, "Acoustic cry characteristics of infants exposed to methadone during pregnancy," *Acta Pædiatrica,* vol. 98, pp. 74-79, 2009.

[3] L. L. LaGasse, *et al.*, "Assessment of infant cry: Acoustic cry analysis and parental perception," *Mental Retardation and Developmental Disabilities Research Reviews,* vol. 11, pp. 83-93, 2005.

[4] S. D. C. Ortiz, "A radial basis function network oriented for infant cry classification," *Lecture notes in computer science,* vol. 3287, pp. 374-380, 2004.

[5] A. Protopapas, "Perceptual differences in infant cries revealed by modifications of acoustic features," *The Journal of the Acoustical Society of America,* vol. 102, pp. 3723-3734, 1997.

[6] O. F. R. Galaviz, "Infant cry classification to identify hypo acoustics and asphyxia comparing an evolutionary-neural system with a neural network system," *Lecture notes in computer science,* vol. 3789, pp. 949-958, 2005.

[7] J. Orozco and C. A. R. Garcia, "Detecting pathologies from infant cry applying scaled conjugate gradient neural networks," presented at the Proc. European Symposium on Artificial Neural Networks, Bruges-Belgium, 2003.

[8] G. Varallyay and Z. Benyo, "Melody Shape - A Suggested Novel Attribute for the Biomedical Analysis of the Infant Cry," presented at the Engineering in Medicine and Biology Society. 29th Annual International Conference of the IEEE, 2007.

[9] P. d. l. Cuadra, *et al.*, "Efficient pitch detection techniques for interactive music," presented at the Proc. International Computer Music Conference, Havana-Cuba, 2001.