WAVELET ANALYSIS OF NEWBORN INFANT CRY

S. Orlandi[1], A. Bandini[1,2], A. Perrella[1], J. Marjouee[1], G.P. Donzelli[3], C. Manfredi[1]

[1]Department of Information Engineering, Università degli Studi di Firenze, Firenze, Italy
[2]Department of Electrical, Electronic and Information Engineering (DEI) "Guglielmo Marconi", Università di Bologna, Bologna, Italy
[3]Department of Neuroscience, Psychology, Medicines and Child Health, Università degli Studi di Firenze, Firenze
{silvia.orlandi, andrea.bandini, gianpaolo.donzelli, claudia.manfredi}@unifi.it

Abstract: **The acoustical analysis of the infant cry is a non-invasive approach to assist the clinical specialist in the detection of abnormalities in infants with possible neurological disorders. Along with the perceptual analysis, the automatic analysis of the cry is often carried out through commercial or free software tools. However, the neonatal cry is a signal extremely difficult to analyze with standard techniques due to its quasi-stationarity and the very high range of frequencies of interest. To address this issue, we present a new fully automatic method that exploits the wavelets high time-frequency resolution and low computing time properties for the estimation of the fundamental frequency F_0 and vocal tract resonance frequencies F_1-F_3. The method is tested on synthetic signals giving results comparable to existing tools. It is also applied to a set of 1669 newborn cry units (CU) coming from 10 very preterm babies and to a set of 3514 CUs of 20 full-term infants.**

Keywords : **acoustical analysis, wavelet transform, newborn infant cry, fundamental frequency, resonance frequencies**

I. INTRODUCTION

The acoustical analysis of the infant cry is a non-invasive approach to assist the clinical specialist in the detection of abnormalities in infants with possible neurological disorders. A brain dysfunction may lead to disorders in the vibration of the vocal folds and in the coordination of the larynx, pharynx and vocal tract.

The main parameters of the newborn cry are the fundamental frequency (F_0), the frequency of vibration of the vocal folds, and the first three resonance frequencies (RFs) of the vocal tract (F_1, F_2 and F_3) related to the varying shape of the vocal tract during the vocal emission. Indeed, in the newborn it is more appropriate to refer to resonance frequencies (RFs) rather than formants. In fact, the vocal tract is almost flat, the mobility of the oral cavity is reduced and the baby is unable to articulate vowel or consonant sounds, as the pharynx is too short and not wide enough for that purpose. For infants F_0 values are usually in the range 200 Hz - 800 Hz (in the case of hyperphonation they can reach and exceed 1000 Hz) [1-2]. Typical values for the first three RFs are approximately 1000 Hz, 3000 Hz and 5000 Hz [3]. Significant deviations from these ranges may be related to pathological conditions of the central nervous system.

The study of neonatal cry has its origins several decades ago, when the technology was limited and it was therefore mainly based on the perceptual analysis made by the clinician through listening to the cry signal and visually analyzing the recorded signal and its FFT-based spectrogram [4]. This approach is implemented in the MDVP™, the first and still used commercial tool, though developed for adult voices [5]. Currently, many researchers use PRAAT [6, 7] freely available on line. As MDVP, it was developed for the adult's voice and requires a careful manual setting of some parameters [7]. In the last years, a fully automatic adaptive parametric approach for the crying analysis was developed, named BioVoice [8, 9]. As for F_0, the difficulty in the estimation of the RFs is mainly linked to the quasi-stationarity and the very high range of frequencies of interest in the newborn cry, which requires sophisticated adaptive numerical techniques characterized by high time-frequency resolution.

To overcome such problems, this paper presents a new fully automatic method based on wavelet transforms specifically developed for the estimation of F_0 and the RFs of newborn cry that does not require any manual setting to be made by the user. The wavelet approach seems particularly suited to the study of neonatal cry thanks to its time-frequency high resolution characteristics and low computing time. In [10] a Continuous Wavelet Transform (CWT) with the Mexican hat was used on adult voice signals and in [11] the complex Morlet mother wavelet were applied.

Claudia Manfredi (edited by), *Models and analysis of vocal emissions for biomedical applications : 9 th international workshop : September 2-4, 2015*, ISBN 978-88-6655-792-0 (print) ISBN 978-88-6655-793-7 (online)

This paper presents the first attempt to apply wavelets to the analysis of newborn cry. The implemented approach, named InCA (Infant Cry Analizer) is currently implemented in MATLAB, but it is easily adaptable for any embedded processor. InCA is tested and compared on synthetic signals with BioVoice and PRAAT. Results are comparable as far as F_0, F_1 and F_2 are concerned while F_3 is slightly overestimated. InCA is also applied to a set of newborn cries coming from 10 preterm infants and 20 full-term infants for a total of 5183 Cry-Units (CU).

II. Methods

A. Pre-processing

The analyzed signals, both simulated and real, were sampled at 44100 Hz and the time duration of the analysis window was chosen equal to 10 ms (441 samples). As compared to the use of longer windows, that might not take into account the variability of the signal, this leads to improved accuracy of the estimates.

The next step is the detection of the vocalic parts of the signal (the so-called "crying episodes" or Cry Units - CU) where F_0 and RFs are estimated. For the selection of CUs, the proposed approach takes advantage of the procedure developed in BioVoice whose higher robustness with respect to other software tools has been demonstrated [9].

An audio recording of crying usually includes several CUs. In the literature, different time lengths are considered for CUs, ranging from 60 to 500 ms [12]. However, CUs of very short duration do not allow the assessment of some relevant features such as their melodic shape. Moreover, inspiratory sounds that have duration less than 200 ms must be disregarded [12]. For these reasons, audio analysis is performed here on CUs longer than 260 ms.

B. Continuous wavelet transform

The wavelet transform filters a signal *f*(t) with a shifted and scaled version of a prototype function ψ(t), the so-called “mother wavelet”, a continuous function in both the time domain and the frequency domain [13].

The scale parameter *a* of a Continuous Wavelet Transform (CWT) is related to the width of the analysis window: it either dilates or compresses the signal. The shift parameter *b* locates the wavelet in time. Varying *a* and *b* allows locating the wavelet at the desired frequency and time instant [13]. The relationship between *a* and the frequency is given by the so-called pseudo-frequency (F_a) in Hz, defined by the following equation:

$$F_a = \frac{F_c}{a\Delta} \quad (1)$$

where Δ is the sampling period, and F_c is the wavelet central frequency.

For F_0 estimation, a Mexican Hat CWT is used.

For each time window and in the frequency band of interest for F_0 [200-800], the highest coefficient of the CWT matrix is found. The autocorrelation (AC) is computed on the row of the matrix that contains this value, which corresponds to the optimal scale. F_0 is given by:

$$F_0 = F_s / \tau \quad (2)$$

Where τ refers to the position (lag) of the maximum of the AC.

The estimation of $F_1 – F_3$ is performed in a similar way, with different ranges for the band-pass filter as reported in Table 1 with a complex Morlet wavelet as prototype [13]. For this wavelet is defined:

$$\omega_c = 2\pi F_c \quad (3)$$

where ω_c as the center frequency of the wavelet; σ_t is the standard deviation (STD), that is the scale parameter which determines the amplitude of the wavelet. In fact $\omega_c\sigma_t$ sets the link between the bandwidth of the wavelet and its frequency F_c. For the Morlet wavelet, the latter must assume values such that [10, 11]:

$$\omega_c\sigma_t \geq 5 \quad (4)$$

Moreover, the following relationship is taken into account:

$$F_b = 2{\sigma_t}^2 \quad (5)$$

Where F_b is the bandwidth of the wavelet. Comparing the frequency ranges and on analogy to [10] the values of F_c and the corresponding values of F_b were set as in Table 1. Specifically, for each F_c relative to each frequency band, F_b was computed with $\omega_c\sigma_t = 5$ and according to Eq. (3) and (5).

Table 1. Frequency bands of interest in newborn cry, center frequency Fc and bandwidth Fb for the complex Morlet

Frequency band [Hz]	F_c [Hz]	F_b [Hz]
F_1 [800 - 2100]	0.8	1.98
F_2 [1500 - 3500]	0.75	2.25
F_3 [3400 - 5500]	1.5	0.56

C. Estimation of F_0

For F_0 estimation, the proposed method involves the following steps:

1. Band-pass filtering FIR with Kaiser window [200-800]Hz;
2. Mexican Hat CWT of the signal. A pxq matrix M of coefficients is obtained, where p = maximum value of the scale and q = number of frames of the signal;

3. Location of the scale (line) of M corresponding to the coefficients of maximum modulus and estimation of F_0 according to eq.(4).

On each time window the CWT scale parameter *a* was allowed to vary in the range 1÷55. This choice is related to a reasonable frequency range for F_0: 200 Hz-1050 Hz [1]. Therefore the Mexican Hat CWT was applied with a = 55, $\Delta = 1/F_s = 1/44.1$ s, Fc = 0.25 Hz. Consequently Fa = 200 Hz according to Eq. (1).

D. Estimation of RFs

The estimation of F_1–F_3 is carried out with a procedure similar to that used for F_0 but with different ranges for the band-pass filter, according to Table 1 and Complex Morlet as mother wavelet.

III. Results

A. Synthetic signals

The method for the F_0 estimation was tested on a sine wave at 450Hz:

$$y(t) = \sin(450\ t) + e(t) \tag{8}$$

The RFs F_1-F_3 estimation method was tested on a sum of three sinusoids on analogy to [11]:

$$y(t)=5\sin(1000t)+10\sin(3000t)+15\sin(5000t)+e(t) \tag{9}$$

White noise e(t) set at 5% of the signal amplitude was superimposed through the Audacity® open source tool. Signals were sampled at Fs=44.1 kHz. Results were compared with those obtained with BioVoice and PRAAT.

BioVoice implements a robust method for the selection of the voiced parts of the signal (CUs) [9] and a variable window length for analysis: the higher the F_0 the shorter the analysis window. RFs F_1-F_3 are obtained by peak picking in a parametric PSD (AR models) whose variable order is estimated on the varying time windows previously found. Instead, PRAAT implements a method for the F_0 estimation based on the AC applied to a time window of fixed size while Linear Predictive Coding is applied for the RFs estimation. For proper use, and especially with newborn cry RFs, it requires the manual setting of some parameters. Therefore, its use must be made with caution [7]. Thus in this work the best parameters for PRAAT were preliminarily tested and set. Specifically, the range for F_0 was set at 200-800 Hz while for F_1-F_3 the maximum range was set up to 11025 (Fs/4) with the estimation of 5 formants instead of 3. The use of default values (5500 Hz and 3 formants) leads to wrong results.

To compare the three approaches for F_0 estimation a preliminary test was carried out on the sinusoid in Eq. (8). First results show that the CWT Mexican Hat allows to obtain better results than the other methods.

Table 2 shows the results obtained with the three approaches. The CWT has the best performance, as well as PRAAT (set with optimal parameters) though with a slightly higher standard deviation (STD), while BioVoice slightly underestimates F_0 (0.26%).

Table 2 – F_0 estimation. Comparison of BioVoice, PRAAT and CWT Mexican Hat on a synthetic signal (sinusoid at 450Hz with 5% white noise)

Method	**F_0 mean**	**STD**
Mexican Hat	450.00	0.00
BioVoice	448.81	2.08
Praat	450.00	0.88

On analogy to F_0, for F1-F3 a preliminary test was made with the synthetic signal in Eq.(9) with CWT Complex Morlet with 5% white noise. Table 3 shows that the CWT Complex Morlet provides good results especially for F_1 and F_2. All methods give comparable results although with significant differences on STD. BioVoice gives the best results, with the lowest STD for all RFs.

Table 3 – F1-F3 estimation. Comparison of BioVoice, PRAAT and CWT Complex Morlet

Method	**F_1 mean STD**	**F_2 mean STD**	**F_3 mean STD**
Morlet CWT	1024.09 91.00	2971.27 170.41	5163.62 411.32
BioVoice	985.47 5.12	2956.42 8.11	5050.56 11.20
Praat	1120.50 387.37	3068.69 346.26	5019.35 147.02

B. Real signals

Results concern spontaneous hunger cry of 20 full-term newborns (TN, 10 male and 10 female) and 10 preterm infants (PN, 5 male and 5 female). Gestational age (g.a.) of TN at birth was between 37 weeks and 2 days and 42 weeks; the weight was between 2400g and 4250g. Gestational age of PN at birth was between 23 weeks and 5 days and 34 weeks. The weight at birth was between 590g and 2700g. At the recording time (20-30 days after birth) the PN gestational age was between 35 weeks and 1 day and 43 weeks and 1 day; the weight ranged between 1380g and 2430g.

The TN infants were recorded within the first two days of life, while PN newborns could be recorded only about 20–30 days after birth, due to their long staying in the incubator. Specifically, the PN infants were recorded within the first 45 days after the normal end of pregnancy (37 weeks).

We collected an audio recording for each infant of at least 1 hour of duration consisting of at least 10% of crying. From the whole recording we manually selected 2 or 3 minutes of crying.
A total of 5183 CUs were extracted with BioVoice and the analysis performed with InCA. Table 4 summarizes the results.

Table 4 – Mean and STD values for F_0, F_1, F_2 and F_3 obtained with InCA.

	F_0 [Hz]	F_1 [Hz]	F_2 [Hz]	F_3 [Hz]
PN mean	481.9	1188.2	2743.6	4395.2
STD	65.2	204.0	535.5	736.0
TN mean	461.1	1090.0	3037.0	4324.2
STD	44,2	204.3	711.7	843.5

IV. DISCUSSION

In this work an innovative method named InCA, based on the wavelet transform, is presented for the study of the acoustical features of the neonatal cry. Unlike most commonly used software tools, this method has been developed specifically for this kind of signals, characterized by high fundamental frequency F_0 and quasi-stationarity.
According to a careful selection of the wavelets, tested on synthetic signals, InCA implements the CWT Mexican Hat for F_0 estimation and the CWT complex Morlet for the RFs estimation.
The computing time is comparable to PRAAT: for 1 s of recording InCA requires 0.9 s for the estimation of F_0 and 2.8 s for the estimation of RFs, against less than 0.5 and approximately 2 s respectively with PRAAT. However, the CUs obtained with PRAAT are less reliable [9] and a careful manual setting of ranges and thresholds is required to avoid meaningless results especially for RFs [7].
InCA is applied to a quite large real data set coming from preterm and full-term newborns. Results are promising. The estimated values of F_0 and RFs are in the ranges reported in the literature.
The crying of newborns and infants is a functional expression of basic biological needs, emotional or psychological conditions such as hunger, cold, pain, cramps and even joy. It requires a coordinated effort of several brain regions, mainly brainstem and limbic system and is linked to the breath system. Its characteristics reflect the development and the integrity of the central nervous system. Thus, infant cry analysis is a suitable non-invasive complementary tool to assess the physical state of infants particularly important in the case of preterm neonates. Specifically, the distinction between a regular wailing and one with anomalies is of clinical interest. Preterm infants and infants with neurological conditions may have different cry characteristics when compared to healthy full-term infant.

For this reason is important to set up an efficient method for automatic cry analysis.
An automatic method for the estimation of crying acoustical characteristics provides a support to the perceptive analysis made by the clinician reducing the required amount of time often prohibitive in daily clinical practice.

REFERENCES

[1] H. Rothganger, “Analysis of the sounds of the child in the first year of age and a comparison to the language,” *Early Hum Dev*, vol. 75, pp. 55-69, 2003.
[2] P.S. Zeskind, “Infant crying and the synchrony of arousal,” in *The Evolution of Emotional Communication: From Sounds in Nonhuman Mammals to Speech and Music in Man*. Oxford University Press, Oxford, 2013; pp. 155-174.
[3] A. Fort, A. Ismaelli, C. Manfredi, P. Bruscaglioni, “Parametric and non-parametric estimation of speech formants: application to infant cry,” *Med Eng Phys,* vol. 18(8), pp. 677-691, 1996.
[4] P. Sirviö, K. Michelsson, “Sound-spectrographic cry analysis of normal and abnormal newborn infants,” *Folia Phoniatrica et Logopaedica*, vol. 28(3), pp. 161-173, 1976.
[5] http://www.kayelemetrics.com
[6] http://www.fon.hum.uva.nl/praat
[7] P. Boersma, *Acoustic analysis, in Research methods in linguistics*, R. Podesva and D. Sharma (eds.), Cambridge University Press, 2014.
[8] C. Manfredi, L. Bocchi, S. Orlandi, L. Spaccaterra, G.P. Donzelli, “High-resolution cry analysis in preterm newborn infants,” *Med Eng Phys*, vol. 31(5), pp. 528-532, 2009.
[9] S. Orlandi, P.H. Dejonckere, J. Schoentgen, J. Lebacq, N. Rruqja, C. Manfredi, “Effective pre-processing of long term noisy audio recordings. An aid to clinical monitoring,” *Biomed Signal Proces*, vol. 8(6), pp. 799-810, 2013.
[10] L. Cnockaert, J. Schoentgen, P. Auzou, C. Ozsancak, L. Defebvre, F. Grenez, “Low-frequency vocal modulations in vowels produced by Parkinsonian subjects,” *Speech Commun,* vol. 50(4), pp. 288-300, 2003.
[11] L. Falek, A. Amrouche, L. Fergani, H. Teffahi, A. Djeradi, “Formantic Analysis of Speech Signal by Wavelet Transform,” Conf. Proc- of the World Congress on Engineering, vol. 2, pp. 1572-1576, 2011.
[12] M. A. Ruíz Díaz, C.A. Reyes García, L.C. Altamirano Robles, J.E. Xalteno Altamirano, A. Verduzco Mendoza, “Automatic infant cry analysis for the identification of qualitative features to help opportune diagnosis,” *Biomed Signal Proces*, vol. 7, pp. 43-49, 2012.
[13] Chui C. *An introduction to wavelets*, Academic Press, 1995.