Biomedical Signal Processing and Control 37 (2017) 16–22

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

Testing software tools for newborn cry analysis using synthetic signals

S. Orlandi [a,*], A. Bandini [a,b], F.F. Fiaschi [a], C. Manfredi [a]

[a] Department of Information Engineering, Università degli Studi di Firenze, Firenze, Italy
[b] Department of Electrical, Electronic and Information Engineering (DEI) "Guglielmo Marconi", Università di Bologna, Bologna, Italy

ARTICLE INFO

Article history:
Received 7 June 2016
Received in revised form
20 September 2016
Accepted 15 December 2016
Available online 21 December 2016

Keywords:
Infant cry
Acoustical analysis
Autoregressive models
Wavelet transform
Fundamental frequency
Resonance frequencies

ABSTRACT

Contactless techniques are of increasing clinical interest as they can provide advantages in terms of comfort and safety of the patient with respect to sensor-based methods. Therefore, they are particularly well suited for vulnerable patients such as newborns. Specifically the acoustical analysis of the infant cry is a contactless approach to assist the clinical specialist in the detection of abnormalities in infants with possible neurological disorders. Along with the perceptual analysis, the automated analysis of infant cry is usually performed through software tools that however might not be devoted to this specific signal. The newborn cry is a signal extremely difficult to analyze with standard techniques due to its quasi-stationarity and to very high range of frequencies of interest. Therefore software tools should be specifically set and used with caution. To address this issue three methods are tested and compared, one freely available and other two specifically built using different approaches: autoregressive adaptive models and wavelets. The three methods are compared using synthetic signals coming from a synthesizer developed for the generation of basic melodic shapes of the newborn cry. Results point out strengths and weaknesses of each method, thus suggesting their most appropriate use according to the goals of the analysis.

1. Introduction

Crying is the first and primary method of communication among humans. It is a functional expression of basic biological needs, emotional or psychological conditions such as hunger, cold, pain, cramps and even joy [1,2]. It involves activation of the central nervous system and requires a coordinated effort of several brain regions, mainly brainstem and limbic system. Therefore 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, giving rise to an abnormal cry [3,4]. To date newborn cry analysis is most often performed with a perceptive examination based on listening to the cry and visually inspecting the signal waveform and its spectrogram. This approach is operator-dependent and requires a considerable amount of time often prohibitive in daily clinical practice. An accurate and automated acoustic analysis of newborn cry could be helpful to assist the clinician to detect risk markers of neurodevelopmental disorders. Specifically, the distinction between a regular and an abnormal crying can be very useful in the clinical practice. Therefore the scientific community is paying special attention to techniques devoted to an accurate automated analysis of the newborn cry [5–7].

The most significant acoustical parameters of infant crying are the fundamental frequency (F_0) and the first three formant frequencies of the vocal tract (F_1, F_2 and F_3). F_0 reflects the regularity of the vibration of the vocal folds while F_1, F_2 and F_3 are related to the varying shape and length of the vocal tract during phonation and thus to its control [8]. Actually, it is more appropriate to refer to resonance frequencies (RFs) rather than formants in newborns. 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. Therefore hereafter we will refer to F_1-F_3 as of resonance frequencies (RFs).

Infant F_0 values are usually in the range 200 Hz–800 Hz (in the case of hyperphonation they can reach and exceed 1000 Hz) [5,9]. This range is quite wide including both healthy and pathological newborns. Indeed no precise ranges are validated in the literature for differentiating the two categories, the possible diseases or categories being of very different nature (i.e., deaf, asphyxiated, gastroschistic, preterm, etc). Typical values for the first three RFs are around 1000 Hz, 3000 Hz and 5000 Hz, respectively [10]. The RFs estimation is very difficult to perform considering their high

* Corresponding author.
E-mail address: silvia.orlandi@unifi.it (S. Orlandi).

http://dx.doi.org/10.1016/j.bspc.2016.12.012
1746-8094/

variability. Significant deviations from these ranges may be related to pathological conditions of the central nervous system [11,12].

The automated analysis of infant cry has its origins several decades ago when the technology was limited. Therefore, it was mainly based on the perceptual analysis made by clinicians through listening to the cry signal and visual inspection of spectrogram estimated with the Fast Fourier Transform (FFT) [13]. This approach is implemented in the Multidimensional Voice Program ($MDVP_{TM}$), the first and still used commercial tool, though developed for adult voices [14]. Currently, many researchers use PRAAT [15,16] freely available on line. As $MDVP_{TM}$, it was developed for the adult voice. It requires a careful manual setting of some parameters and thus some technical skills [16].

Thus, since early studies, it was highlighted the need to develop dedicated software tools that could provide automatically the main parameters of the infant cry. In the last twenty years, most of the research was devoted to F_0 estimation with traditional approaches such as FFT, correlation and cepstrum [13,17–20]. Instead, few papers addressed the RFs estimation: in some papers F1 was estimated with FFT [10,18,21,22] and recently FFT is applied for F_1-F_3 estimation [20]. In Robb et al. [22], FFT and Linear Prediction (LP) methods were compared, with results comparable only for F1.

Several approaches were tested on synthetic and real signals [7,10,11,23]. An automatic adaptive parametric approach was developed, called BioVoice [24], and successfully applied to newborn cry [7,25,26]. As well as for F_0, the difficulty in the estimation of the RFs is mainly due 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 the authors' knowledge only two software tools exist specifically designed for the automatic analysis of infant crying. The above mentioned BioVoice [7,23,25–27], based on an innovative adaptive parametric approach for F_0 and formants estimation [10] and successfully tested against other software tools [11,28] and a software tool recently proposed by Reggiannini et al. [20] that estimates F_0 by means of a cepstrum approach. [11]. It was shown that the cepstrum approach is faster than a parametric approach, but less robust against noise in frequencies (F_0 and RFs) estimation [11].

Taking into account the high variability of the signal the wavelet approach could be particularly suited to the study of neonatal cry thanks to its time-frequency varying resolution and low computing time. Therefore a new automated method based on the wavelet transform is proposed here for the estimation of F_0 and the RFs of newborn cry that, like the BioVoice tool, does not require any manual setting to be made by the user. Based on the literature the Mexican hat Continuous Wavelet Transform (CWT) is applied for F_0 estimation and the complex Morlet for RFs estimation [29,30]. The proposed approach, named WInCA (Wavelet Infant Cry Analyzer) is currently implemented in MATLAB but can be easily adapted to any embedded processor.

A new synthesizer capable to reproduce variable melodic shapes of newborn cry was developed, based on the methods described in [8,31]. This synthesizer is used to test WInCA against PRAAT and BioVoice.

This paper presents the first attempt to apply wavelets to the analysis of newborn cry, therefore the method will be described in the next section along with the new synthesizer.

The innovative features of this paper are: a synthesizer specifically developed to reproduce the melodic shape of the neonatal cry; a new method for newborn cry analysis based on the wavelet transform and a comparison of three different methods of automated analysis of cry on synthetic signals. Advantages and limits of the three methods are highlighted and the optimal use of each of them is suggested.

2. Methods

2.1. The newborn cry synthesizer

The synthesizer was developed under Matlab computing environment. It is capable of synthesizing newborn cry signals with different melody shapes. The synthesizer, based on a method developed for adult male voices [8,31] is composed by two blocks: a pulse train generator and a vocal tract filter, according to [32].

2.1.1. Pulse train generator

The pulse train generator, based on additive synthesis, builds a glottal pulse sequence. It approximates a periodic signal through a linear combination of sine waves whose frequencies are in harmonic ratio with each other. Thus, the first step of the infant cry synthesizer assumes the glottal pulse sequence as a pulse train with period T that is the reciprocal of the fundamental frequency F_0 ($F_0 = 1/T$).

The synthesis of each harmonic component is obtained through an oscillator block. Each oscillator is driven by two control vectors: the amplitude (a[n]) and the frequency (f[n]) of the output sine waves, where n is the n-th element of the control vectors ($1 \le n \le N$) and N is related to the duration (in s) of the synthesized signal and to the number of frames in which the synthesized signal is segmented. Assuming a sampling frequency $F_s = 44.1$ kHz and a frame duration of 10 ms, we get 441 samples per frame (SpF). Thus, the frame rate (the frequency of the control vector) is given by the ratio F_s/SpF (100 Hz). With a synthesized signal of 2 s of duration N = 201 samples.

In the first harmonic $a_0[n] = 1$ and $f_0[n] = F_0$. Higher harmonics are obtained with the same amplitude vector of the first harmonic ($a_0[n]$) and frequency vectors given by multiples of F_0:

$$f_k[n] = kf_0[n] \tag{1}$$

where $1 \le k \le K$ and K = 30 is the number of digital oscillators considered in the synthesizer. From the literature [5] we know that the range of F_0 for newborn cry is around 200–800 Hz and the first three resonance frequencies can reach values up to 10 kHz. Assuming an average $F_0 = 400$ Hz, with K = 30 we are able to synthesize signals with frequencies up to 12 kHz, thus covering the range of interest. Signals obtained by each oscillator are then summed up giving rise to the glottal pulse train $s_s[n]$. According to the additive synthesis described above, F_0 is kept constant throughout the whole vocal emission. Thus, to get a time varying F_0 in the synthesized signal, a modification of this approach is proposed here through the modulation of the first harmonic. Modulation is obtained driving the oscillator that generates the first harmonic with a vector $f_0[n]$ made up by variable values, as it will be discussed in subsection A3. According to Eq. (1), the other oscillators (that generate higher order harmonics) will give rise to frequency modulated signals.

2.1.2. Vocal tract filter

Vocal tract resonances are given by a filter bank with 3 parallel resonance filters [31]. Each filter is given by a 2nd-order all-pole filter with frequency response:

$$H(z) = \frac{b_1}{1 - a_2 z^{-1} + a_3 z^{-2}} \tag{2}$$

Given the bandwidth (B) of the filter, the resonance central frequency (f_c) and $\omega_c = 2\pi f_c / F_s$, the filter coefficients are obtained through the following relationships [33]:

$$r = e^{\frac{-\pi B}{F_s}} \tag{3}$$

$$a_2 = -2r\cos\left(\frac{2\pi f_c}{F_s}\right) \tag{4}$$

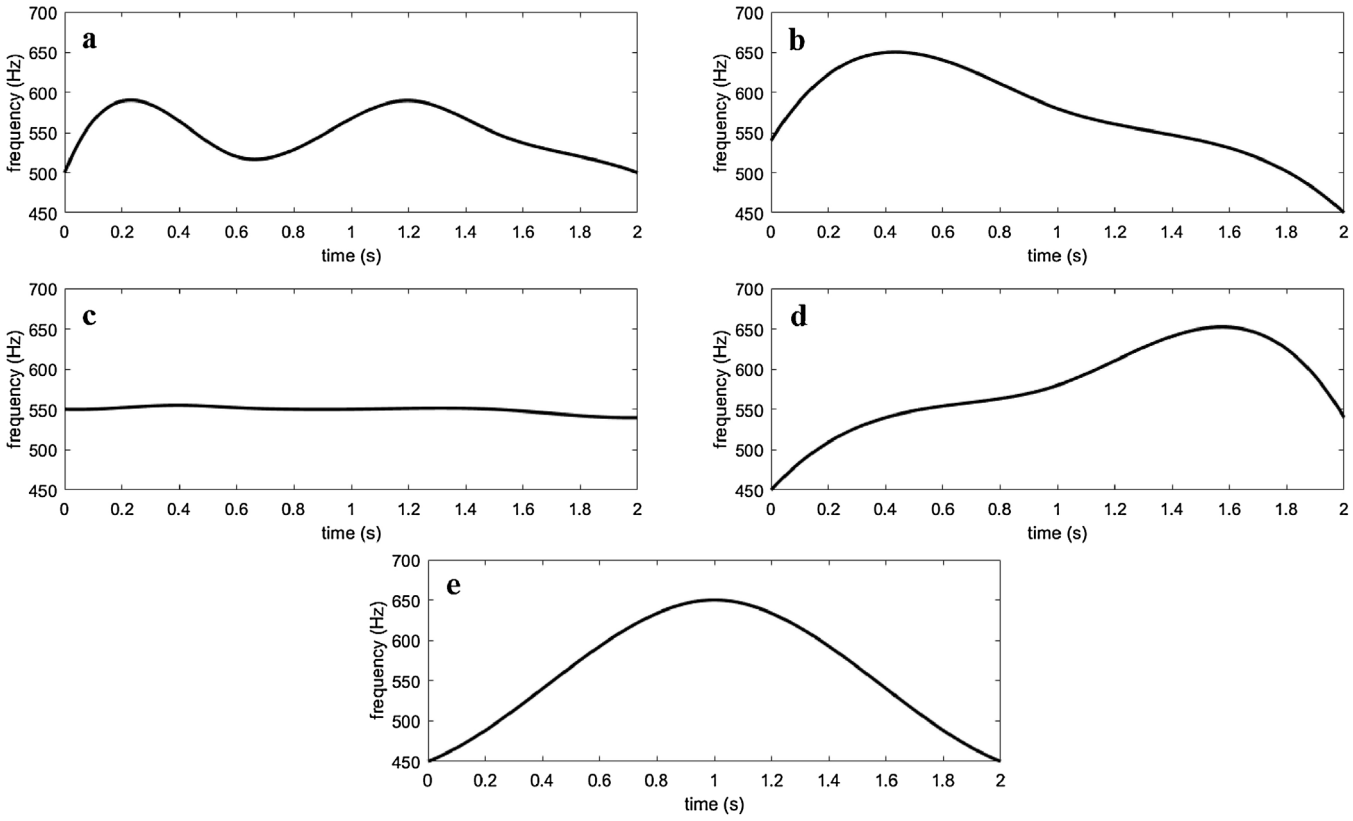

Fig. 1. Basic melody shapes of the newborn cry obtained through the newborn cry synthesizer: a) *Complex* (double-arch melody shape); b) *Falling* (single arch melody shape with a rapid F_0 increase followed by a slow decrease, with asymmetric shape skewed to the left); c) *Plateau* (flat melody profile); d) *Rising* (single-arch melody shape with a slow F_0 increase followed by a rapid decrease, with asymmetric shape skewed to the right); e) *Symmetric* (single arch melody shape almost symmetric with respect to its midpoint).

$$a_3 = r^2 \tag{5}$$

$$b_1 = (1-r)\sqrt{1-2r\cos(\omega_c)+r^2} \tag{6}$$

Thus the frequency response (2) can be written as:

$$H(Z) = \frac{b_1}{1-2r\cos(\omega_c)z^{-1}+r^2z^{-2}} \tag{7}$$

The central frequencies of the three parallel filters (f_c) correspond to the vocal tract formant frequencies (F_1–F_3). Thus, the proposed synthesizer can be controlled by setting the value of the fundamental frequency (F_0) and of the first three formant frequencies (F_1–F_3).

2.1.3. Infant cry synthesis – melody shapes and vocal tract resonances

According to [5] the fundamental frequency F_0 of healthy newborns usually varies between 200 Hz and 800 Hz. Thus, the F_0 values of the synthesized signals are allowed to vary in this range. F_0 variations were set according to the basic melody shapes of the newborn cry [23,34–37]:

- *Complex*: double-arch melody shape (Fig. 1a);
- *Falling*: single arch melody shape with a rapid F_0 increase followed by a slow decrease, with asymmetric shape skewed to the left (Fig. 1b);
- *Plateau*: flat melody profile (Fig. 1c);
- *Rising*: single arch melody shape with a slow F_0 increase followed by a rapid decrease, with asymmetric shape skewed to the right (Fig. 1d);
- *Symmetric*: single arch melody shape almost symmetric with respect to its midpoint (Fig. 1e).

As mentioned above, the synthesized signals have fixed duration equal to 2 s. Thus, the control vectors that drive the oscillators are made up by 201 samples. To get control vectors with varying frequency for synthesizing the 5 basic melody shapes of F_0, a spline interpolation was implemented. The interpolation points (nodes) were set to obtain melodic shapes within the range 400 Hz–650 Hz, which is the central interval of the range reported in [5]. Specifically, for each basic shape, the frequency control vectors are obtained as follows:

- The complex shape is defined as composed by multiple arcs (at least 2) [34]. We synthesized a double-arc shape whose maxima (with frequency equal to 590 Hz) are located at 0.25 s and 1.25 s respectively. Between these maxima, a local minimum (with frequency equal to 520 Hz) is located at 0.6 s, in order to make the second peak slower with respect to the first one (Fig. 1a). The first and last nodes of the vector to be interpolated were both set at 500 Hz
- The falling shape was obtained setting a maximum frequency (650 Hz) at 0.4 s, that allow for a slow decrease towards 450 Hz (Fig. 1b). The first node was set equal to 550 Hz
- The plateau shape was obtained by varying F_0 within a quite narrow range (540–555 Hz) around an average F_0 value equal to 550 Hz. In particular, there is a slightly ascending section up to 555 Hz, followed by a decrease towards 540 Hz (Fig. 1c).
- The rising shape was obtained by reversing the falling shape. The maximum (650 Hz) was set equal to 1.5 s, in order to get a fast

decrease towards 550 Hz (Fig. 1d). The first node was set equal to 450 Hz
- The symmetric shape was obtained setting the maximum (650 Hz) in the midpoint of the vector to be interpolated (thus at 1 s). The first and last nodes were set equal to 450 Hz (Fig. 1e).

The amplitude control vectors were obtained in the same way for all the shapes: increasing amplitude is set in the first part of the signal, followed by a nearly constant part and then decreasing amplitude at the end of the signal.

Resonance frequency ranges were set according to [38], where the first three RFs of 28 healthy term newborns were found in the range: F1 [700–1400] Hz; F2 [2000–4000] Hz; F3 [4000–7000] Hz. Specifically here we set: F_1 = 1100 Hz, F_2 = 3000 Hz and F_3 = 5500 Hz.

To test the proposed method, additive white noise of increasing amplitude (1%, 5% and 10% of the signal maximum amplitude) was added to the synthesized signals through the Audacity tool [39]. Noise was added to have more realistic melodic shapes of cries in healthy or sick infants, since it is always characterized by irregularities. The five melodic shapes described above were synthesized without noise and with each one of the three levels of added noise. Therefore twenty synthetic signals were analyzed.

2.2. WInCA: wavelet method

In this section the novel wavelet-based method for estimating F_0 and RFs is explained. The new tool named WInCA has been developed specifically for the acoustical analysis of newborn cry, characterized by high fundamental frequency F_0 and quasi-stationarity. To date no other application of the wavelet transform to newborn cry exists. Therefore, the choice of the mother wavelet is based on results obtained with adult voice signals. Specifically the Mexican Hat continuous wavelet is applied for F_0 estimation [29] and the complex Morlet for RFs estimation [30].

2.2.1. 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 [40].

The Continuous Wavelet Transform (CWT) of f(t) is defined as [40]:

$$CWT(a,b;f(t),\psi(t)) = \int_{-\infty}^{+\infty} f(t)\frac{1}{\sqrt{a}}\psi^*\left(\frac{t-b}{a}\right)dt \quad (8)$$

The scale parameter *a* of a 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 [40]. The relationship between *a* and the frequency is given by the so-called pseudo-frequency (F_a) in Hz, defined by the following equation:

$$F_a = \frac{F_c}{a\Delta} \quad (9)$$

Where Δ is the sampling period and F_c is the wavelet central frequency.

For F_0 estimation, a Mexican Hat CWT is used. The Mexican Hat CWT is defined as:

$$\psi(t) = \frac{2}{\sqrt{3}}\pi^{-\frac{1}{4}}(1-x^2)e^{-\frac{x^2}{2}} \quad (10)$$

For each time window and within the F_0 frequency range (200–800 Hz), 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:

$$F_0 = F_s/\tau \quad (11)$$

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 [40]. The CWT provides a more accurate representation of the oscillatory components within the signal without introducing fluctuations in the coefficients [40]. The complex Morlet wavelet is described by the following relationship [41]:

$$\Psi(t) = \frac{1}{\pi^{\frac{1}{4}}}e^{j\omega_c t}e^{-\frac{t^2}{2\sigma_t^2}} \quad (12)$$

where $1/\pi^{1/4}$ is the normalization term of the energy; for this wavelet the center frequency is defined as $\omega_c = 2\pi F_c$. The scale parameter σ_t is the standard deviation (STD). It 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 [29,30]:

$$\omega_c\sigma_t \geq 5 \quad (13)$$

Moreover, the following relationship is taken into account:

$$F_b = 2{\sigma_t}^2 \quad (14)$$

Where F_b is the bandwidth of the wavelet. Comparing the frequency ranges and on analogy to [29] 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 Eqs. (13) and (14).

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

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

2.2.2. Estimation of F_0

For the estimation of the fundamental frequency F_0, the proposed method involves the following steps:

1. Band-pass FIR filtering with the Kaiser window [200–800] Hz;
2. Mexican Hat CWT of the signal. A matrix M (dimensions pxq) of coefficients is obtained, where p is the maximum value of the scale and q is the number of frames on which the scaled versions of the mother wavelet are shifted. For a signal duration of 2 s and a fixed shift of 10 ms, q = 200.
3. Location of the scale (line) of M corresponding to the coefficients of maximum modulus and estimation of F_0 according to Eq. (11). An F_0 estimate is computed every 10 ms.

On each time window the CWT scale parameter *a* was allowed to vary in the range $1 \div 55$. This choice allows including the whole range of infant cry F_0 [5]. Therefore the Mexican Hat CWT was applied with $a = 55$, $\Delta = 1/F_s = 1/44.1$ s, $F_c = 0.25$ Hz. Consequently $F_a = 200$ Hz according to Eq. (9).

2.2.3. 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 fil-

Table 2
RMSE for the three methods (WInCA, PRAAT and BioVoice) applied to the 20 synthetic melodic shapes: complex, falling, rising, symmetric and plateau with no added noise, 1%, 5% and 10% added noise. Best results are highlighted in bold.

RMSE [Hz]	Noise	WInCA				PRAAT				BIOVOICE			
		F0	F1	F2	F3	F0	F1	F2	F3	F0	F1	F2	F3
Complex	0.00	6.29	55.27	227.47	600	**2.33**	50.47	82.53	1050.79	5.87	**47.77**	**65.38**	**89.35**
	0.01	6.24	55.27	227.47	600	**2.33**	50.51	82.56	1049.77	5.86	**55.62**	**77.31**	**87.91**
	0.05	6.24	55.27	226.9	600	**2.40**	50.50	82.37	1025.91	6.10	**49.99**	**79.81**	**91.70**
	0.10	6.48	55.91	203.35	600	**3.20**	51.40	82.59	1007.15	6.10	**51.06**	**76.27**	**90.74**
Falling	0.00	6.62	**96.94**	467.58	600	**2.07**	318.30	96.01	1098.48	7.02	120.14	**94.51**	**134.61**
	0.01	6.62	**96.94**	467.58	600	**2.06**	310.11	96.07	1051.83	6.97	102.55	**97.15**	**135.30**
	0.05	6.62	**96.90**	464.32	600	**2.12**	310.11	96.07	1051.83	7.30	107.67	**78.41**	**112.56**
	0.10	6.62	**96.51**	441.46	600	**2.38**	279.06	95.53	987.23	6.98	102.55	**97.15**	**135.30**
Rising	0.00	8.99	**103.43**	495.63	600	**2.15**	337.57	99.48	1096.02	3.95	119.25	**87.68**	**119.40**
	0.01	9.04	**103.43**	495.63	600	**2.15**	338.22	99.46	1093.86	3.90	105.96	**82.37**	**108.57**
	0.05	8.85	**103.76**	474.08	600	**2.13**	330.36	99.42	1076.90	4.08	102.35	**71.76**	**117.54**
	0.10	9.34	**102.36**	463.95	600	**2.14**	305.95	97.99	1007.47	3.96	104.84	**81.57**	**105.41**
Symmetric	0.00	8.21	**120.77**	388.45	600	**2.30**	311.86	91.28	1126.26	8.89	138.80	**84.80**	**116.46**
	0.01	8.28	**120.77**	388.45	600	**2.31**	311.62	91.31	1130.12	8.85	128.64	**89.26**	**109.04**
	0.05	8.53	**120.47**	362.32	600	**2.31**	299.99	90.85	1080.83	8.83	128.92	**68.46**	**110.74**
	0.10	8.59	**121.89**	308.44	600	**2.33**	303.84	102.52	996.42	8.92	121.00	**73.09**	**111.82**
Plateau	0.00	6.51	24.39	148.42	600	**0.17**	**6.98**	89.91	1479.89	3.80	18.85	**75.49**	**46.73**
	0.01	6.51	24.39	148.42	600	**0.17**	**6.99**	90.00	1478.08	3.71	18.80	**60.00**	**45.52**
	0.05	6.51	24.39	148.42	600	**0.24**	**7.03**	90.29	1449.22	3.85	20.73	**62.52**	**47.36**
	0.10	6.51	24.39	150	600	**0.23**	**7.13**	90.26	1365.96	3.89	6.31	**73.01**	**28.76**

ter, according to Table 1 and using the Complex Morlet as mother wavelet.

2.3. Estimation through existing methods and performance comparison

WInCA was compared with PRAAT and BioVoice in F_0 and RFs estimation.

With BioVoice F_0 estimation is performed by means of a two-step algorithm. First, the Simple Inverse Filter Tracking (SIFT) is applied to time windows of short and fixed length; afterwards, F_0 is adaptively estimated on signal frames of variable length through the Average Magnitude Difference Function (AMDF) within the range provided by the SIFT [7,24,42]. Resonance frequencies are obtained by peak picking in the power spectral density (PSD) evaluated on the same adaptive time windows used for F_0 estimation. On each varying time window, an autoregressive model of order q = 22 (half of the sampling frequency in kHz) is applied. This choice for q comes from acoustic and physiological constraints, giving an enough detailed spectrum while preventing spectral smoothing and consequent loss of spectral peaks [23,43].

In PRAAT F_0 is estimated through the autocorrelation of the signal calculated on short frames while Linear Predictive Coding (LPC) is applied for the RFs estimation [15,16]. We underline here that using the default values of PRAAT for both F_0 and RFs ranges as well as their number definitely incorrect results were obtained. Therefore a careful selection of the best values was carried out, necessarily in an empirical way. Specifically we chose the F_0 range equal to 200–800 Hz and for RFs up to 10000 Hz. The number of RFs was set to 5 but we considered only the first three frequencies [16].

Once that F_0–F_3 were estimated with the three methods (WInCA, Biovoice and PRAAT), the F_0 vectors were resampled at 100 Hz (sampling frequency of the control vectors of the synthesized signals). In this way, the comparison with the reference values was possible through the calculation of the root-mean-square error (RMSE) in Hz, defined as:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - y_i)^2} \quad (15)$$

where N = 201 is the number of samples of the control vectors (length of the F_0 vector of the synthesized signal), x_i is the i-th F_0 value of the synthetic signal and y_i is the corresponding i-th estimated F_0 values. As the synthesized signals had fixed RFs, the interpolation process was not performed on the RFs vectors (F_1–F_3). Thus N in Eq. (15) is equal to the number of samples contained in the RFs vector estimated with the three methods and x_i corresponds to the RFs values considered for the synthesized signals (F_1 = 1100 Hz, F_2 = 3000 Hz, F_3 = 5500 Hz).

3. Results

Table 2 shows the RMSE values for each method and the 20 synthetic melodic shapes: complex, falling, rising, symmetric and plateau with: no added noise, 1%, 5% and 10% added noise. Best results are highlighted in bold.

To provide an overall picture, Table 3 summarizes the mean values of RMSE over all the melody shapes and all noise levels for the three methods.

4. Discussion

The comparison among the three methods reported in this paper is made according to synthetic signals that simulate the main melodic shapes of newborn cry. To test the robustness of the methods increasing noise levels are added to the signals. The proposed

Table 3
Overall RMSE of F_0–F_3 estimates with the three methods. Best results are highlighted in bold.

	RMSE F0 [Hz]	RMSE F1 [Hz]	**RMSE F2 [Hz]**	**RMSE F3 [Hz]**
WInCA	7.38 ± 1.16	**80.17 ± 36.13**	334.92 ± 135.48	600 ± 0.00
PRAAT	**1.88 ± 0.89**	199.40 ± 144.11	92.33 ± 6.28	1135.20 ± 164.39
BioVoice	5.94 ± 1.96	82.59 ± 43.72	**78.80 ± 10.78**	**97.24 ± 31.75**

synthesizer represents a breakthrough with respect to existing literature.

To give a measure of the matching capability of the three methods the RMSE is computed between the synthetic signals and the estimated ones. Results (Tables 2 and 3) show that:

– PRAAT gives slightly better results as far as F_0 estimation is concerned, with a mean error less than 2 Hz. However, it fails in the RFs estimation (except for plateau) with a mean error ranging from 92 Hz (for F_2) up to 1135 Hz for F_3.

We point out again that these results are obtained after a careful manual selection of ranges and thresholds for F_0 and for the RFs, according to literature. Much worse results are obtained if the default values were used (40–500 Hz for F_0, up to 5500 Hz for RFs, considering 5 formants) [16]. Thus PRAAT must be used carefully.

An additional advantage of PRAAT is undoubtedly the low computing time: the analysis of a signal of 1 s of duration requires less than 0.5 s for F_0 estimation and about 2 s for the RFs.

– WInCA: F_0 gives an acceptable error but higher than PRAAT. Concerning RFs, quite good results are obtained for F_1 estimation only. The computing time is comparable to PRAAT: for 1 s of recording WInCA requires 0.9 s for the estimation of F_0 and 2.8 s for the estimation of RFs. Another advantage is that WInCA does not require any manual setting. If integrated into BioVoice (or provided with the pre-processing step) it could be a suitable alternative to other tools.
– BioVoice: Quite good results for F_0 (mean error <6 Hz) and better results for RFs as compared to PRAAT. For F_2 and F_3 it gives the best results among the three tools (errors less than 100 Hz), with an error similar to that of WInCA for F_1 (82 Hz against 80 Hz with WInCA). Moreover it does not require any manual setting parameters and implements the pre-processing step. Finally, it allows the analysis of multiple signals simultaneously, building folders with exportable graphs and tables. These advantages along with the sophisticated implemented techniques for F_0 and RFs estimation represent its main drawback that is the computational complexity (for 1 s of recoding it requires 3 s for F_0 estimation and 4.7 s for the RFs).

Thus, the use of one technique or another one should therefore be carefully evaluated according to the specific needs in order to avoid errors and unreliable results.

Further suggestions include the choice of the analysis window and the minimum duration of cry units to analyze. The time duration of the analysis window should be set as short as possible. In fact, we applied a window of 10 ms in WInCA for the estimation of F_0 and RFs. However, most of the traditional analysis techniques do not allow adequate frequency resolution for frames of such short duration and typically much longer windows (even 50 ms) are used leading to a poor temporal resolution. As a compromise it is therefore suggested to use windows up to a maximum of 20 ms of duration (used in BioVoice).

Among the techniques presented here PRAAT and BioVoice have already been applied to real signals [7,25,44], while the application of WInCA will be the subject of future studies.

5. Conclusions

This paper presents a comparison of three different techniques for the automated acoustical analysis of newborn cry. Specifically two existing tools (PRAAT and BioVoice) are compared to a newly developed one, named WInCA (Wavelet Infant Cry Analyzer). We point out that the infant cry is a signal extremely difficult to analyze with standard techniques due to its quasi-stationarity and to the high range of frequencies of interest. This study could represent a guideline for the automated acoustical analysis of infant cry, in order to include these objective methods in the clinical practice along with perceptual assessments.

The crying of newborns is a functional expression of basic biological needs. It is based on a coordinated effort of several brain regions, mainly brainstem and limbic system and is linked to the breath system. Therefore cry features reflect the development and the integrity of the central nervous system.

In addition to other clinical tests the automated infant cry analysis, when properly applied, is a suitable cheap and contactless approach for an early assessment of the neurological state of infants. A reliable automated method for the estimation of crying acoustical characteristics could provide a support to the perceptive analysis made by clinicians reducing the required amount of time often prohibitive in daily clinical practice.

In this work, a newborn cry synthesizer and a new wavelet-based method for newborn cry analysis are presented. Synthetic signals are obtained implementing a new accurate simulation model that takes into account the variability of the signal under study as well as the basic shapes of the newborn cry melody. A database of the synthetic signal is available on request to the authors.

The new analysis tool named WInCA has been developed specifically for the acoustical analysis of newborn cry, characterized by high fundamental frequency F_0 and for being quasi-stationary. The wavelet approach is in fact particularly suited to the study of infant cry thanks to its time frequency high resolution characteristics and low computing time. To assess its performance WInCA is compared to the already existing tools PRAAT and BioVoice on a set of synthetic melodic F_0 shapes corrupted by increasing noise levels.

Results show that PRAAT gives slightly better F_0 results but requires proper settings. WInCA is best suited for F_1 estimation and BioVoice gives best results for F_2 and F_3. It is also the most robust against increasing noise.

As specified in the introduction, the aim of this work is to propose for the first time a comparison among objective methods for newborn cry analysis. This study was performed on synthetic signals because we need to know the exact reference values of signals to obtain an accurate comparison.

The synthesizer proposed in this work is capable to generate the five basic waveforms of the newborn cry that are pointed out in the literature: rising, falling, symmetric, complex and plateau. The synthetic signals have a good match with the real cries also from the perceptual point of view. Of course, they do not represent the wide variety of real cries, which include tens of shapes, but are nevertheless useful, since no references exist to date. Upon request the authors will provide the synthesized signals to test new methods for infant cry analysis.

The application to real signals will be the subject of future works. In particular, the advantages of Biovoice and WInCA will be merged for the analysis of infant cry in order to detect clinically useful parameters and ranges. Moreover, it could be applied to any category of sick newborns or to compare preterm and full term infants as well.

Acknowledgements

This work was partially carried on under Project PGR00202 “Analysis and classification of voice and facial expression: application to neurological disorders in neonates and adults”, Italian Ministry of Foreign Affairs – Progetti Grande Rilevanza Italy-Mexico.

References

[1] O. Wasz-Höckert, E. Valanne, V. Vuorenkoski, K. Michelsson, A. Sovijarvi, Analysis of some types of vocalization in the newborn and in early infancy, Ann. Paediatr. Fenn. 9 (1963) 1–10.
[2] O. Wasz-Höckert, T.J. Partanen, V. Vuorenkoski, K. Michelsson, E. Valanne, The identification of some specific meanings in infant vocalization, Experientia 20 (3) (1964) 154.
[3] K. Michelsson, K. Eklund, P. Leppanen, H. Lyytinen, Cry characteristics of 172 healthy 1-to 7-day-old infantsm, Folia Phoniatr. Logop. 5 (2002) 190–200.
[4] H.E. Baeck, M.N. de Souza, Longitudinal study of the fundamental frequency of hunger cries along the first 6 months of healthy babies, J. Voice 5 (2007) 551–559.
[5] H. Rothganger, Analysis of the sounds of the child in the first year of age and a comparison to the language, Early Hum. Dev. 75 (2003) 55–69.
[6] O.F. Reyes-Galaviz, E.A. Tirado, C.A. Reyes-Garcia, et al., Classification of infant crying to identify pathologies in recently born babies with ANFIS, in: Klaus Miesenberger (Ed.), Computers Helping People with Special Needs, Springer, Berlin, 2004, pp. 408–415.
[7] C. Manfredi, L. Bocchi, S. Orlandi, L. Spaccaterra, G.P. Donzell i, High-resolution cry analysis in preterm newborn infants, Med. Eng. Phys. 31 (5) (2009) 528–532.
[8] J.R. Deller, J.H. Hansen, J.G. Proakis, Discrete-Time Processing of Speech Signal, McMillan Publishing Company, New York, 1993, pp. 724–728.
[9] 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.
[10] A. Fort, A. Ismaelli, C. Manfredi, P. Bruscaglioni, Parametric and non-parametric estimation of speech formants: application to infant cry, Med. Eng. Phys. 18 (8) (1996) 677–691.
[11] A. Fort, C. Manfredi, Acoustic analysis of newborn infant cry signals, Med. Eng. Phys. 20 (1998) 432–442.
[12] S.J. Sheinkopf, J.M. Iverson, B.M. Lester, Atypical cry acoustics in 6-month-old infants at risk for autism spectrum disorder, Autism Res. 5 (5) (2012) 331–339.
[13] P. Sirviö, K. Michelsson, Sound-spectrographic cry analysis of normal and abnormal newborn infants, Folia Phoniatr. Logop. 28 (3) (1976) 161–173.
[14] Corp Kay Elemetrics, Multi-Dimensional Voice Program (MDVP): Software Instruction Manual, Kay Elemetrics, Lincon Park, 2003.
[15] P. Boersma, D. Weenink, Praat: Doing Phonetics by Computer [Computer Program] 2014 Version 5. 3 74, 2014, http://www.praat.org/, Accessed: 07 June 2016.
[16] P. Boersma, Praat, a system for doing phonetics by computer, Glot. Int. 5 (9/10) (2002) 341–345.
[17] M.Z. Laufer, Y. Horii, Fundamental frequency characteristics of infant non-distress vocalization during the first twenty-four weeks, J. Child Lang. 4 (02) (1977) 171–184.
[18] K. Michelsson, O. Michelsson, Phonation in the newborn, infant cry, Int. J. Pediatr. Otorhinol. 49 (1999) S297–S301.
[19] M.P. Robb, D.H. Crowell, P. Dunn-Rankin, Sudden infant death syndrome: cry characteristics, Int. J. Pediatr. Otorhinol. 77 (8) (2013) 1263–1267.
[20] B. Reggiannini, S.J. Sheinkopf, H.F. Silverman, X. Li, B.M. Lester, A flexible analysis tool for the quantitative acoustic assessment of infant cry, J. Speech Lang. Hear. Res. 6 (5) (2013) 1416–1428.
[21] Y. Kheddache, C. Tadj, Resonance frequencies behavior in pathologic cries of newborns, J. Voice 29 (1) (2014) 1–12.
[22] M.P. Robb, A.T. Cacace, Estimation of formant frequencies in infant cry, Int. J. Pediatr. Otorhinol. 32 (1) (1995) 57–67.
[23] K. Wermke, W. Mende, C. Manfredi, P. Bruscaglioni, Developmental aspects of infant's cry melody and formants, Med. Eng. Phys. 24 (2002) 501–514.
[24] C. Manfredi, L. Bocchi, G. Cantarella, A multipurpose user-friendly tool for voice analysis: application to pathological adult voices, Biomed. Signal Process. Control 4 (2009) 212–220.
[25] S. Orlandi, L. Bocchi, G.P. Donzelli, C. Manfredi, Central blood oxygen saturation vs crying in preterm newborns, Biomed. Sig. Proc. Control 7 (2012) 88–92.
[26] 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 Process. Control 8 (6) (2013) 799–810.
[27] C. Manfredi, V. Tocchioni, L. Bocchi, A robust tool for newborn infant cry analysis, in: Conf. Proc. IEEE Eng. Med. Biol. Soc., New York, NY, 2006, pp. 509–512.
[28] N. Rruqja, P.H. Dejonckere, G. Cantarella, J. Schoentgen, S. Orlandi, S.D. Barbagallo, C. Manfredi, Testing software tools with synthesized deviant voices for medicolegal assessment of occupational dysphonia, Biomed. Signal Process. Control 13 (2014) 71–78.
[29] L. Cnockaert, J. Schoentgen, P. Auzou, C. Ozsancak, L. Defebvre, F. Grenez, Low-frequency vocal modulations in vowels produced by Parkinsonian subjects, Speech Commun. 50 (4) (2003) 288–300.
[30] L. Falek, A. Amrouche, L. Fergani, H. Teffahi, A. Djeradi, Formantic analysis of speech signal by wavelet transform, Conf. Proc. World Congr. Eng. vol. 2 (2011) 1572–1576.
[31] G. De Poli, C. Drioli, F. Avanzini, Sintesi Dei Segnali Audio, 1999, http://www.dei.unipd.it/~musica/Dispense/cap5.pdf. Accessed: 07 June 2016.
[32] J.D. Markel, A.H. Gray, Linear Prediction of Speech, Springer – Verlag, Berlin, Heidelberg, New York, 2011.
[33] G. De Poli, F. Avanzini, Sound modeling: signal-based approaches. In: Algorithms for Sound and Music Computing, 2009.
[34] K. Wermke, W. Mende, Musical elements in human infants' cries: in the beginning is the melody, Musicae Scientiae 13 (Suppl. 2) (2009) 151–175.
[35] G. Varallyay, The melody of crying, Int. J. Pediatr. Otorhinol. 71 (11) (2007) 1699–1708.
[36] E. Amaro-Camargo, C.A. Reyes-García, Applying statistical vectors of acoustic characteristics for the automatic classification of infant cry, in: Advanced Intelligent Computing Theories and Applications. With Aspects of Theoretical and Methodological Issues, Springer Berlin, Heidelberg, 2007, pp. 1078–1085.
[37] K. Wermke, D. Leising, A. Stellzig-Eisenhauer, Relation of melody complexity in infants' cries to language outcome in the second year of life: a longitudinal study, Clin. Linguist. Phonet. 21 (2007) 961–973.
[38] S. Orlandi, C.A. Reyes-Garcia, A. Bandini, G.P. Donzelli, C. Manfredi, Application of pattern recognition techniques to the classification of full-Term and preterm infant cry, J. Voice 30 (6) (2016) 656–663.
[39] http://www.audacityteam.org/, Accessed: 07 June 2016.
[40] C. Chui, An Introduction to Wavelets, Academic Press, San Diego, CA, USA, 1995.
[41] P.S. Addison, The Illustrated Wavelet Transform Handbook: Introductory Theory and Applications in Science, Engineering, Medicine and Finance, Institute of Physics Publishing, London, 2002.
[42] C. Manfredi, G. Peretti, A new insight into post-surgical objective voice quality evaluation. Application to thyroplastic medialisation, IEEE Trans. Biomed. Eng. 53 (3) (2006) 442–451.
[43] C. Manfredi, Adaptive noise energy estimation in pathological speech signals, IEEE Trans. Biomed. Eng. 47 (2000) 1538–1542.
[44] G. Esposito, M. del Carmen Rostagno, P. Rostagno, P. Venuti, J.D. Haltigan, D.S. Messinger, Brief report: atypical expression of distress during the separation phase of the strange situation procedure in infant siblings at high risk for ASD, J. Autism Dev. Disord. (2013) 1–6.