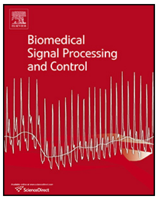
Biomedical Signal Processing and Control xxx (2016) xxx–xxx

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 xxx

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

Please cite this article in press as: S. Orlandi, et al., Testing software tools for newborn cry analysis using synthetic signals, Biomed. Signal Process. Control (2016), http://dx.doi.org/10.1016/j.bspc.2016.12.012

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 ($1 \leq n \leq N$) 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 $1 \leq k \leq K$ 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]:

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

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

Please cite this article in press as: S. Orlandi, et al., Testing software tools for newborn cry analysis using synthetic signals, Biomed. Signal Process. Control (2016), http://dx.doi.org/10.1016/j.bspc.2016.12.012