Medical Engineering & Physics 31 (2009) 528–532

Contents lists available at ScienceDirect

Medical Engineering & Physics

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

High-resolution cry analysis in preterm newborn infants

C. Manfredi [a,*], L. Bocchi [a], S. Orlandi [a], L. Spaccaterra [b], G.P. Donzelli [b]

[a] Department of Electronics and Telecommunications, Università degli Studi di Firenze, Via S. Marta 3, 50139 Firenze, Italy
[b] Department of Paediatrics, Children Hospital A. Meyer, Università degli Studi di Firenze, Firenze, Italy

ARTICLE INFO

Article history:
Received 21 March 2008
Received in revised form 15 October 2008
Accepted 16 October 2008

Keywords:
Newborn infant cry analysis
High-resolution resonance frequencies analysis
Fundamental frequency estimation
Preterm newborn infants
User-friendly analysis tool

ABSTRACT

Infant monitoring is a common procedure in clinical practice in neonatal critical care units. A number of vital functions are monitored, such as heart beat, breathing, blood flow, etc. Specifically, preterm and/or low-birth-weight infants often present respiratory problems that require monitoring. These may range from insufficient ventilation to apnoea. One of the most common events that may affect the respiratory flow is crying, a physiological action made by the infant to communicate and draw attention, but, for a preterm infant, this action requires great effort, which may cause distress and even may have an adverse impact on blood oxygenation. Acoustic analysis of newborn infant cry is thus of importance, since it is related to other basic neuro-physiological parameters. Being easy to perform, cheap and completely non-invasive, it can be successfully applied in many circumstances.

The newborn infant cry is characterised by very high fundamental frequency (F_0) and resonance frequency (RFs) values, with abrupt changes and voiced/unvoiced features of very short duration in a single utterance. To deal with such signals, a new user-friendly software tool has been developed, that allows robust tracking of main acoustic parameters on very short and time-varying signal frames. The software developed provides the user with a high-resolution picture of the cry signal characteristics.

1. Introduction

With the increased survival of very preterm infants, there is a growing concern for their developmental and socio-emotional outcomes. Infant cry characteristics reflect the development and possibly the integrity of the central nervous system. Previous studies have shown that preterm infants and infants with neurological conditions have different cry characteristics such as fundamental frequency, when compared to healthy full-term infants. Research has been developed to study possible differences between full-term and preterm infants in their neuro-physiological maturity and the subsequent impact on their speech development [1–5]. Also, the anatomical and physiological bases of inspiratory cry phonation are of relevance [6]. Newborn infant cry characteristics change with increasing conceptual age: the older the child the more the cry pattern resembles that of full term infant [7,8]. Acoustic analysis of preterm infant cry signals can thus give an aid to clinical diagnosis and prevention of distress since it is easy to perform, cheap and completely non-invasive [9–13].

Difficulties arise, however, as far as the reliability of results is concerned, due to the signals under study. In fact, the newborn infant cry is characterised by a very high fundamental frequency, F_0, with abrupt changes and voiced/unvoiced features of very short duration, in a single utterance. Moreover, vocal tract resonance frequencies (RFs), F_i, need accurate high-resolution tracking. Following RFs evolution in time as well as during the first months of life can give useful clinical information pertaining to preterm and full-term newborn phonatory capability development [7,8]. Classical FFT methods can fail in such cases, because the signal is quasi-stationarity only on very short time frames (often even less than 10 ms).

In this paper, first results are presented concerning a new software tool for voice analysis named BioVoice. It is applicable to high-pitched quasi-stationary voice signals, and performs robust tracking of the main acoustic parameters. The tool is equipped with a user-friendly interface, which allows selection of age, sex and type of vocal emission for each patient, no manual setting to be made by the user (Fig. 1). BioVoice automatically adjusts internal settings for optimal frame length and frequency range for analysis and plots. Specifically, the interface allows for:

- Choosing the voice type, ranging from high-pitched newborn and singers voices to adult voices (40 Hz < F_0 < 1300 Hz).
- Selecting the kind of analysis: single audio file or two files (for comparison purposes).
- Loading (and listening to) data (.wav files).

* Corresponding author. Tel.: +39 055 4796410.
E-mail address: manfredi@det.unifi.it (C. Manfredi).

1350-4533/$ – see front matter © 2008 IPEM. Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.medengphy.2008.10.003

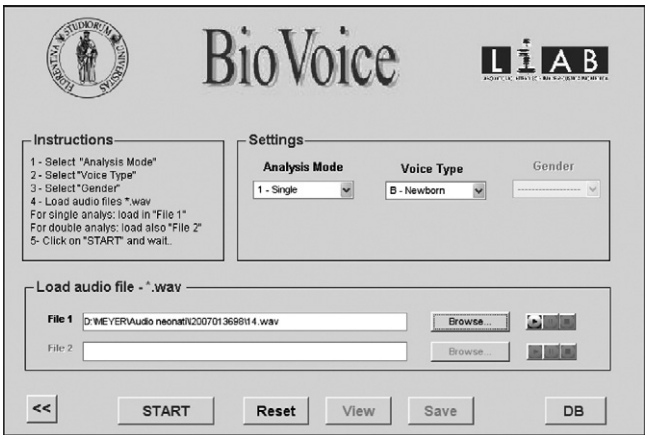

Fig. 1. The user interface for the new voice analysis tool named BioVoice. The original picture is in the colour map.

- Viewing and saving results (in printable format) and storing information in a data base.

A notice is added concerning computer time required: for long files (>5 s) and high sampling frequency (>40 kHz) the total time could approach 1–2 min. A moving bar shows the residual time during computations. For each voice type, a number of ad hoc plots and tables is displayed and saved in printable format for a visual comparison of results. Specifically, for an infant cry, coloured map plots are completed, showing F_0, V/UV frames and a spectrogram with the first three RFs F_1–F_3 superimposed. A set of tables summarise the mean, standard deviation (std), maximum and minimum values for F_0 and F_1–F_3. Also, the number of cry episodes, cry length and the corresponding maximum energy is displayed. These parameters are, in fact, considered among the most meaningful in newborn cry analysis [7–9].

2. Materials and methods

As pointed out in Section 1, the newborn infant cry is characterised by high fundamental frequency and RFs values rapidly varying in time. Thus, a reliable analysis requires high-resolution techniques to be used. In this section, details are given concerning the implemented techniques.

2.1. F_0 estimation

The fundamental frequency, F_0, was estimated with a two-step procedure. Simple inverse filter tracking (SIFT) was applied first [14–16], to signal time windows of short and fixed length. The window length was chosen as $M = 3F_s/F_{min}$, where F_s is the signal sampling frequency and F_{min} is the minimum allowed F_0 value for the signal under consideration (for newborn cry: $F_{min} = 150$ Hz). Moreover, instead of a low and fixed one, an adaptive choice for the SIFT filter order was applied, that allowed the detection of varying signal characteristics. The choice was based on singular value decomposition (SVD) of suitable data matrices that required selecting the "size" k of the signal subspace (i.e. the minimum number of eigenvectors spanning the clean data). This was achieved by separating the largest (squared) singular values σ_i^2 from the smallest ones by means of a variable threshold, which was based on the dynamic mean evaluation (DME) criterion [11]. Typically with DME, $2 \le k \le 6$ during the utterance and the larger the estimated k, the more the signal varied. From the first step, a raw F_0 tracking was

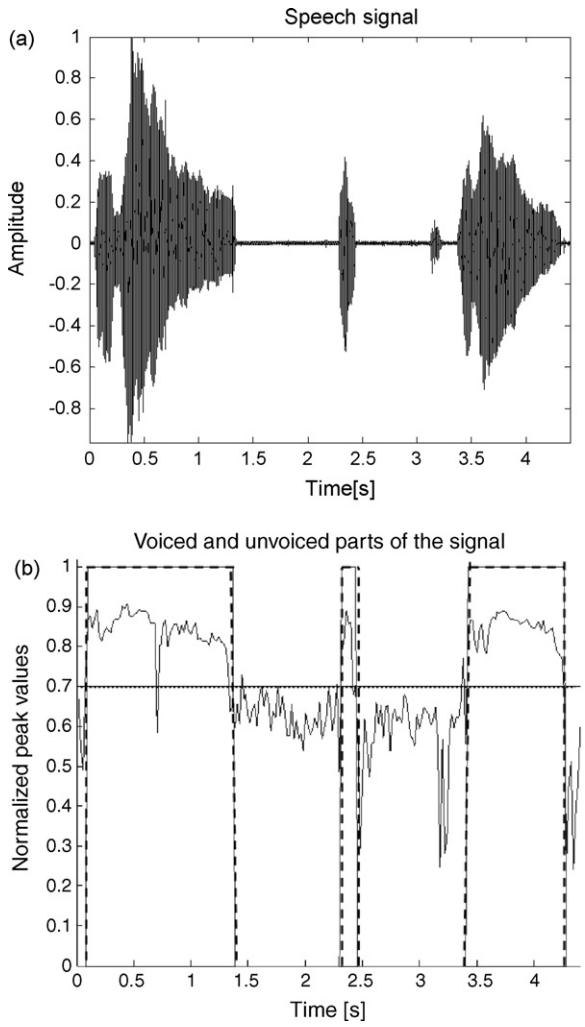

Fig. 2. Voiced/UnVoiced (V/UV) selection for a preterm infant as obtained with the proposed approach. Even very short voiced frames (few tenths of ms) are recovered.

obtained along with its range of variation $[F_l, F_h]$, where F_l = lowest F_0 value and F_h = highest F_0 value.

In order to disregard voiceless parts of the signal, a *voiced/unvoiced decision* (V/UV) was applied. It was based on the approach proposed previously in [14] and was suitably modified for our purposes here. Basically, a signal frame is selected as voiced if the maximum of the autocorrelation function on that frame, γ_{max}, is larger than a threshold value, linked to F_0. Modifications were introduced, in order to exclude possible wrong V/UV choices, through a number of controls made on adjacent frames [13]. For a newborn cry, it was commonly found that $\gamma_{max} \ge 0.7$. An example is reported in Fig. 2, relative to the preterm newborn infant described later in Section 3 and in Fig. 4. Fig. 2a shows the normalised signal amplitude where three voiced parts can be easily detected by eye inspection roughly corresponding to 0.2–1.4 s, 2.4–2.5 s, and 3.4–4.4 s. In Fig. 2b, the corresponding successful V/UV detection is shown where the solid line represents the signal autocorrelation, the dashed line shows the selected voiced parts, and the dotted line corresponds to the threshold value $\gamma_{max} = 0.7$. Notice that the V/UV procedure was capable of detecting the very short voiced frame, lasting about 0.1 s.

In the second step, F_0 was adaptively estimated inside $[F_l, F_h]$. This allowed for a more precise F_0 estimation. According to [11,15], a variable window length for analysis was applied, that was inversely

proportional to the changing F_0. Very short time windows, ranging from 5 to 15 ms, were thus obtained, and they were locally dependent on the signal variability. Over each time window, the signal was band-pass filtered (for newborn cry the range was settled to 150–900 Hz. This range can be easily adapted to other situations, e.g. hyperphonated cry, where F_0 could rise above 1000 Hz) with the Mexican hat continuous wavelet transform, and the signal periodicity was extracted by means of the average magnitude difference function (AMDF) approach [16]. The choice of the AMDF instead of the autocorrelation sequence (AS) was due to the non-stationarity and amplitude modulation of the signals under study that were shown to cause an incorrect estimation of the true signal periodicity. In case of fast and abrupt F_0 changes, this procedure was shown to increase the robustness of the F_0 estimation, giving enhanced results with respect to standard methods [15,17].

2.2. Resonance frequencies

Even if vowel frequencies cannot be found in newborn cries, resonance frequencies (RFs) reflect important acoustical characteristics of the infant vocal tract. For RFs estimation and tracking, a robust parametric technique was proposed, obtained by peak picking in the power spectral density (PSD) plot. This was evaluated on the same adaptive time windows as previously described. For PSD estimation, autoregressive (AR) models were used. The model order q varied according to the signal characteristics. The "modified covariance method" was applied, as it was shown to give the best results for the reduction of spectral line splitting and bias of the frequency estimations [16].

One of the main advantages of parametric spectral analysis over classical approaches relies in its high-resolution capability, as the model extrapolates data outside the analysed window. AR spectral estimators are very sensitive to order selection, however, in the case of overestimated model order q, formant splitting can occur, while underestimation smoothes the spectrum and causes misallocation of spectral peaks. Many criteria have been defined for finding the best model order q, but they were shown to be almost unreliable for short data frames due to long-term convergence properties [16]. In this paper, the relation $q \cong 0.5F_s$ (in kHz) was found to be optimal for obtaining an enough detailed spectrum, while preventing spectral smoothing and consequent loss of spectral peaks. This relation comes from the physical constraint: $q = 2LF_s/c$, where L is the length of the vocal tract ($L \cong 8.5$ cm for newborns) and $c = 340$ m/s is the speed of sound [18]. This choice has already been proven effective in many applications, with enhanced results as far as resolution is concerned [13,17]. Notice that most commercial software tools require manual setting of q, thus implying some expertise required by the user in order to obtain reliable results. Default values are, in fact, usually not suited for this application, and could strongly distort results.

Co-ordinates of PSD maxima on each time window, as well as their mean and std value on the whole signal, were also evaluated. Thus, details were given about RFs evolution in time as related to energy. The first three RFs, F_1, F_2, and F_3, were considered here, commonly between 1 and 6 kHz.

3. Experimental results

The analysis has been carried out on a group of 19 preterm infants, with a gestational age ranging from 23 to 38 weeks and a birth weight between 590 and 3020 g. No relevant pathology was found among the analysed infant group. Infants were selected by physicians from patients at the Critical Care Unit of the Children Hospital A. Meyer, in Firenze, Italy.

A small control group composed of 2 full-term healthy infants (gestational age > 37 weeks, weight at birth >3000 g) has been used as a reference set. As at present, in our knowledge, reference values are not available for preterm newborns, we compared our results with F_0–F_3 reference values reported in literature for full-term newborns [1–9].

Cry recording was performed by collecting data with a unidirectional microphone (Shure SM58) equipped with Tascam US-144 portable audio/MIDI interface (96 kHz/24-bit recording). A multimedia laptop allowed for acquiring a single channel audio track with a sampling rate of 44 kHz. Hence, according to the discussion carried on in the previous section, the AR model order for RFs estimation was fixed at $q = 22$.

Several audio recordings for each infant were acquired lasting about 30 min each. From these recordings, the three most significant cry episodes were selected for each infant, and lasted approximately 5–6 s each. A cry episode was defined here as a sequence of high energy utterances, which were possibly separated by few milliseconds of silence (the inspiratory action). Cry episodes were manually selected, by listening to the recorded audio signal. An automatic procedure for cry episode extraction from the whole recording is under development. The choice of the cry episodes was also made taking into account other factors, mainly the quality of recordings. The signal should contain the lowest possible amount of background noise coming from other sensors and devices connected to the newborn. Notice that the selected cries were chosen from three different recordings for each newborn, to have independent data sets. Hence, a set of about 60 cry episodes was collected.

Finally, it should be pointed out that, in some cases, recordings were made some weeks after the birth, because the newborn was required to be in an incubator for an extended time period. This was taken into account in the analysis of the results.

The aim of this work was to understand if gestational age and/or weight at birth can influence newborn cry. To this aim, first we performed cry analysis for all available data to find out any possible grouping of the newborns under analysis. This analysis gave us an almost consistent indication for a cut off point of 34 weeks for gestational age, g.a. and 2500 g for weight at birth, w.a.b.

Hence, newborns were divided into two classes, based on g.a. (less or greater than 34 weeks) and w.a.b. (below or above 2500 g). Notice that a low-birth weight newborn does not necessarily correspond to a preterm one; hence, different results could be obtained with the two classes.

In order to find possible relationships between F_i, $i = 0, 1, 2, 3$, and w.a.b. and/or g.a., t-test statistical analysis was applied to all available data (21 cases, e.g. 63 independent data sets) and to the set of preterm newborns only (19 cases, e.g. 57 independent data sets). Possible correlations between F_0 and F_i, $i = 1, 2$ values have been exploited. F_3 was excluded from the comparison due to its high variability. A highly significant statistical difference was found in both classes ($p \ll 0.01$).

Furthermore, to make the data more homogeneous, full-term newborns were excluded from comparison. Also, one preterm infant was excluded due to his/her long-term incubation period, which did not allow for cry recordings until 1 month after birth. Hence, a total of 18 cases, divided according to g.a. and w.a.b., were considered. Two sets of 9 cases each were obtained. Both classes showed a decrease in F_0, F_1 and F_2 mean, maximum and minimum values with increasing age/weight. F_3 was highly unstable in most cases, which gave almost unreliable results for comparison.

Fig. 3 shows box plots comparing F_0, F_1, and F_2 (a–c, respectively) for the newborn cry data divided according to g.a. and w.a.b. The figure consistently shows a decrease in frequency with increasing age or weight for all the parameters. This result, in agreement with literature [1,3,5,7,9,11], can be explained with increasing length and

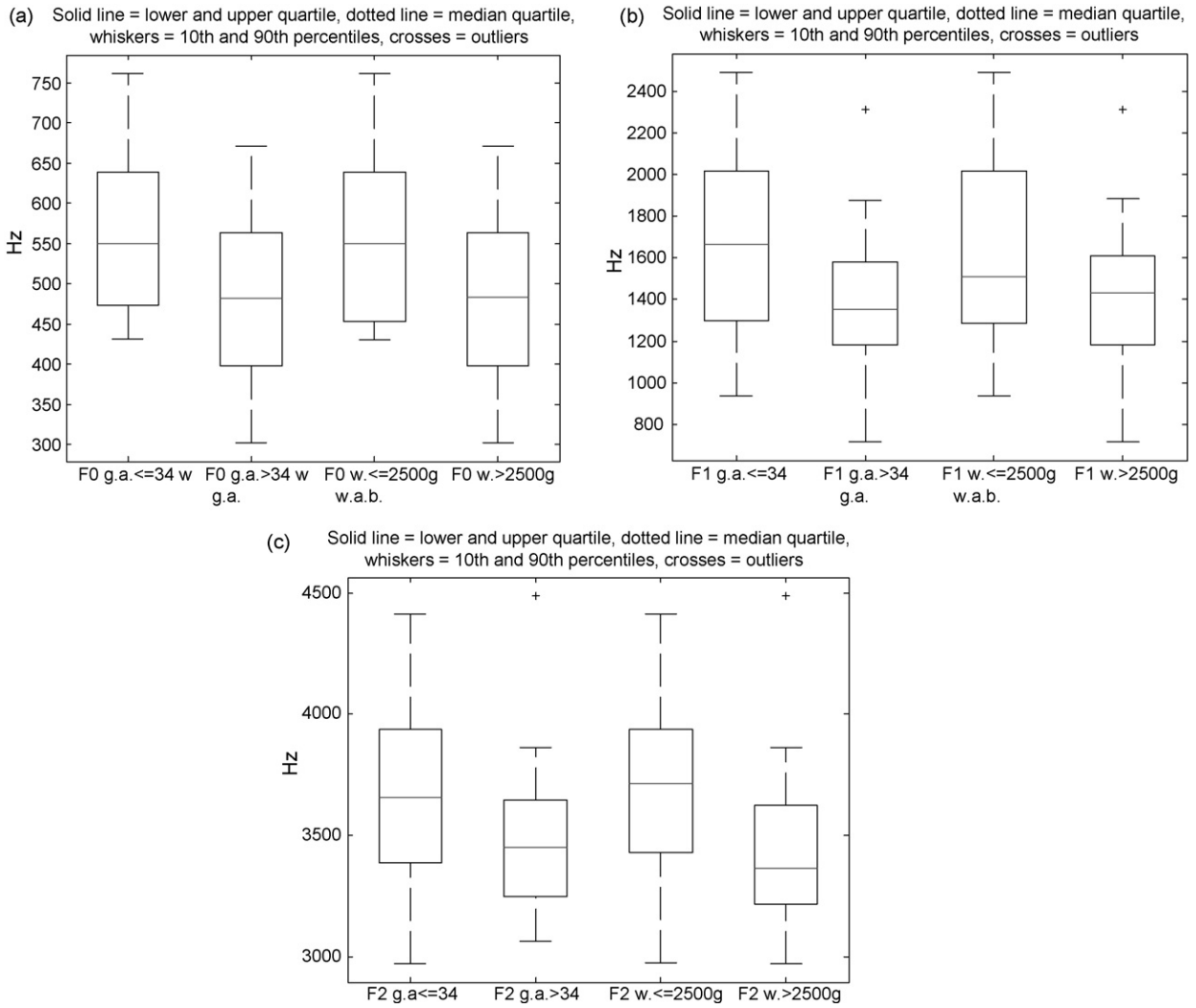

Fig. 3. Box plots comparing F_0, F_1, and F_2 for the newborn cry data, divided according to gestational age and weight at birth (a–c, respectively). Results consistently show a decrease in frequency with increasing age or weight for all the parameters.

structural changes of both vocal folds and vocal tract for increasing g.a. and w.a.b.

Furthermore, statistical comparisons between F_0, F_1, and F_2 have been performed for newborns divided according to g.a and w.a.b. Results are reported in Table 1. Highly significant statistical differences ($p \ll 0.001$) were always found between F_0–F_1, F_0–F_2, and F_1–F_2 comparison (rows 1–3). Also, comparing F_0–F_0 and F_1–F_1 for newborns divided according to gestational age (bold) had statistical differences. Hence, gestational age seems more relevant than weight at birth. Significant statistical differences ($p < 0.05$) were found in almost all other cases. This result could be explained with the high variability of F_1 and F_2 values, which were possibly related to the incomplete physiological development of the vocal tract in preterm newborn infants.

Table 1
Results of the statistical comparisons between F_0, F_1 and F_2 for the newborn cry data divided by gestational age and weight at birth. Highly significant statistical differences ($p \ll 0.01$) were always found in the F_0–F_1, F_0–F_2 and F_1–F_2 comparisons as well as between F_0–F_0 and F_1–F_1 for newborns divided according to gestational age (bold). Significant statistical differences ($p < 0.05$) were found in almost all other cases.

t-Test	g.a. < 34 week	g.a. > 34 week	w.a.b. < 2500 g	w.a.b. > 2500 g
F_0–F_1	$p \ll 0.0001$	$p \ll 0.0001$	$p \ll 0.0001$	$p \ll 0.0001$
F_0–F_2	$p \ll 0.0001$	$p \ll 0.0001$	$p \ll 0.0001$	$p \ll 0.0001$
F_1–F_2	$p \ll 0.0001$	$p \ll 0.0001$	$p \ll 0.0001$	$p \ll 0.0001$
F_0–F_0	**0.006**		0.029	
F_1–F_1	**0.005**		0.125	
F_2–F_2	0.058		0.012	

To show the performance of the BioVoice tool, an example is reported in Fig. 4. It shows a cry episode coming from a preterm, low-birth-weight newborn (g.a: 27 weeks, w.a.b.: 800 g). As in most very preterm, very low-weight-at-birth cases, however, recordings were made later (32nd–35th week) due to the long duration that the newborn was required to be in the incubator. The original pictures are in the colour map. The normalised signal and V/UV selection are reported in Fig. 2. F_0 tracking is shown in Fig. 4a. Notice almost irregular and short time duration of each utterance. In Fig. 4b, the spectrogram with the tracking of the first three RFs superimposed is displayed. Some basic statistics (mean, std) are also reported. Notice the almost irregular shape of the third RFs, which was almost unrecoverable in most cases under study. This could be linked to the still incomplete vocal tract structure in the preterm newborn.

In the case reported here, loss of periodicity and stability was observed, as well as a high energy of non-harmonic spectral com-

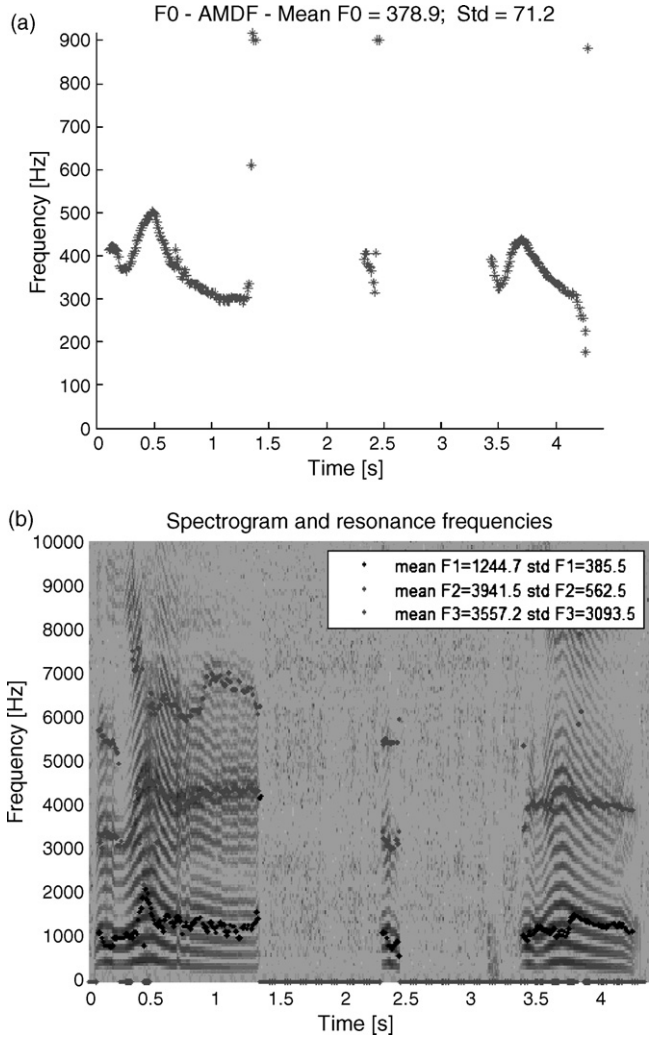

Fig. 4. BioVoice analysis of a preterm, low weight at birth, newborn (gestational age: 27 weeks, weight at birth: 800 g): (a) F_0 tracking with mean and std and (b) spectrogram and F_1–F_3 tracking with mean and std. Original pictures are in the colour map.

ponents. These results could correspond to immature or deficient neurological control in this subject.

4. Conclusions

A new robust high-resolution tool for newborn infant cry analysis was presented. Being completely automatic and adaptive, the proposed software can be successfully used in a wide range of applications, and in the case of highly varying signals as those under study, it does not require any manual setting to be made by the user.

First results on a set of 21 newborn infants have shown that the tool is capable of extracting basic parameters such as fundamental frequency and vocal tract resonances with high reliability. Statistical analysis shows a relationship among such parameters, as related to increasing gestational age and weight at birth. This could give useful information to clinicians working in children hospital critical care units, about possible relationships between crying and the central nervous system development and control.

In future work, more parameters will be added as well as optimisation of the tool under the required computer processing time point of view.

Finally, an automatic procedure for cry episode extraction from long lasting recording is under construction. A data base, that could be useful for exploiting possible correlations and differences among signals, will be completed for diagnosis and classification purposes in neonatal medicine.

Relationships among cry and other parameters are under investigation. Specifically, possible correlations will be investigated among some cry parameters (e.g. F_0, F_1, length of cry episodes, melody and intensity) and the drop in cerebral oxygenation level, to give a measure of distress, also to highlight possible differences in the development of neuro-pyhsiological control between full- and pre-term newborn infants [19].

Acknowledgements

This work has been partially supported by “Ente Cassa di Risparmio di Firenze” under the project: *n.* 2006.1517 “Analisi di segnali ed immagini vocali per applicazioni biomediche” and by COST Action 2103 “Advanced voice quality assessment”.

Conflict of interest

I declare that there is no conflict of interest.

References

[1] Goberman AM, Robb MP. Acoustic examination of preterm and full-term infant cries—the long-time average spectrum. J Speech Lang Hear Res 1999;42:850–61.
[2] Michelsson K, Järvenpää AL, Rinne A. Sound spectrographic analysis of pain cry in preterm infants. Early Hum Dev 1983;8(2):141–9.
[3] Lester BM. Developmental outcome prediction from acoustic cry analysis in term and preterm infants. Paediatrics 1987;80(4):529–34.
[4] Bellieni CV, Sisto R, Cordelli DM, Buonocore G. Cry features reflect pain intensity in term newborns: an alarm threshold. Pediatr Res 2004;55(1):142–6.
[5] Robb MP, Goberman AM, Cacace AT. An acoustic template of newborn infant crying. Folia Phoniatr Logop 1997;49(1):35–41.
[6] Grau SM, Robb MP, Cacace AT. Acoustic correlates of inspiratory phonation during infant cry. J Speech Hear Res 1995;38:373–81.
[7] Wermke K, Mende W, Manfredi C, Bruscaglioni P. Developmental aspects of infant’s cry melody and formants. Med Eng Phys 2002;24(7–8):501–14.
[8] Nicollas R, Ouaknine M, Giovanni A, Berger J, To JP, Dumoulin D, et al. Physiology of vocal production in the newborn, 2003. In: Proc. 3rd Int. Workshop on models and analysis of vocal emissions for biomedical applications. 2003. p. 51–4.
[9] Escobedo D, Cano S, Collo E, Regueiferos L, Capdevila L. Rising shift of pitch frequency in the infant cry of some pathologic case, 2001. In: Proc. 2nd Int. Workshop on models and analysis of vocal emissions for biomedical applications. 2001.
[10] Donzelli GP, Rapisardi G, Moroni M, Scarano E, Ismaelli A, Bruscaglioni P. Neurodevelopmental (ND) prognostic significance of a new method of infant cry analysis in the preterm baby. Biol Neonate 1994;66:153.
[11] Fort A, Ismaelli A, Manfredi C, Bruscaglioni P. Parametric and non parametric estimation of speech formants: application to infant cry. Med Eng Phys 1996;18(8):677–91.
[12] Fort A, Manfredi C. Acoustic analysis of new-born infant cry signals. Med Eng Phys 1998;20(6):432–42.
[13] Manfredi C, Tocchioni V, Bocchi L. A robust tool for newborn infant cry analysis. In: Proc. EMBC Conf. 2006.
[14] Markel JD. The SIFT algorithm for fundamental frequency estimation. IEEE Trans Audio Electroac 1972;20:367–77.
[15] Manfredi C, D’Aniello M, Bruscaglioni P, Ismaelli A. A comparative analysis of fundamental frequency estimation methods with application to pathological voices. Med Eng Phys 2000;22(2):135–47.
[16] Marple SL. Digital spectral analysis with applications. Englewood Cliffs, NJ, USA: Prentice Hall; 1987.
[17] Manfredi C. Adaptive noise energy estimation in pathological speech signals. IEEE Trans Biomed Eng 2000;47:1538–42.
[18] Markel JD, Gray AH. Linear prediction of speech. Berlin DE: Springer-Verlag; 1982.
[19] Manfredi C, Bocchi L, Orlandi S, Calisti M, Spaccaterra L, Donzelli GP. Non-invasive distress evaluation in preterm newborn infants. In: Proc. 30th IEEE EMBS Annual Int. Conf. 2008.