Proceedings of the 28th IEEE
EMBS Annual International Conference
New York City, USA, Aug 30-Sept 3, 2006

A Robust Tool for Newborn Infant Cry Analysis

Claudia Manfredi, *Member, IEEE*, Valentina Tocchioni, Leonardo Bocchi

Abstract - **In this paper, a new robust adaptive tool for new-born infant cry analysis is proposed, characterised by high tracking capability, well suited for the signals under study. It performs F_0, noise and resonance frequencies tracking, on signal frames of varying length (even few ms), adaptively tailored to varying signal characteristics. Moreover, voiced/unvoiced separation is implemented, allowing disregarding unvoiced parts of the signal where misleading results could be obtained. Plots of F_0 and its harmonics, noise tracking, spectrogram with resonance frequencies superimposed, are presented in a coloured-scale. Some added statistics allow further understanding and comparison of results. The new software tool is completely automatic, working with any sampling frequency and F_0, and also with strongly corrupted signals, and does not need any manual setting of whatever option to be made by the user, thus being easily usable also by non-experts. Some examples are reported, concerning both healthy and pathological new-born infant cries.**

I. INTRODUCTION

Acoustic analysis of new-born infant cry signals is of importance, as an aid to clinical diagnosis. Being easy to perform, cheap and completely non-invasive, it can be successfully applied in many circumstances. However, difficulties are encountered as far as reliable results are concerned, due to the signals under study. In fact, new-born infant cry is characterised by very high fundamental frequency F_0, with abrupt changes and voiced/unvoiced features of very short duration, within a single utterance. Moreover, vocal tract resonance frequencies, that in most cases are highly varying, need accurate tracking. In fact, following their evolution in time and during the first months of life can give useful information both as far as new-born phonatory capability evolution and possible CNS dysfunctions are concerned [1]-[7].

In this paper, first results are presented concerning a new software tool for high-pitched voice signals, which allow robust tracking of main acoustic parameters without requiring any manual setting by the user.

Manuscript received April 3, 2006.

C.Manfredi is with the Department of Electronics and Telecommunications, Università degli Studi di Firenze, Via S. Marta 3 – 50139 Firenze, Italy (Corresponding author. Phone: +39-055-4796410; fax: +39-055-494569; e-mail: manfredi@det.unifi.it).

V.Tocchioni is with the Department of Electronics and Telecommunications, Università degli Studi di Firenze, Via S. Marta 3 – 50139 Firenze, Italy. (e-mail:valinatoc@inwind.it).

L.Bocchi is with the Department of Electronics and Telecommunications, Università degli Studi di Firenze, Via S. Marta 3 – 50139 Firenze, Italy. (e-mail: leo@asp.det.unifi.it).

F0, noise, resonance frequencies, and harmonics are evaluated and tracked, on very short time windows. Any possible variation, even within few ms, is measured, thus allowing the user a high-resolution picture of the signal characteristics. The new tool has been successfully applied both to healthy and pathological new-born infant cry signals.

II. MATERIALS AND METHODS

F_0 estimation - In the proposed approach, the fundamental frequency F_0 is estimated by means of a two-step procedure. Simple Inverse Filter Tracking (SIFT) is applied first [8], on signal time windows of short and fixed length. The window length is 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 (here: F_{min}=150Hz). Such short time windows are required, due to high non-stationarity of the signals under study. Moreover, instead of a low and fixed one, an adaptive choice for the filter order is applied, that allows following varying signal characteristics. The choice is based on Singular Value Decomposition (SVD) of suitable data matrices that requires selecting the "size" p of the signal subspace, i.e. the minimum number of eigenvectors spanning the clean data. This is achieved by separating the largest (squared) singular values σ_i^2 from the smallest ones by means of a variable threshold, based on the Dynamic Mean Evaluation (DME) criterion [1], [2]. Typically, with DME, $2 \le p \le 6$ during the utterance: the larger the estimated p, the more varying the signal. From the first step, a first raw F_0 tracking is obtained, along with its range of variation $[F_l, F_h]$.

Moreover, in order to disregard voiceless parts of the signal, a *voiced/unvoiced decision* (V/UV) is applied, based on the approach proposed in [8], suitably modified. 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 are introduced, in order to exclude possible spurious V/UV choices, by means of a two-step selection to optimise the threshold. Details will be reported elsewhere. For new-born cry, it was found that commonly $\gamma_{max} \ge 0.7$. An example is reported in Fig.1, relative to the healthy cry reported in Sect. III.

In the second step, F_0 is adaptively estimated inside $[F_l, F_h]$, thus allowing for a more precise estimation. According to [3]-[5], a variable window length for analysis is used, inversely proportional to varying F_0. On each time window, the signal is band-pass filtered (150Hz-900Hz) with the Mexican hat Continuous Wavelet Transform (empirical choice) and its periodicity is extracted by means of the Average Magnitude Difference Function (AMDF) approach

Authorized licensed use limited to: Universita degli Studi di Firenze. Downloaded on December 22,2020 at 17:52:14 UTC from IEEE Xplore. Restrictions apply.

[9]. The choice of the AMDF instead of the autocorrelation sequence (AS) is due to the non-stationarity and amplitude modulation of the signals under study that were shown to often cause misestimating of the true signal periodicity. In case of fast and abrupt F_0 changes, this procedure was shown to increase robustness in F_0 estimation, giving enhanced results with respect to standard ones [9], [11].

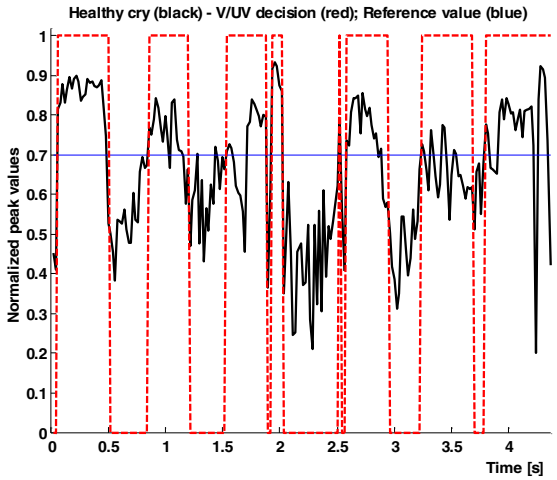

Fig.1. V/UV decision as obtained with the proposed approach.

Noise estimation – A new adaptive noise estimation technique is proposed, that allows tracking varying noise level during phonation, possibly related to pathology (dysphonic and/or dull sound). The method, named ANNE (Adaptive NNE), is based on the Normalised Noise Energy (NNE) comb filtering approach [10], optimised in order to deal with data windows of varying length. The new method relies on choosing an arbitrary, but pre-specified, number of noise lines in spectral 'dip' regions (i.e. between two successive harmonics), thus resulting in accurate noise estimation, as it allows avoiding empty 'dip' regions along the frequency axis. This is achieved by properly defining the optimal time window length for the analysis, adaptively tailored to varying F_0. The method has already been successfully applied to pathological voices, coming from vocal fold operated patients [11]. The ANNE is thus suited to give the physician an objective tracking of voice "hoarseness" due to disease. Specifically, large negative ANNE values correspond to good voice quality, while values close to zero reflect the presence of strong noise.

Resonance frequencies - Even if vowel frequencies cannot be found in cries, resonance frequencies (RF) reflect important acoustical characteristics of the vocal tract of the infant. For RF estimation and tracking, a robust parametric technique is proposed, obtained by peak picking in the Power Spectral Density (PSD), evaluated on the same adaptive time windows as obtained before [1]-[5]. For PSD estimation, the "modified covariance method" is applied, as it was shown to give the best results as far as reduction of spectral line splitting and bias of the frequency estimate are concerned [12].

One of the main advantages of parametric spectral analysis over classical approaches consists in its high-resolution capability, as the model extrapolates data outside the analysed window. However, AR spectral estimators are very sensitive to order selection: in case of overestimated model order p, formant splitting may occur, while underestimation smoothes the spectrum and causes misallocation of spectral peaks. Many criteria have been defined for finding the best model order p, but, they were shown to be almost unreliable for short data frames, due to long-term convergence properties [12]. In this paper, the relation $p \cong 0.5F_s$ (in kHz) was found the best one for obtaining an enough detailed spectrum, while preventing from spectral smoothing and consequently loss of spectral peaks. This relation comes from the physical constraint: $p=2LF_s/c$, where L is the length of the vocal tract ($L \cong 8.5$cm for new-borns) and c=34cm/ms is the speed of sound [13]. This choice has already been proved effective in many applications, with enhanced results as far as resolution is concerned [5], [11].

Co-ordinates of PSD maxima on each time window, as well as their mean and STD value on the whole signal, are also evaluated, thus giving details about RFs evolution in time, as related to energy. The first three RFs, F1, F2, F3 are considered here, below 8 kHz.

Finally, RFs *tuning (TUP) and transition (TRP) periods around F_0 harmonics* are evaluated, and displayed on the spectrogram. A TUP is defined as the time (>20ms) during which a RF remains close (distance <100Hz) to a harmonic. A TRP is the time between two subsequent TUPs. Their duration and evolution seems in fact to be related to the development of phonatory and articulatory capabilities [14]-[16]. Any irregular behaviour (flatness or vibrato) could thus give information about possible neurological dysfunction and /or malformation.

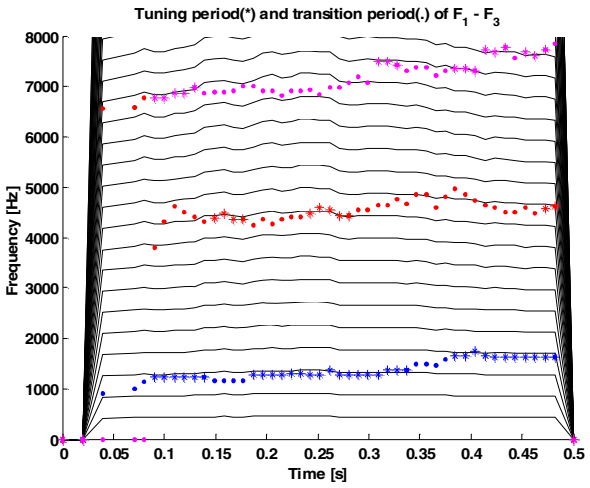

Fig.2. Tuning and transition periods for a healthy cry.

Fig.2 shows almost regular tuning and transition periods for F1-F3, for the healthy cry described in the following section. For clarity reasons, only the first 0.5 s are displayed.

Authorized licensed use limited to: Universita degli Studi di Firenze. Downloaded on December 22,2020 at 17:52:14 UTC from IEEE Xplore. Restrictions apply.

III. EXPERIMENTAL RESULTS

To test the approach, first results are presented concerning one healthy and two pathological cries. Audio signals (sampled at F_s=44.1 kHz) come from a data base available on the web (DISAT, Università degli Studi di Milano-Bicocca, Milano, Italy) [16]. Further work will concern both healthy and pathological recordings, coming form the main Firenze Children Hospital A. Meyer, as a part of a scientific co-operation. The AR model order for RFs estimation is fixed at p=22.

Figures 3-4 show results concerning a healthy cry, caused by sudden movements of the new-born infant during a neurological examination. In Fig.3, the signal (upper plot), F_0 along with the varying time window length (middle plot) and noise tracking (lowest plot) are shown. Almost regular F_0 is found, around 430Hz. Notice the short window length, around 10ms, that adaptively follows F_0 variations. Noise is rather low, around -17dB, as expected with healthy cry.

Fig.4 shows the signal spectrogram with resonance frequencies F1-F3 superimposed. Mean values of F1-F3 are in the usual range for new-born cries, and show a rising-falling shape, typical of healthy babies.

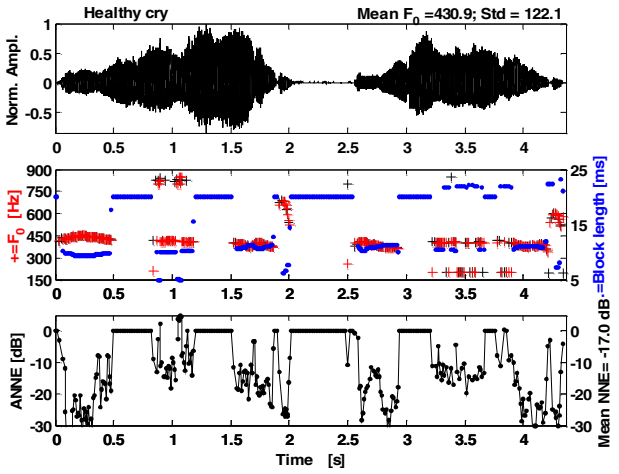

Fig. 3. Normalised signal (black, upper plot), F_0 (red, middle plot) with adaptive window length (blue, middle plot) and noise tracking (black, lowest plot) for a healthy cry.

The following example concerns a premature cry. F_0 and noise are reported in Fig.5. Notice more irregular and shorter time duration of each utterance, as compared to the healthy cry, and higher noise level (around -10dB). Also, notice the very short block length (around 7ms), that allows tracking high F_0 (580Hz). Fig.6 reports the spectrogram with F1-F3 superimposed for the same cry. In this case, loss of periodicity and stability is observed, as well as a prevalence of non-harmonic spectral components. Moreover, F1-F3 are set to lower frequencies with respect to the healthy cry.

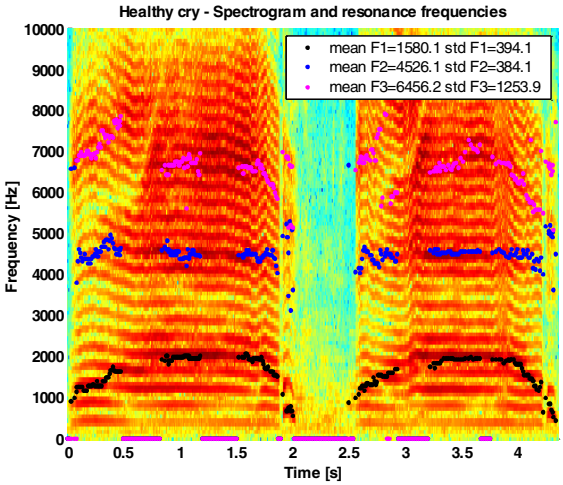

Fig.4. Spectrogram and F1-F3 plot for the healthy cry of Fig.1.

The last example concerns a new-born infant affected by phenylketonuria (PKU). As shown in Fig. 7, F0 is highly unstable and irregular, varying from 350Hz up to 700Hz and more. Noise is low, as can be seen from the spectrogram in Fig. 8 that shows a very irregular spectrum, with few harmonic regions and absence of periodicity for F1-F3, settled at considerably lower frequency values than for the healthy case. Also, notice large STD values, denoting strong irregularity. These results confirm reduced neurological control in this subject.

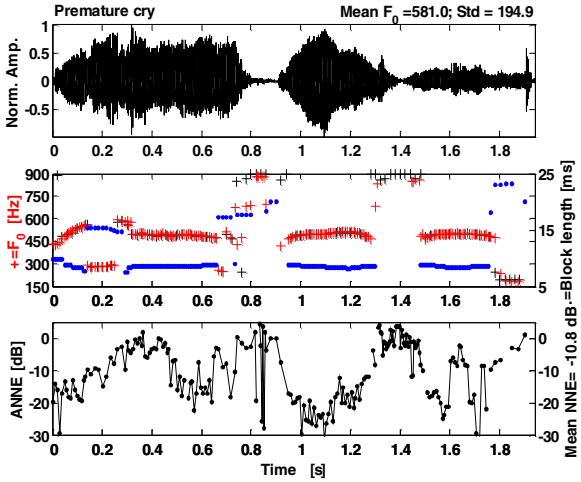

Fig. 5. Normalised signal (black, upper plot), F_0 (red, middle plot) with adaptive window length (blue, middle plot) and noise tracking (black, lowest plot) for a premature cry.

IV. FINAL REMARKS

In this paper, a new robust tool for new-born infant cry analysis is presented. Being completely automatic and adaptive, the proposed software can be successfully used in a wide range of applications, also in case of highly varying signals, without requiring any manual setting to be made by the user. First results are in agreement with [16]. As compared to MDVP® (Kay Elem.) commercial software [17], result have shown better performance for the new tool,

Authorized licensed use limited to: Universita degli Studi di Firenze. Downloaded on December 22,2020 at 17:52:14 UTC from IEEE Xplore. Restrictions apply.

both as far as robustness and visual display are concerned. Details will be reported elsewere.

Future work will concern adding more parameters, as well as further optimising existing ones. A data base is under construction, in co-operation with the Children Hospital A. Meyer, Firenze, Italy, with the aim of searching for possible correlations and differences among signals, for diagnosis and classification purposes.

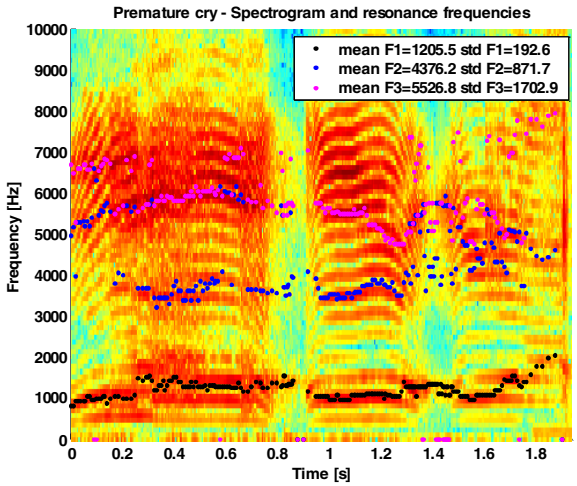

Fig.6. Spectrogram and F1-F3 plot for the premature cry of Fig.5

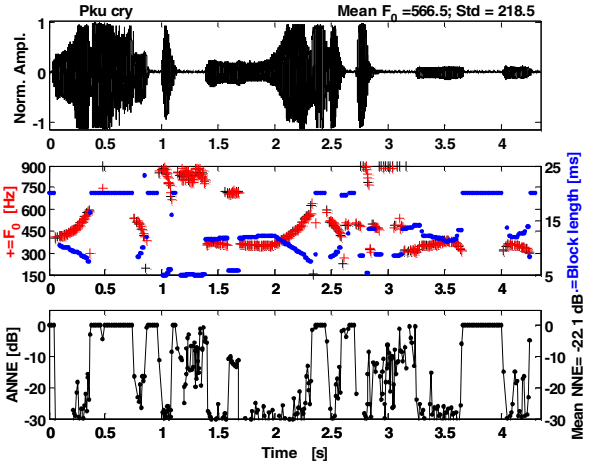

Fig. 7. Normalised signal (black, upper plot), F_0 (red, middle plot) with adaptive window length (blue, middle plot) and noise tracking (black, lowest plot) for a PKU cry.

REFERENCES

[1] A. Fort, A. Ismaelli, C. Manfredi, P. Bruscaglioni, "Parametric and non parametric estimation of speech formants: application to infant cry", *Medical Engineering and Physics*, vol.18, n.8,pp.677-691, 1996.

[2] A. Fort, C. Manfredi, "Acoustic analysis of new-born infant cry signals", *Medical Engineering and Physics*, vol.20, n.6, pp.432-442, 1998.

[3] K. Wermke, W. Mende, C. Manfredi, P. Bruscaglioni, "Developmental Aspects of infant's Cry melody and Formants", *Medical Engineering and Physics*, vol.24, n.7-8, pp..501-514, 2002.

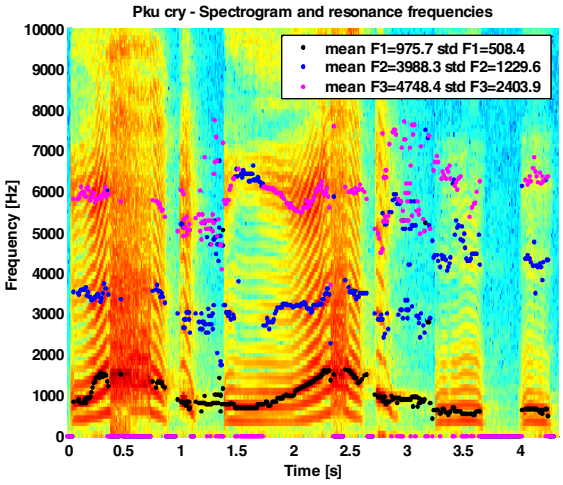

Fig.8. Spectrogram and F1-F3 plot for the PKU cry of Fig.7.

[4] C. Manfredi, K. Wermke, W. Mende, P. Bruscaglioni, "Analysis of the development of acoustic parameters in infant cry", in *Proc. 2nd Europ. Med. Biol. Eng.Conf.*, Vienna, Austria, 4-10 December, 2002, vol.1, pp.474-475.

[5] C. Manfredi, W. Mende, P. Bruscaglioni, K.Wermke, "Resonance development and formant tuning phenomena in infant's crying", in *Proc. 3rd Int. Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications*, Firenze, Italy, 10-12 December 2003, pp. 35-38.

[6] R. Nicollas, M. Ouaknine, A. Giovanni, J. Berger, J.P. To, D. Dumoulin, J.M. Triglia, " Physiology of vocal production in the newborn", in *Proc. 3rd Int. Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications*, Firenze, Italy, 10-12 December 2003, pp. 51-54.

[7] D. Escobedo, S. Cano, E. Collo, L. Regueiferos, L. Capdevila, "Rising shift of pitch frequency in the infant cry of some pathologic case", in *Proc. 2nd Int. Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications*, Firenze, Italy, September 2001 (CD-ROM).

[8] J.D.Markel,"The SIFT algorithm for fundamental frequency estimation", *IEEE Trans. Audio & Electroac*,20, pp.367-377, 1972.

[9] C. Manfredi, M. D'Aniello, P. Bruscaglioni, A. Ismaelli, "A Comparative Analysis of Fundamental Frequency Estimation Methods with Application to Pathological Voices", *Medical Engineering and Physics*, vol.22, n.2, pp.135-147, 2000.

[10] H. Kasuya, S. Ogawa, K. Mashima, S. Ebihara, "Normalised Noise Energy as an Acoustic Measure to Evaluate Pathologic Voice", *J. Acoust. Soc. Am.*, vol. 80, n.5, p.1329-1334, 1986.

[11] C. Manfredi, "Adaptive Noise Energy Estimation in Pathological Speech Signals", *IEEE Trans. Biomed. Eng.*, 47, p.1538-1542, 2000.

[12] S.L. Marple, *Digital spectral analysis with applications*, Englewood Cliffs, NJ, U.S.A.: Prentice Hall, 1987.

[13] J.D. Markel, A.H. Gray, *Linear prediction of speech*, Berlin, DE: Spriger-Verlag, 1982.

[14] K.Wermke, W.Mende, C.Manfredi, P.Bruscaglioni, A.Stellzig-Eisenhauer,"Tuning phenomena of Melodies and Resonance frequencies ('formants') in infants' pre-speech utterances", ASA Conference, 2005, March 2005,Vancouver, Canada.

[15] K.Wermke, W. Mende, A. Kempf, C. Manfredi, P. Bruscaglioni, A. Stellzig-Eisenhauer, "Interaction patterns between melodies and resonance frequencies in infants' pre-speech utterances", in *Proc. 4th Int. Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications*, Firenze, Italy, October 2005, pp. 87-190.

[16] http://www.disat.unimib.it/bioacoustics/it

[17] Kay Elemetrics Corp., "Multi Dimensional Voice Program (MDVP). Operations manual", 1994.

Authorized licensed use limited to: Universita degli Studi di Firenze. Downloaded on December 22,2020 at 17:52:14 UTC from IEEE Xplore. Restrictions apply.