Spectrogram-based fundamental frequency tracking of spontaneous cries in preterm newborns

Bertille Met-Montot
Univ Rennes, Inserm, LTSI - UMR 1099
Rennes, France
bertille.met-montot@univ-rennes1.fr

Sandie Cabon
Univ Rennes, Inserm, LTSI - UMR 1099
Rennes, France
sandie.cabon@univ-rennes1.fr

Guy Carrault
Univ Rennes, Inserm, LTSI - UMR 1099
Rennes, France
guy.carrault@univ-rennes1.fr

Fabienne Porée
Univ Rennes, Inserm, LTSI - UMR 1099
Rennes, France
fabienne.poree@univ-rennes1.fr

***Abstract*—Cry analysis of preterm newborns has proven to be relevant for prediction of pathologies or for comparison with full-term newborns. In this paper we propose a new approach for the automated detection and tracking of the fundamental frequency in cries, based on the processing of the spectrogram. A first step automatically detects the frequency bounds including the fundamental frequency along each cry. Then, the tracking of the fundamental frequency is obtained after a contour detection. Results showed that this new approach allows to process efficiently all types of cries. This whole procedure applied to a database including 1889 cries from 14 babies, at term-equivalent age, highlighted differences between extremely, very and late preterm as well as full-term newborns. In addition, we observed a decrease of the mean fundamental frequency with increasing gestational age, a result in accordance with the literature.**

***Index Terms*—prematurity, monitoring, cry analysis, fundamental frequency tracking, melody**

I. Introduction

Worldwide, 15 million babies are born prematurely each year and this number is increasing [1]. Therefore prematurity remains a main public health concern. As most of death or injury could be prevented with routine interventions, it is essential to witness newborns' neonatal development in daily care. The European project Digi-NewB aims to address this point by developing a new generation of monitoring system in Neonatal Intensive Care Unit (NICU), that will combine signals from different sources, including electro-physiological, video and audio data.

This work deals with sound aspects and has for first objective to automatically extract and characterize preterm newborn spontaneous cries in long recordings. If cry analysis of preterm newborns has proven to be relevant for prediction of pathologies or for comparison with full-term newborns in a lot of studies (see [2] for review), the analysis of spontaneous cries in newborns is recent [3]. In general, adopted approaches rely on the computation of features in two domains: time and frequency [4]. Yet, the most relevant clinical parameter to date is the fundamental frequency (F_0) and its trend over a cry (also called melody). A lot of F_0 estimation methods have been conducted for adult voices but the topic remains an open-issue regarding children crying in light of recent papers on the subject [5]–[8].

Indeed, fundamental frequency estimation in newborn cries is a highly difficult task because of their great variability and irregularity leading to a wide range of frequencies. Existing methods generally search F_0 in a fixed frequency band, but no consensus exist concerning its bounds. This band needs to be large enough in order to cover all possible types of cries. Conversely, the wider is the frequency range, the greater is the risk of false F_0 estimation in favour of high-energy harmonics (formants). In practice, this band is approximately located in the interval [200, 1000] Hz, but other values can be encountered. Moreover, it is worthwhile specifying that some cries (called hyperphonation) have a fundamental frequency beyond 1000 Hz [9].

Several types of approaches have been employed for the estimation of the fundamental frequency: in time domain from the autocorrelation function, in spectral domain from the spectrogram or in the cepstral domain [2]. Furthermore, additional strategies have been developed to track the evolution of F_0 along a cry and several different melodies have been identified according to their shape (falling, rising, plateau, symmetric or complex) [6], [7].

The purpose of this communication is to propose a new method for the fundamental frequency tracking of spontaneous cries in preterm newborns. The first step is based on the seek in the spectrogram of an adaptive frequency range of analysis for each cry, in order to reduce the band to its minimum and avoid risk of jump to higher harmonics. Once the area covered by the fundamental frequency is well delimited, the second step consists in extracting the melody of the cries using a contour detection. Related approach has been observed in the processing of mammals whistles [10], but has never been encountered with newborn cries.

The paper is organized as follows. The processing steps of the proposed methodology are described in Section II. In Section III, results computed over 1889 cries show the F_0 tracking performed over different melody patterns but also study the evolution of the F_0 values as a function of the

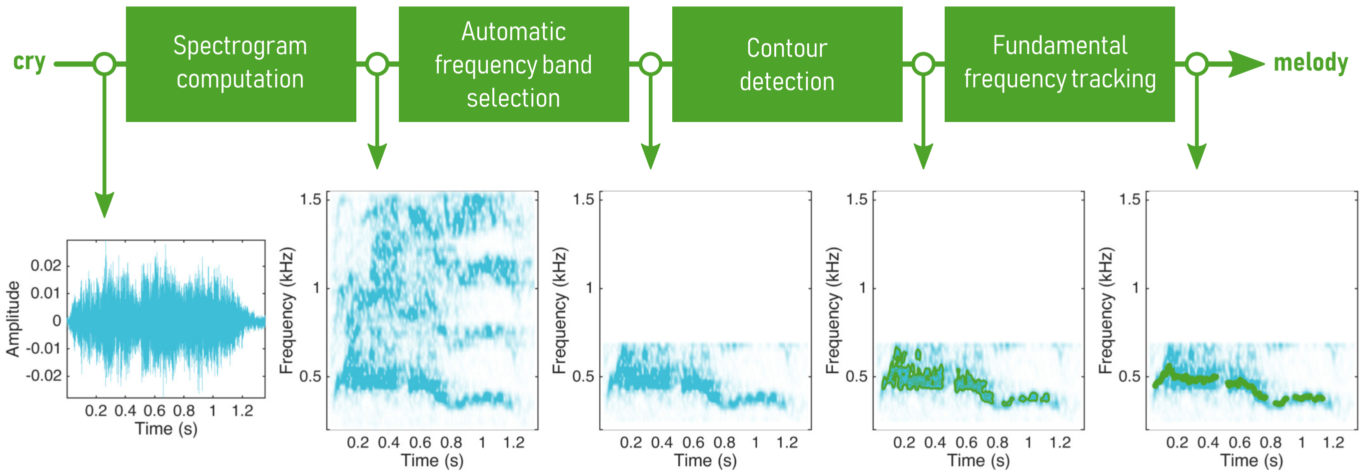

Fig. 1: Illustration of the frequency estimation algorithm processing steps.

severity of the prematurity. Finally, conclusion is presented as well as suggestions and direction for further research.

II. Methods

This section describes the processing steps to automatically track fundamental frequency of newborn cries within a non-fixed analysis band using spectrograms. For that purpose, four main steps, illustrated in Figure 1, are involved:

- Spectrogram computation, visual representation of the spectrum of frequencies of cry unit as it varies with time;
- Automatic frequency band selection, narrowing the image to be analyzed around the frequencies of interest;
- Contour detection in two steps: computation of all the contours in the spectrogram and selection of the relevant ones;
- Fundamental frequency tracking over the cry.

A. Spectrogram computation

First of all, the cry unit is filtered by a Kaiser Window FIR (Finite Impulse Response) filter with cut-off frequencies: 250-1500 Hz. This band corresponds to a wide variability range of the expected cry fundamental frequency. Then, spectrogram of the signal is computed using short-time Fourier transform of successive 0.04 ms long (1000 samples) Hamming-windowed frames with 90% overlap and a sample rate of the cry sampling frequency (24 kHz). The configuration provides n =53 frequency bins per frame of the spectrogram with a frequency resolution of 23.4 Hz and a time resolution of 4.2 ms.

B. Automatic frequency band selection

It can be decomposed into four steps illustrated in Figure 2.

1) Spectrogram' maximums extraction: For each frequency row of the spectrogram, maximum point is extracted. Result is called the maximum curve (c_{max}) which is, therefore, of the same size n than the frequency vector.

2) Local maxima detection: Local maxima are found in the c_{max} curve. In order to avoid the detection of noise peaks among the main frequency components, only peaks with an amplitude greater than 5% of the maximum amplitude and more than 300 Hz apart are retained.

3) Local minima detection: Local minima are also found in the c_{max} curve. In order to avoid detection of irrelevant minima, a threshold is chosen as 20% of the amplitude of the selected maximum peak. Local minima surrounding this peak are chosen to be the new limits for the chosen frequency band. When the first maximum has been detected close to the initial bound (250 or 1500Hz), one surrounding minimum might be missing. In this case, the missing bound is set to the initial corresponding one (see example in Figure 2).

4) Spectrogram filtering: Spectrogram values with corresponding frequency above or under the new limits are set to 0.

C. Contour detection

Here two steps are performed: contour computation and contour selection.

Contour computation: Isolines of the filtered spectrogram are computed, as illustrated in Figure 3. For each detected isoline, two types of variable are returned:

- the height-level,
- and the position coordinates.

Contour selection: First, isolines are computed over cross-sections of the spectrogram relief. Therefore, main spectral components contours are obtained by selecting the low-height isolines. Moreover, as the melody tends to be continuous over the cry unit width, contours of duration less than $\Delta_{dur} = 0.05s$ are disregarded. At last, contours included inside another contour are neglected.

D. Fundamental frequency tracking

Fundamental frequency tracking involves computation of the contour trends and concatenation of the relevant ones.

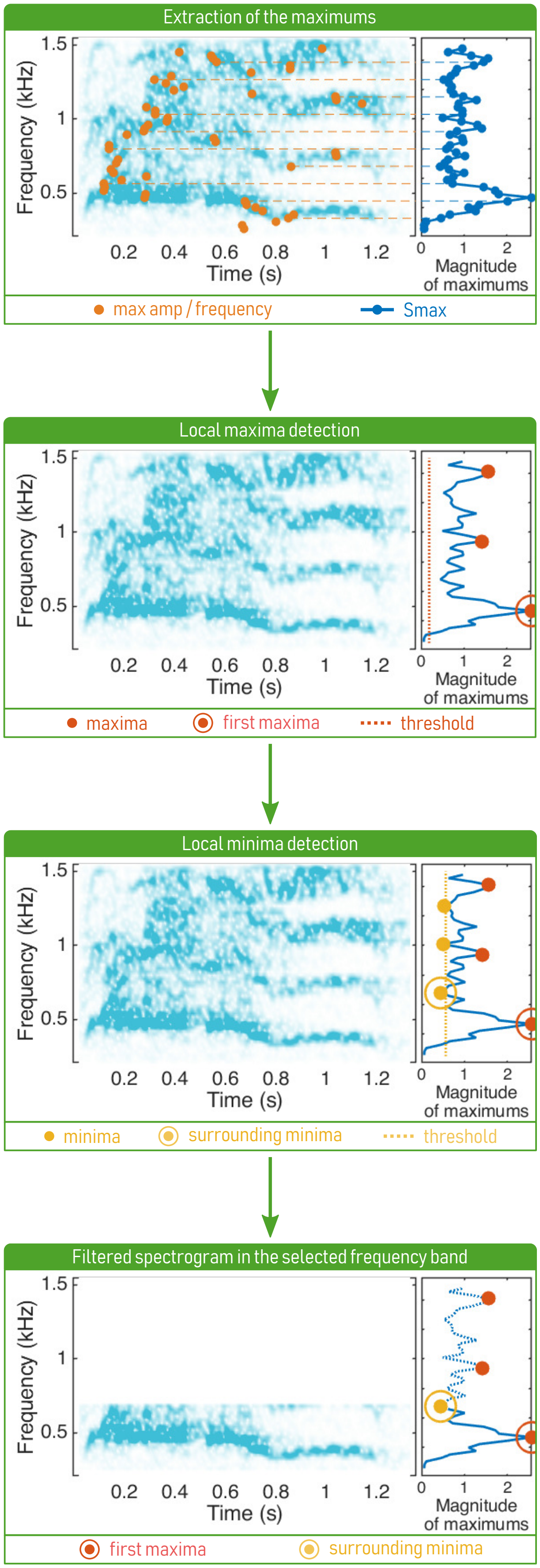

Fig. 2: Illustration of intermediate steps for automatic frequency band selection.

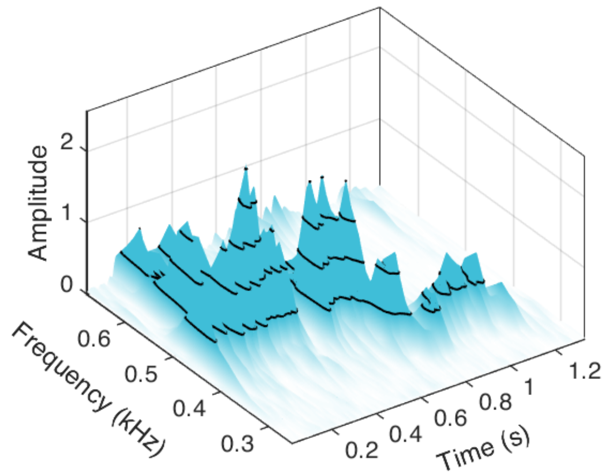

Fig. 3: Illustration of isolines computation over the filtered spectrogram.

Contour trends computation: To begin with, points of one contour are arranged in two vectors containing, respectively, top and bottom values 4(a)). The smallest step that can contain at least one point of the upper vector and one point of the lower one is calculated and defined for all cries as $\Delta_{step} = 0.005s$. For each step, average values of the upper and lower points are calculated separately 4(b)). Finally, the contour trend is the mean value of the just computed high and low averages (Fig. 4(c)). This procedure is applied to all previously selected contours.

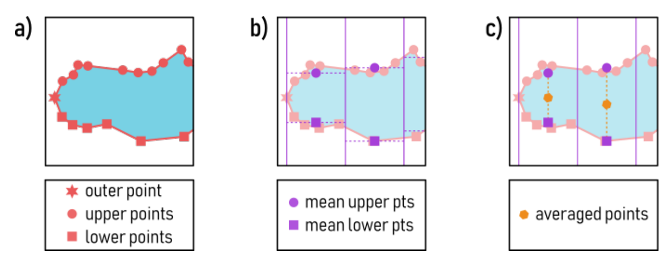

Fig. 4: Illustration of contour trend computation steps : a) arrangement of top and bottom points; b) separated top and bottom averages for each step; c) mean of the averages.

Concatenation: Finally, a vector of the size of the cry unit is created and filled with :

- Contour trends as they appear over time;
- Not-a-number values otherwise.

When two or more contours are temporally overlapping, only the wider one is kept.

III. Results

A. Database

In the framework of the European project DigiNew-B, cry recordings were performed in six french hospitals. Spontaneous cries of each baby were recorded by a multimodal recording system containing an omni-directional Knowles Acoustic microphone (FG-23329-P07) sampled at 24 kHz. Microphones were set less than one meter away from the newborn's head.

A total of 14 babies (six girls and eight boys) were enrolled in this study including:

- three extremely preterm (EP): gestational age less than 28 weeks, weight at birth less than 1000 g, with a total of 486 cries,
- four very preterm (VP): gestational age between 28 and 32 weeks, weight at birth less than 1500 g, with a total of 290 cries,
- four late preterm (LP): gestational age between 32 and 37 weeks, weight at birth less than 3000 g, with a total of 1039 cries,
- and three full-term healthy newborns (FT): gestational age greater or equal to 37 weeks, weight at birth greater than 2400 g with a total of 74 cries.

Recordings were performed at term-equivalent age, i.e. at post-menstrual age between 37 and 42 weeks. Approximately 120 hours were processed and a total of 1889 cries were automatically extracted by a classification approach based on a k-Nearest Neighbors (k-NN) described in [11].

B. Band selection and melody tracking

The whole process was applied to the 1889 cries of the database.

Figure 5 shows the result of the fundamental frequency tracking (in green) for five different melody shapes and for a hyperphonation cry. Horizontal lines shows the frequency bands automatically detected, delimiting the area of research for the contour detection. It can be observed that, for each type of melody, the band is correctly located over the interesting area and that the fundamental frequency is well-tracked. In the last example (hyperphonation), the high value of the upper bound can be noticed.

To supplement these results, in Table I, we present the distribution of the bound values automatically detected in terms of lower, upper and bandwidth values. It can be noted for example that the range of the upper bound goes up to 1500 (that might correspond to a hyperphonation), which is higher than the value generally chosen in the literature. Furthermore, the mean bandwidth is equal to 288 Hz with a minimal value at 55 Hz (that might correspond to a plateau melody), which is narrower than in the literature.

TABLE I: Lower, upper and bandwidth values of the automatic frequency bands.

	mean	s.d	range
lower bound F_{low} (Hz)	304	70	250-797
upper bound F_{up} (Hz)	592	177	305-1500
bandwidth $F_{up} - F_{low}$ (Hz)	288	161	55-1250

C. Clinical application

In this section we propose to evaluate the evolution of the fundamental frequency as a function of the prematurity severity. For this purpose, after the fundamental frequency tracking, F_0 values were averaged for each cry. Then, for each baby, means of these values were computed. Figure 6 shows that the fundamental frequency obtained is higher for the extremely preterm newborns and tends to decrease as the gestational age increases. These results shows a good agreement with those in [3], the only paper addressing this question to our knowledge up to date. It can be noted that the last two full-term babies have a higher fundamental frequency than the trend. These results may be due to the low number of analyzed cries that probably don't represent all the different kinds of crying that these babies can produce.

IV. Discussion and Conclusion

In this paper we proposed a new approach for the automated detection and tracking of the fundamental frequency in cries of premature newborns, based on the processing of the spectrogram. A first step automatically detects the frequency band including the fundamental frequency along each cry. Then, the tracking of the fundamental frequency is obtained after a contour detection. Such a strategy has never been encountered in the processing of newborn cries [2].

Results showed that this new approach allows to process efficiently all types of cries. This whole procedure applied to a database including 1889 cries from 14 babies, at term-equivalent age, highlighted differences between extremely preterm, very preterm, late preterm and full-term newborns. In addition, we observed a decreasing of the mean fundamental frequency with increasing gestational age, a result in accordance with the literature.

Further works will concern the processing of cries from a higher number of babies, with diverse post-menstrual ages.

Acknowledgment

Results incorporated in this publication received funding from the European Union's Horizon 2020 research and innovation program under grant agreement N°689260 (Digi-NewB project).

References

[1] OMS, https://www.who.int/fr/news-room/fact-sheets/detail/preterm-birth.

[2] S. Cabon, F. Porée, A. Simon, O. Rosec, P. Pladys, and G. Carrault, "Video and audio processing in paediatrics: a review," *Physiological measurement*, vol. 40, no. 2, p. 02TR02, 2019.

[3] Y. Shinya, M. Kawai, F. Niwa, and M. Myowa-Yamakoshi, "Preterm birth is associated with an increased fundamental frequency of spontaneous crying in human infants at term-equivalent age," *Biology Letters*, vol. 10, no. 8, 2014.

[4] L. L. LaGasse, A. R. Neal, and B. M. Lester, "Assessment of infant cry: Acoustic cry analysis and parental perception," *Mental Retardation and Developmental Disabilities Research Reviews*, vol. 11, no. 1, pp. 83–93, 2005.

[5] S. Orlandi, A. Guzzetta, A. Bandini, V. Belmonti, S. D. Barbagallo, G. Tealdi, S. Mazzotti, M. L. Scattoni, and C. Manfredi, "AVIM - A contactless system for infant data acquisition and analysis: Software architecture and first results," *Biomedical Signal Processing and Control*, vol. 20, pp. 85–99, 2015.

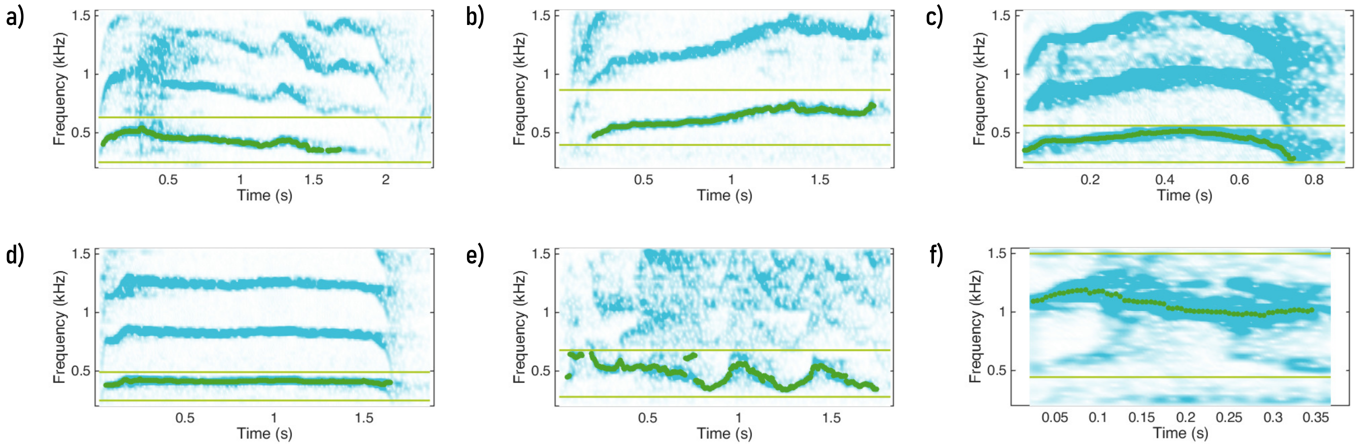

Fig. 5: Examples of fundamental frequency tracking (in green) results for different melody shapes: a) falling, b) rising, c) rising-falling, d) flat, e) complex, and for f) an hyperphonation cry. Horizontal lines shows the frequency bands automatically detected.

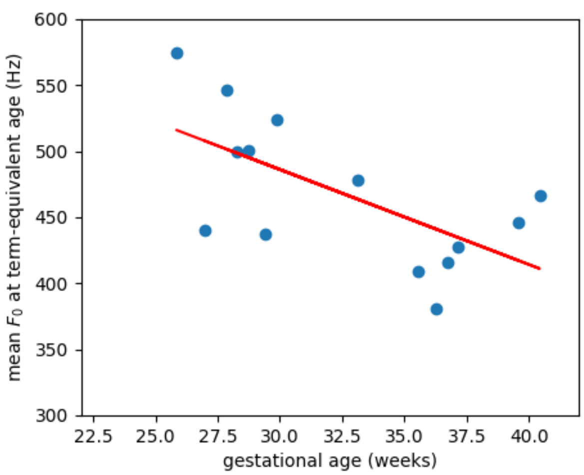

Fig. 6: Scatter plot showing the relationship between gestational age and mean fundamental frequency (F_0) of spontaneous cries at term-equivalent age for all infants.

[6] S. Orlandi, A. Bandini, F. Fiaschi, and C. Manfredi, "Testing software tools for newborn cry analysis using synthetic signals," *Biomedical Signal Processing and Control*, vol. 37, pp. 16–22, 2017.

[7] C. Manfredi, A. Bandini, D. Melino, R. Viellevoye, M. Kalenga, and S. Orlandi, "Automated detection and classification of basic shapes of newborn cry melody," *Biomedical Signal Processing and Control*, vol. 45, pp. 174–181, 2018.

[8] Y. Kheddache and C. Tadj, "Identification of diseases in newborns using advanced acoustic features of cry signals," *Biomedical Signal Processing and Control*, vol. 50, pp. 35–44, 2019.

[9] B. M. Lester and C. Z. Boukydis, *Infant crying: Theoretical and research perspectives*. Springer, 1985.

[10] M. A. Roch, T. Scott Brandes, B. Patel, Y. Barkley, S. Baumann-Pickering, and M. S. Soldevilla, "Automated extraction of odontocete whistle contours," *The Journal of the Acoustical Society of America*, vol. 130, no. 4, pp. 2212–2223, 2011.

[11] S. Cabon, "Monitoring of premature newborns by video and audio analyses," Ph.D. dissertation, Université de Rennes 1, 2019.