BABIES' VOICES: A COLLABORATIVE RESEARCH PROGRAM ON THE AUTOMATED ACOUSTICAL ANALYSIS OF THE PRETERM NEWBORN CRY

R. Viellevoye[1], D. Melino[2], S. Orlandi[2], G. Pieraccini[2], G. Donzelli[3], A.Torres-García[4], C.A. Reyes García[4],C. Manfredi [2]

[1]Department of Pediatrics, Neonatal Intensive Care Unit, University of Liege, Liege, Belgium
[2]Department of Information Engineering, Università degli Studi di Firenze, Firenze, Italy
[3]Department of fetal and neonatal medicine, Università degli Studi di Firenze, Firenze, Italy
[4]Computer Science Department, Instituto Nacional de Astrofisica, Optica y Electronica (INAOE), Puebla, Mexico
renaud.viellevoye@chrcitadelle.be;donatellamelino@gmail.com;orlandisilvia85@gmail.com;
gianandrea.pieraccini@stud.unifi.it;alejandro.torres@ccc.inaoep.mx;kargaxxi@inaoep.mx;
claudia.manfredi@unifi.it

Abstract: **The aim of this collaborative work is to provide the automated assessment of the melodic shape of the newborn cry with the BioVoice software tool. The method was tested on synthetic signals with 100% matching. Acoustical parameters of cries obtained from preterm and term newborns in Liege (Belgium) and Firenze (Italy) were estimated with BioVoice. The automated classification was first compared to the perceptual (visual) analysis considered as the gold standard on a set of healthy at term newborns with a matching up to 85%. Then, significant differences were found between at term and preterm babies up to 85%. Our study suggests that some melodic characteristics of the newborn cry could be detected to predict the belonging to term/preterm group of patients with an acceptable accuracy.**

Keywords: **Newborn, cry, preterm newborn, cry melody, automated analysis.**

I. INTRODUCTION

The acoustical analysis of infant crying is a promising non-intrusive and cheap approach as an aid to early diagnosis of neurological disorders [1]. The most relevant clinical parameter is the fundamental frequency f0, which reflects the regularity of the vibration of the vocal folds of the newborn. To date, the analysis of the infant crying melody, that is the temporal trend of f0 over time, is carried out by the paediatrician/neurologist with a perceptive examination based on listening to the cry and visually inspecting the fundamental frequency f0 shape. This approach is not widespread as the procedure is operator-dependent and requires a considerable amount of time often prohibitive in daily clinical practice [2]. The aim of our collaborative work is to provide a fast and fully automated method for assessing the melody shape of the newborn cry that could be used routinely to assess at risk newborns such as preterm infants.

Indeed, preterm newborns are at high risk for developing cerebral palsy, cognitive impairment, behavioural difficulties and/or neurosensory disabilities [3]. Early diagnosis of neurological impairment is crucial to initiate neuromodulatory interventions supporting cerebral plasticity, while delayed recognition increases the risk for co-morbidities and poor outcome. Systematic automated analysis of the newborn cry performed at term-equivalent age could thus help clinicians to identify particular pattern of cry that could be predictive of poor neurological outcome.

Furthermore, cry is a developmental process influenced by acoustical environment and stimulations. Premature birth causes a sudden transition from the physiological intrauterine environment towards the noisy world of the neonatal intensive care unit (NICU), depriving the baby of the biological maternal voice. The cry, as the first way of communication experienced by the newborn to elicit caretaking, could be negatively influenced by this modified environment. Developmental care is a broad category of interventions designed to minimize the stress of the NICU environment [3, 4]. These interventions may include elements such as control of external auditory stimuli and facilitation of parental involvement. Again, a systematic analysis of the newborn cry could identify adequate strategies that could minimize the impact of the postnatal environment on neurodevelopment, in particular speech and language acquisition.

II. METHODS

BioVoice is a multi-purpose voice analysis tool developed under Matlab® at the Biomedical Engineering Lab., Department of Information

Claudia Manfredi (edited by), *Models and analysis of vocal emissions for biomedical applications : 10 th international workshop : December 13-15, 2017*, ISBN 978-88-6453-606-4 (print) ISBN 978-88-6453-607-1 (online)

Engineering, Università degli Stud di Firenze [5-8]. Newborn infant cry recordings, that may last even several minutes, are made up by a number of "cry units" (CUs), commonly of different length and separated by "silence" frames. A CU is defined here as a high-energy frame lasting >260ms. The purpose is to automatically perform the classification of the cry melody of each CU detected within a recording. Detection of CUs is performed using a robust Voiced/Unvoiced (V/UV) detection procedure with variable energy thresholds that avoids incorrect splitting of a single event into several intervals [8,9].

BioVoice performs the estimation of several acoustic parameters that gained great scientific interest in the last years, including the fundamental frequency f0 and the first three resonance frequencies of the vocal tract, along with their std. To successfully assess the CU melodic shape, the first step is the accurate removal of the f0 outliers, that are irregularly distributed within the CU and that could distort the shape identification. This is performed through several steps each one based on specific conditions. Afterwards, BioVoice allows the classification among 5 basic melodic shapes: Plateau (P), Rising (R), Falling (F), Symmetric (S) and Complex (C). To perform the classification each CU is subdivided into 12 equally spaced time segments each one described by its f0 mean value plus the first f0 value (13 Perc). To find the best number of segments, the same procedure is applied with 20 equally spaced points plus the first f0 value (21 Perc).

First, the method was tested on synthetic signals. To this aim a new synthesizer was developed made up of a pulse train generator and a vocal tract filter. To get variable frequency control vectors capable of synthesizing the 5 basic melody shapes, a spline interpolation was implemented with a time varying f0[n]. Settings allowed obtaining melodic shapes close to newborn cry, within the range 400 Hz - 650 Hz, and f0 mean between 500 Hz and 550 Hz. To test the proposed method synthetic white noise of increasing amplitude (0%, 1%, 5% and 10% of the signal maximum amplitude) was added to the synthesized signals [10].

The automated classification has then been applied to cry recordings coming from at term healthy newborns. Recordings were performed with a Shure SM58 microphone and a Tascam US144MK2 sound card. The microphone was kept at fixed distance (15 cm) from the newborn's mouth and cry signals were recorded at the awakening of the baby before feeding, thus they were supposed to be feeding cries. Results have been compared to the perceptual analysis (considered as the gold standard) performed by trained raters. A melodic shape was classified as belonging to one of the five categories (P, F, R, S and C) if at least two out of the three raters agreed with the same shape. If the three raters disagreed, the CU was defined as "borderline" and temporarily eliminated.

Once the method was validated, it was applied to cry episodes recorded from both preterm and at term newborns in Liege, Belgium (Neonatal Intensive Care Unit, CHR Liege) and Firenze, Italy (Neonatal Intensive Care Unit, Meyer Children Hospital and Neonatal Unit, Azienda Sanitaria Ospedale San Giovanni di Dio) at term-equivalent age. Percentage of the different melodic shapes and several acoustic parameters such as mean duration of CUs, mean fundamental frequency f0 and standard deviation were calculated by automated analysis using BioVoice. Perceptual analysis was performed by trained raters. Results were compared between Belgian and Italian newborns as well as between at term and preterm babies among the entire and local populations. Statistical analysis was performed using a t-test and statistical significance was considered for p-value <0.05.

Finally, a classification technique based on the 10-fold cross-validation method was applied to the set of estimated melodic shapes in order to obtain the automated classification of a cry episode as belonging to term vs preterm infant [11].

III. RESULTS

A. Synthetic data

Both 13 and 21 Perc were tested. For all the synthetic melodic shapes and all the levels of added noise the fitting procedure gave the best results with the 4th order polynomial for which the R-square parameter R^2 (ranging between 0 and 1 where 1 is the best fitting):

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \quad (1)$$

was found as: $R^2 \cong 0.993$ for 13 Perc and $R^2 \cong 0.995$ for 21 Perc. Both perceptual and automated melodic classifications were successful at 100%: all melody shapes were correctly classified [10, 11].

B. Real data

CUs from 6 healthy at term newborns (3 male and 3 female) were recorded in the maternity unit in CHR Liège (Belgium). Each cry episode lasts 1-2 minutes and consists of several CUs. A total of 466 CUs was collected. 48 CUs were excluded because of the absence of consensus between the three raters. Moreover, other 116 CUs were not perceptually recognized by any rater as belonging to one of the five basic shape considered here and were also excluded from analysis.

Considering all the 5 shapes, the automated analysis with 21 percentiles matched the perceptual one in 89.5% of cases while only in 80.3% with 13 percentiles. Excluding the C shape, the match increases to 96.7% with 21 percentiles and to 89.3% with 13 percentiles. Therefore, the best number of Perc proved to be equal to 21. We point out that these results concern the case of full agreement among the three raters. Table I summarizes the results.

TABLE I – Automated vs perceptual classification of the melodic shapes. Results for the full agreement among the three raters are presented. Best result: 96.7% match with 21 Perc and the 4 basic shapes P, R, F and S (C excluded).

3/3 raters	**P, F, R, S, C**	**P, R, F, S**
Automated – 13 Perc	80.3%	89.3%
Automated – 21 Perc	89.5%	**96.7%**

C. Newborn cry melody in term and preterm newborn

A larger data set was analyzed consisting of a total of 9 preterm infants (21 cry episodes and 382 CUs) and 24 term infants (41 cry episodes and 2532 CUs) delivered from Belgian mothers in the maternity and neonatal intensive care units at CHR Liege. Moreover, 9 preterm infants (24 cry episodes and 1787 CUs) and 25 term infants (70 cry episodes and 5187 CUs) delivered from Italian mothers were recorded in the neonatal intensive care unit at Meyer Children Hospital and San Giovanni di Dio Hospital, Firenze, respectively. Melody of each CU was then classified as belonging to one of the five main categories (P, F, R, S and C) or to additional categories (LU – Low Up, UL – Up Low, FS – Frequency step, D – Double, U – Unstructured, NC – Not a cry or O – Other) using both automated and perceptual analysis. Details about shapes can be found in [10, 11].

Matching between automated and perceptual analysis was 58%, 65%, 59% and 61% for Belgian preterm, at term, Italian preterm and at term newborns respectively, which was lower than the results obtained in our preliminary study, even when limited to the five main basic shapes. This was mainly due to the low level of experience of the operator(s) that however increased during the testing phase: a deeper training brought to a better matching rate. Moreover, percentage of occurrence of each of these five categories was not statistically different between groups regarding to the method of analysis we used.

The mean value of f0 was not statistically different between term and preterm infant in Belgium ($P = 0.22$), Italy ($P = 0.30$), or from both countries ($P = 0.45$). Mean CU duration was significantly shorter for term newborns than for preterm infants (2,57s vs 5,04s respectively, $p<0.05$). We did not find any significant differences between groups regarding to the percentage of each of the five main categories considering both automated and perceptual analysis: Belgian vs Italian preterm infants (Pauto=0.25; Pperc=0.40) or term infants (Pauto=0.11; Pperc=0.50), term vs preterm infant in Belgium (Pauto=0.28; Pperc=0.46) and Italy (Pauto=0.44; Pperc=0.23), term vs preterm infants from both countries (Pauto=0.45; Pperc=0.24) or Belgian vs Italian newborns (Pauto=0.10; Pperc=0.45). However, preterm newborns seemed to have a trend in favour of more C and NC shapes as compared to term infants who more frequently present the P pattern, while other categories show similar frequencies in each population. Table II summarizes these results.

TABLE II – Percentage of C, NC and P shapes in Belgian, Italian and overall at term and preterm newborns.

Shape	**Population**	**Preterm**	**Term**
Complex	Liege	32.5%	23.7%
	Firenze	29.8%	25.4%
	Total	**33%**	24.9%
Not a cry	Liege	7.9%	6.5%
	Firenze	9%	4.6%
	Total	**8.8%**	5.3%
Plateau	Liege	11.3%	28.7%
	Firenze	17.8%	28.3%
	Total	16.6%	**28.4%**

Finally, results obtained with automated classification of cry episodes using the 10-folds cross-validation method were encouraging. The method, applied to the whole set of melodic shapes, was able to discriminate between preterm or at term infants with an accuracy ranging from 74.47 to 85.48%. (Table III).

Table III – Accuracy of the automated classification to discriminate between at term and preterm newborns

PRETERM vs AT TERM	**Correct**	**Not correct**
Liege	**85,48%**	14.52%
Firenze	74,47%	25.53%
Liege + Firenze	80.77%	19.23%

IV. Discussion

This work presents a methodological approach to the classification of the newborn cry melody whose features are considered clinically relevant for the assessment of the neurological status of the newborn at birth. The method has the advantage of being totally contactless and thus applicable also to very delicate subjects such as newborn babies are. The required equipment is low cost and easy to use and therefore easily implementable in any paediatric clinic both public and private.

Results obtained with the BioVoice software for the automated classification of newborn cry compared to the perceptual analysis are encouraging. We assumed as gold standard the blinded perceptual (visual) analysis made by a panel of trained raters. The automated analysis with 21 percentiles better matched the perceptual one than the automated analysis with 13 percentiles. However, imperfect matching still remains, especially if more than the five basic categories of melodic shapes are considered. We point out that the inter-observer reliability was not perfect in our study, maybe due to different levels of experience between raters. For a more reliable gold standard reference, future developments should consider a larger and more experienced group of evaluators.

Comparison of the main acoustical characteristics and the pattern of melodic shapes between at term and preterm newborns did not reach statistical significance, except for the mean duration of CU that was found shorter in term newborns. However, our study shows some trends between at term and preterm babies in the percentage of some categories of melodic shapes.

Finally, the results obtained with the automatic classification suggest that some characteristics of the newborn cry could be selected to predict the belonging to a defined group of patients with an acceptable accuracy. Systematic recording of cry from preterm newborn at term-equivalent ages is currently under progress. This could make it possible to retrospectively characterize the acoustical parameters and melodic patterns of infants with typical development compared to infants with abnormal neurological development such as cerebral palsy, cognitive or speech delay in order to develop a predictive model based on the most relevant features.

V. Conclusion

This methodological work is a first step towards the establishment of procedures for the analysis of infant cry. The overall matching percentage between automated and perceptual analysis was found around 60%. This was mainly due to the low level of experience of the operator(s) that however increased during the testing phase: a deeper training brought to a better matching rate.

In future work the automated analysis will be improved by a refined control on the f0 shape variations and on the estimation of frequencies steps.

These results could be compared with the patients' follow-up, especially for preterm babies, in order to track their development and to study relationships between the automated/perceptual results and possible diseases in the central nervous system for such delicate patients. Therefore, the automated analysis of newborn cry melody could become a reliable support to the "time-consuming and subjective perceptual analysis and, if properly assessed, could even replace it and become part of clinical standards in the neonatal screening.

References

[1] Corwin MJ, Lester BM, Golub HL. The infant cry: what can it tell us, Curr. Probl. Pediatr. 1996 Oct;26(9):325-34.

[2] Larroque B, Ancel PY, Marret S, Marchand L, Andre M, Arnaud C, Pierrat V, Roze JC, Messer J, Thiriez G, Burquet A, Picaud JC, Breart G, Kaminski M; EPIPAGE Study group. Neurodevelopmental disabilities and special care of 5-year-old children born before 33 weeks of gestation (the EPIPAGE study): a longitudinal cohort study. Lancet 2008; 371(9615): 813-820.

[3] Marx V, Nagy E. Fetal behavioural response to maternal voice and touch. PlosOne 2015;10(6):e0129118

[4] Als H. Developmental care in the newborn intensive care unit. Curr Opin Pediatr. 1998 Apr;10(2):138-42

[5] Wermke, K., Mende, W., Manfredi, C., Bruscaglioni, P., "Developmental aspects of infant's cry melody and formants", (2002) Medical Engineering and Physics, 24 (7-8), pp. 501-514. doi: 10.1016/S1350-4533(02)00061-9

[6] Manfredi, C., Tocchioni, V., Bocchi, L., "A robust tool for newborn infant cry analysis", Conf Proc IEEE Eng Med Biol Soc 2006 ;1:509-12

[7] Manfredi, C., Bocchi, L., Orlandi, S., Spaccaterra, L., Donzelli, G.P., "High-resolution cry analysis in preterm newborn infants", (2009) Medical Engineering and Physics, 31 (5), pp. 528-532. doi: 10.1016/j.medengphy.2008.10.003

[8] Orlandi S, Manfredi C, Bocchi L, Scattoni ML, "Automatic newborn cry analysis: a non-invasive tool to help autism early diagnosis.", Conf Proc IEEE Eng Med Biol Soc. 2012;2012:2953-6. doi: 10.1109/EMBC.2012.6346583.

[9] Orlandi S, Dejonckere PH, Schoentgen J, Lebacq J, Rruqja N, Manfredi C. Effective pre-processing of long term noisy audio recordings: An aid to clinical