RELATIONSHIPS BETWEEN NEWBORNS' CRY MELODY SHAPES AND NATIVE LANGUAGE

C.Manfredi[1], G.Pieraccini[1], R.Viellevoye[2], A.Torres-García[3] C.A. Reyes-García[3]
[1] Department of Information Engineering, Università degli Studi di Firenze, Firenze, Italy
[2]Department of Pediatrics, Neonatal Intensive Care Unit, University of Liege, Liege, Belgium
[3]Computer Science Department, Instituto Nacional de Astrofísica, Óptica y Electrónica (INAOE), Puebla, Mexico
claudia.manfredi@unifi.it; gianandrea.pieraccini@stud.unifi.it; renaud.viellevoye@chrcitadelle.be; alejandro.torres@ccc.inaoep.mx; kargaxxi@inaoep.mx

Abstract: **Recent research studies have shown that since the last trimester of pregnancy the human fetus is able to listen to and possibly memorize auditory stimuli from the external world, both as music and language are concerned. In particular, they exhibit a specific sensitivity to prosodic features such as melody, intensity, and rhythm, that are essential for an infant to learn and develop the native language. This paper presents first results concerning the mother language assessment of a set of about 7.500 cry units coming from French, Arabic and Italian mother-tongue healthy at term newborns. A number of acoustical parameters and 12 different melodic shapes are detected with the BioVoice SW tool and their classification is performed with Random Forest and 4 neuro-fuzzy classifiers. Results show up to 94% differences among the three languages.**
Keywords: **newborn cry melody, mother language, automated acoustical analysis, classification algorithms.**

I. INTRODUCTION

During the last three months of pregnancy the human fetus is able to perceive sounds and distinguish the maternal voice. Adult-like processing of pitch intervals allows newborns to appreciate musical melodies as well as emotional and linguistic prosody and language [1]. Cry is the first means of communication for humans, and is the result of a developmental process influenced by the acoustical environment and stimulations, therefore some studies suggest that the newborn cry melody (the trend of the fundamental frequency with time) could be shaped by the maternal native language [2]. Prosodic features such as melody, intensity, and rhythm are in fact essential for an infant acquiring language dominion [3].

This paper presents first results concerning the mother language assessment of a large set of about 7.500 cry units coming from French, Arabic and Italian mother-tongue newborns. The acoustical parameters and the melodic shapes are detected with BioVoice [3-6] and their classification is performed with Random Forest and 4 neuro-fuzzy classifiers. Results show up to 94% difference among these languages, thus suggesting that newborns pick up acoustic elements of their parents' language before they are even born, and certainly before they start to babble themselves.

II. METHODS

The automated newborn cry analysis is performed with BioVoice, a multi-purpose voice analysis tool developed under Matlab® at the Biomedical Engineering Lab., Department of Information Engineering, Università degli Studi di Firenze [3-6]. Typically newborn infant cry recordings, that may last even several minutes, are made up of a number of "cry units" (CUs) of different length and separated by "silence" frames. A CU is defined as a high energy voiced frame lasting >260ms. With BioVoice the detection of CUs is performed using a robust Voiced/Unvoiced (V/UV) procedure that avoids incorrect splitting of a single event into several intervals [7]. On each CU BioVoice estimates several acoustic parameters, among which the fundamental frequency F0 and the first three resonance frequencies F1-F3, along with their maximum, minimum and standard deviation values, as well as other statistical parameters. It applies autoregressive (AR) parametric techniques, well suited to deal with quasi-stationary high-pitched signals as newborn cries are. After an accurate automated removal of outliers, the melodic assessment of the CU shapes is made through several steps, each one based on specific conditions. BioVoice allows both automated and perceptual classification of a CU among 12 basic melodic shapes: Plateau (P), Rising (R), Falling (F), Symmetric (S), Complex (C), Low-Up (LU), Up-Low (UL), Frequency Step (FS), double (D), Unstructured (U), Not a cry (NC), and Other (O). More details can be found in [8]. Some examples of the above mentioned melodic shapes are reported in Fig.1.

The very simple user interface implemented in BioVoice is shown in Fig.2. Several recordings can be

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)
CC 2017 Firenze University Press

uploaded and analysed sequentially. The plot in the lower part of Fig.2 shows the result of the automated CUs selection (dotted line). Fig.3 shows the interface for the melodic assessment. As an example, a P shape is shown in the picture.

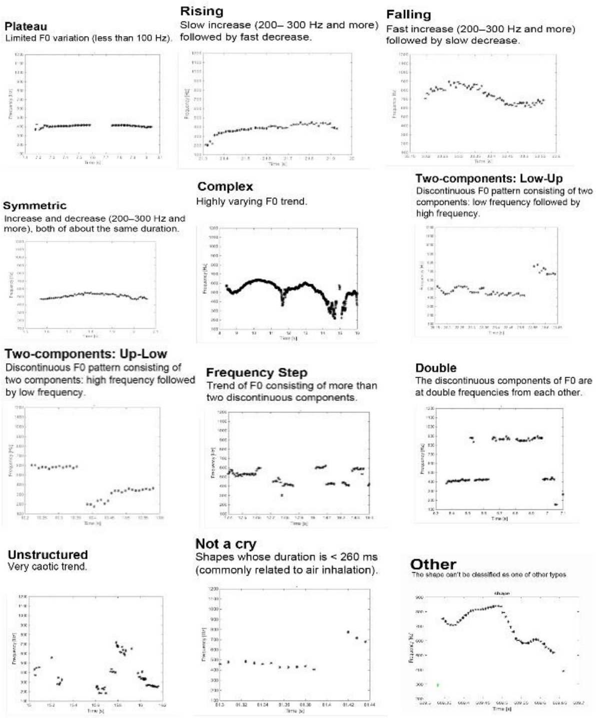

Figure 1 – Examples showing the 12 melodic shapes assessed with BioVoice. For each shape a short description is reported in the legend.

Once the acoustical parameters and the melodic shapes are detected and estimated, the automatic classification consists in detecting which classifier must infer a function (from training data) that allows recognizing new cases (test data) not used during the training process. Specifically, we assessed the Random Forest (RF) classifier, which is an ensemble of classification trees, and 4 neuro-fuzzy classifiers: Adaptive Neuro-Fuzzy with linguistic Hedges (ANFCLH) [9], Adaptive Neuro-Fuzzy with feature selection based on linguistic hedges (ANFCLH-FS) [10], Neuro Fuzzy Classifier (NFC) [11] and Speed-up Scaled Conjugated Gradient Neuro-Fuzzy Classifier (NFC-SCG) [12]. Neuro-Fuzzy models are hybrid systems that combine the capabilities of both representing the knowledge using linguistic expressions of fuzzy systems and learning of neural networks.

Specifically, NFC, NFC-SCG, ANFCLH and ANFCLH-FS create fuzzy rules using the K-means algorithm, whose input membership functions and output functions are later trained as a neural network. The main difference between NFC and NFC-SCG is that the second one implements an improvement for speeding up the scaled conjugate gradient algorithm used for training the NFC classifier. Whereas ANFCLH adds a layer of linguistic hedges for applying to each membership function. Finally, ANFCLH-FS takes advantages of the linguistic hedges for selecting a subset of features, which are used later for the classification stage using ANFCLH.

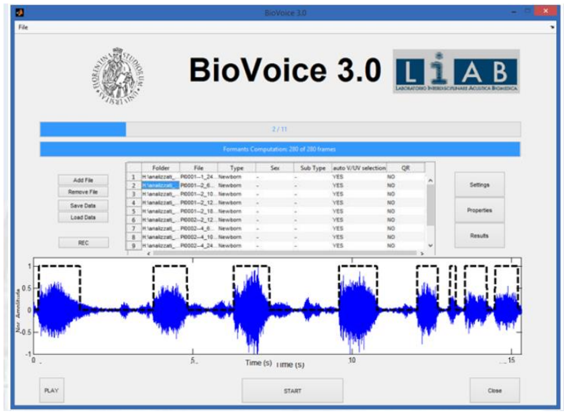

Figure 2 – BioVoice user interface

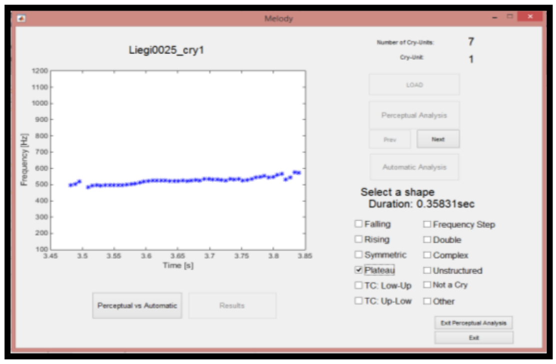

Figure 3 – BioVoice interface for the melodic assessment of each detected CU. A P-shape is shown.

III. RESULTS

For the experiments, we recorded infant cries from healthy at term babies whose mother's native languages are: Arabic, French, and Italian. Specifically, recordings come from two data sets: a set of 24 at term newborns (2532 CUs) from the Children Hospital «La Citadelle», Liege, Belgium (French and Arabic) and 28 at term newborns (5187 CUs) from the San Giovanni di Dio Hospital, Firenze, Italy (Italian). All recordings were made according to the same protocol: a Shure58 microphone was kept at a fixed distance of 20cm from the baby's mouth and connected to a laptop through a Tascam audio board. Recordings, lasting from few seconds to even 1 minute, were made 1-2 days after birth, in a non-noisy environment, before feeding:

therefore, all recorded signals should concern hunger cries.
BioVoice was applied first to raw data, i.e. the overall set of 52 recordings of variable length. Within each recording all the CUs were automatically detected and for each CU about 25 parameters were estimated. Also, the melodic shape of each CU was automatically assessed among the 12 listed in the previous section.
Looking for any relationship between CUs and mother's native language, we assessed the performance obtained from the automatic classifiers. All the infant cry's instances were characterized with the 12 qualitative features described above: falling, rising, symmetrical, plateau, complex, low-up, up-low, frequency step, double, unstructured, not a cry, and other. Also, we tested if adding two quantitative features (mean and standard deviation of the fundamental frequency F0) could help to improve the recognition of native language from infant CUs.
In a first instance, we assessed the performance of the automatic classifiers for the recognition of pairs of languages: Italian vs French, Arabic vs French and Arabic vs Italian. Table 1 shows the performances which are above the chance level for two classes (50%) and show that the use of mixed features is somewhat better than using qualitative features only (for the best performance's cases). Best results show up to 94% of correct classification (Arabic vs Italian).

Table 1. Accuracy percentages obtained for the classifiers using pairs of languages. Q - qualitative features. M - both qualitative and quantitative features.

Classifier	Italian/French		Arabic/French		Arabic/Italian	
	Q	M	Q	M	Q	M
RF	**90.52**	89.47	67.64	70.58	91.35	91.35
NFC	82.22	87.71	72.5	**77.5**	85.13	**93.75**
NFC-SCG	82.22	87.71	72.5	80	90.13	**93.75**
ANFCLH	82.22	85.51	72.5	**77.5**	88.88	87.5
ANFCLH-FS	85.33	88.82	69.16	72.5	88.75	88.88

Moreover, we evaluated which of these performances could be kept, or improved, when all languages are simultaneously classified. Table 2 shows the classification performances for all methods and using three types of features: qualitative (melodic shape), quantitative (acoustical parameters) and mixed. In this case, we observed that the best performances (nearly 84%) were obtained using mixed and qualitative features either with RF or NFC-SCG classifiers. Moreover, almost all performances were above the chance level (33.33% for the 3 classes).

Table 2. Accuracy percentages obtained for the classifiers using simultaneously all languages. Q - qualitative features; q - quantitative features; M - both qualitative and quantitative features. The low ANFCLH-FS q accuracy is obtained applying feature selection in which one feature from two variables is selected This means that one qualitative feature is not good enough for classifying among languages.

Classifier	All languages		
	Q	q	M
RF	82.85	60	**83.80**
NFC	74.18	72.45	76.40
NFC-SCG	73.27	67.72	81.31
ANFCLH	74	63	79.31
ANFCLH-FS	75.18	16	66.01

IV. Discussion

In this paper, a preliminary assessment of the relationship between the native language of the babies' mothers and qualitative features computed from infant cry recordings is presented. Results point out strong differences of newborn cry melody between Italian, Arabic and French mother tongues, even when all languages are simultaneously classified. Furthermore, the methods' performances are above the chance level for 3 classes, highlighting the performances obtained by RF (using qualitative and mixed features) and NFC-SCG (using mixed features). Thanks to the robust estimation and classification techniques the percentage of classification accuracy was found quite high: 90.52% between Italian and French and even higher (93.75%) between Italian and Arabic. The percentages vary according to the language and the classification method applied: the most robust methods are RF and NFC when the Italian language is compared to the other two, nevertheless the percentages are always above 67%.
We notice that the highest percentage was found between Italian and Arabic: maybe this could be related to the quite low number of guttural sounds that are found in Italian with respect to Arabic and also to French language.

V. CONCLUSION

In this paper, first results are presented concerning the automated classification of the newborn's cry melody. The melodic shapes as well as several acoustical parameters of the newborn cry are estimated with the BioVoice software tool, whose performance was tested with synthetic signals [13]. The classification is performed with several methods on a large data set, made up of more than 7.500 cry units, coming from French, Arabic and Italian mother language newborns, according to a specific protocol. Results show differences up to 94% thus suggesting that newborns pick up elements of their parents' language before they are even born, and certainly before they start to babble themselves.

Although the outcomes are promising, an extensive study should be further carried on for a better understanding. Also, improvements in the SW could help with a better assessment of the 12 melodic shapes as well as for the detection and classification of other shapes. Finally, recording a larger set of infant cries, especially the Arabic ones, would help to evaluate if the methods' performances could be kept.

REFERENCES

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

[2] Mampe B, Friederici AD, Christophe A, Wermke K. Newborns' cry melody is shaped by their native language. Curr Biol. 2009 Dec 15;19(23):1994-7. doi: 10.1016/j.cub.2009.09.064. Epub 2009 Nov 5.

[3] Wermke K., Mende W., Manfredi C., Bruscaglioni P. Developmental aspects of infant's cry melody and formants, Medical Engineering and Physics,24,7-8, 2002

[4] Manfredi C., Bocchi L., Orlandi S., Spaccaterra L., Donzelli G.P., High-resolution cry analysis in preterm newborn infants, Medical Engineering and Physics,31, 5, 2009

[5] Manfredi C., Tocchioni V., Bocchi L., A robust tool for newborn infant cry analysis, Annual International Conference of the IEEE Engineering in Medicine and Biology – Proceedings, 2006

[6] Orlandi S., Manfredi C., Bocchi L., Scattoni M.L., Automatic newborn cry analysis: A Non-invasive tool to help autism early diagnosis, Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, 2012.

[7] Orlandi S., Dejonckere P.H., Schoentgen J., Lebacq J., Rruqja N., Manfredi C. Effective pre-processing of long term noisy audio recordings: An aid to clinical monitoring Biomedical Signal Processing and Control,8, 6, 2013.

[8] Orlandi S., Reyes Garcia C.A., Bandini A., Donzelli G.P., Manfredi C., Application of Pattern Recognition Techniques to the Classification of Full-Term and Preterm Infant Cry, Journal of Voice, 30, 6, 656-663, 2016, https://doi.org/10.1016/j.jvoice.2015.08.007.

[9] Sun CT, Jang JSR. A neuro-fuzzy classifier and its applications. Proc. of IEEE Int. Conf. on Fuzzy Systems, San Francisco 1:94–98. Int. Conf. on Fuzzy Systems, San Francisco 1993.

[10] B. Cetişli, A. Barkana. Speeding up the scaled conjugate gradient algorithm and its application in neuro-fuzzy classifier training. Soft Computing 14(4):365–378, 2010.

[11] B. Cetişli. Development of an adaptive neuro-fuzzy classifier using linguistic hedges: Part 1. Expert Systems with Applications, 37(8), pp. 6093-6101, 2010.

[12] B. Cetişli. The effect of linguistic hedges on feature selection: Part 2. Expert Systems with Applications, 37(8), pp 6102-6108, 2010.

[13] Orlandi S., Bandini A., Fiaschi F.F., Manfredi C., Testing software tools for newborn cry analysis using synthetic signals, Biomedical Signal Processing and Control, 37, 16-22, 2017, https://doi.org/10.1016/j.bspc.2016.12.012