On the Development of a Non Invasive Pathologies Identification System by Qualitative and Quantitative Characterization of Infant Crying and the Application of Intelligent Classification Models to be Used in Rural Environments)

Carlos A. Reyes-Garcia
Ciencias y Tecnologías Biomédicas
Instituto Nacional de Astrofísica, Óptica y Electrónica
Tonantzintla, Pue., México
kargaxxi@inaoep.mx

Alejandro. A. Torres-Garcia
Coordinación de Ciencias Computacionales
Instituto Nacional de Astrofísica, Óptica y Electrónica
Tonantzintla, Pue., México
kcobaimskywalker@gmail.com

M. Antonia Ruiz-Diaz
Universidad Politecnica de Tlaxcala
Zacatelco, Tlaxcala, Mexico
mariaantonia.ruiz@uptlax.edu.mx

Abstract— **The detection of pathologies in the early stages of a baby's life has been one of the major challenges to overcome for the medical sciences. Lack of means of interpretation of this normal physical manifestation of the child has made this task extremely complicated. The discovery that the crying wave, as the sole initial means of communication of babies, contains information about their neurophysiological state, has opened the possibility of interpreting that state and to diagnose diseases from a few days of birth. In this paper we present the efforts to develop a practical integral system to automatically identify pathologies in newborn babies by selecting quantitative features and which also highlights different types of qualitative characteristics on the newborn infant crying through appropriate acoustic processes. And, in each case, after the features are selected or identified uses them to recognize the inherent pathology. Once the system is complete and fully tested we pretend to offer rural nurses, general doctors, researchers and scholars a tool as a mean to make noninvasive diagnostics and as an information support to allow them to have a solid perspective on relevant crying events, and to facilitate the development and unification of standards for the assessment or comprehensive description of the crying wave.**

Keywords*— *analysis of infant crying, automatic identification of qualitative characteristics, classification of crying, non-invasive diagnosis.

I. INTRODUCTION

The crying of newborns is a functional expression of basic biological needs, emotional or psychological conditions such as hunger, cold, pain, cramps and even joy [1]. It requires a coordinated effort of several brain regions, mainly brainstem and limbic system, and is related to respiration and pulmonary mechanisms. Its characteristics reflect the development and possibly the integrity of the central nervous system. Therefore, the analysis of infant crying is a suitable non-invasive complementary tool to assess the physical state of infants, particularly important in the case of premature infants. Specifically, the distinction between a regular crying and one with abnormalities is of clinical interest. Being economic and without contact, the study of the crying of the newborn baby has had an outstanding growth in the last decades. Several studies refer to both the subjective auditory analysis of voice and speech and to automatic acoustic analysis in adults. However, with regard to newborn crying, there are few automatic methods, some based on classical approaches such as the Fourier transform and the autocorrelation analysis [1] [2] [3] [4] [5] and others in parametric techniques [6] [7]. These methods allow us to estimate the main acoustic quantitative characteristics, such as the frequency of vibration of the vocal cords, the resonance frequencies of the vocal tract, linear prediction coding (LPC), Mel frequency cepstral coefficients (MFCC), etc. In recent years, several authors propose classification methods for a wide range of pathologies. Reyes et al. [8], [9], [10] have investigated normal, deaf and asphyxiating newborns through neural networks, evolutionary model selection and fuzzy logic, Poel et al [11] present results on the classification of crying in newborns normal disorder and related to hypoxia using radial-based function neural networks with a general classification performance of 85%.

The cry of the newborn reflects the development and possibly the integrity of the central nervous system, so that its analysis is an attractive non-invasive means to assess the physical state of babies from very early stages of life. In the analysis of infantile crying, it is also important to identify the qualitative characteristics, since they provide relevant extra information that allows to identify variations or similarities between normal and pathological crying, as well as to differentiate between different pathologies. Generally the analysis of the qualitative characteristics is done manually, by means of visual perception (inspecting spectrograms) and auditory (listening to the crying recordings) of specialist doctors, who according to what they see and hear can make a diagnosis . This paper presents the approach to develop a system which will integrate a model section method to use quantitative features and a method that allows the automatic detection of crying units and ehich uses a model called "dodecagrama" that allows to automatically identify the melody type of the crying units, and finally with the values of the fundamental frequency. they automatically identify distinctive qualitative characteristics such as shifts, glides and noise concentrations of crying units.

IREHI 2018 : 2nd IEEE International Rural and Elderly Health Informatics Conference

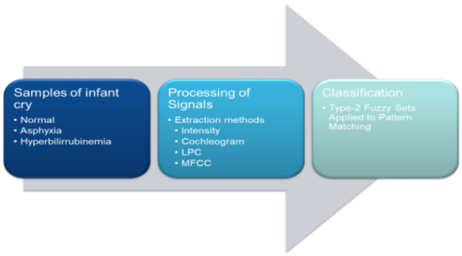

Figure 1. Automatic Infant Cry Recognition (AICR) Process

II. The Automatic Infant Cry Recognition (AICR) Process

This process, in general, is performed through two phases; first, the processing of the signal to obtain the acoustic characteristic vectors and the second phase to identify the type of cry by means of a classifier, Figure 1 shows this process.

2.1 Signal Processing Phase

There are two different processes to extract acoustic features from the infant cry wave. One is to get quantitative features and the other is to obtain the qualitative features.

In the case of the quantitative features, during the signal processing phase, each signal was divided in segments of 1sec. Each segment was subdivided in 50 ms windows, then generating 19 windows out of every one second sample. Later, from each window 16 coefficients MFCC were extracted, with which a total of 304 coefficients by vector were obtained. After adding the label of the class, each vector has 305 attributes. For the experimentes shown in this paper we used 507 samples of normal cry and 879 of deaf cry. At the end the size of the matrices generated for each type of cry were as follows: normal cry 507x305, deaf cry 879x305. The quantitative features extracted for our purposes were Mel Frequency Cepstrum Coefficients (MFCC) which have a frequency similar to the one of the human ear, which is more sensitive to certain frequencies that to others. In this way, an approximation to the form in which the ear perceives the sounds is obtained.

For the qualitative features the extraction process begins with the detection of crying units. This is carried out with the purpose of using them for a later analysis, such is the case of [12], in which the average duration of the crying signals, the mean of the fundamental frequency of the crying as well as its melodic form are analyzed.

In the recordings are sounds that are not useful for the analysis of infant crying, such as the sounds produced by the environment and the inspiratory sounds produced by the infants before emitting a unit of crying, to which the doctors call inspirations. Another important point that should be considered is the variety of environments and devices in which the recordings are acquired as well as the intensity and type of crying of the infant.

In the recordings there can be cries with high pitch, with low pitch, sharp cries, etc. and the variation of the intensity, since the infants can reduce or increase the intensity of their crying at will in the same recording. Crying detection has been carried out manually in various works such as [1], [13].

In this work, the crying unit detection method is part of an interface implemented in MATLAB. Based on experimental tests, crying units smaller than 200ms were eliminated because they are very short-duration sounds that do not provide useful information for further analysis. We also defined an energy threshold (U (e)) applied to the signals, and which, based on [14] and our experiments, is obtained as follows:

$$U(e) = \frac{E_n}{4}$$

where E_n is the energy of the short time signal.

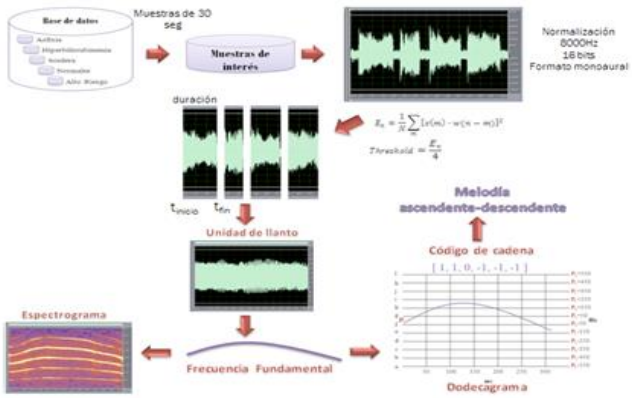

Figure 2. General scheme of the process of extraction and processing of qualitative characteristics

Figure 2 shows step by step the operation of the proposed method starting with the detection of crying units. After the identification of the melody type is made, a modification of the method presented in [15] was developed, which we have called the dodecagram method, which is part of the feature extraction interface. In Figure 2 the dodecagram is shown in which each crying unit is positioned at the center of lines *f* and *g*. The value of the lines is determined by the value of the fundamental frequency of the first window. The next step is to code the unit of crying by means of the following rules: 1 if the value of the fundamental frequency passes to a higher row, 0 if the The value of the fundamental frequency remains in the same row and -1 if the value of the fundamental frequency passes to a lower line. Where the number 1 corresponds to an increase in the fundamental frequency, the 0 without changes, and -1 to a decrease in the fundamental frequency. Therefore, we can see that the unit of crying shown in Fig. 1 has a melodic form of type: descending-ascending.

To identify the *shifts*, the differences of the fundamental frequencies along the signal are measured, if the difference exceeds 100Hz it is considered shift (there can be more than one in a crying unit).

In the same way to identify the *glides* the differences of the fundamental frequencies along the signal are measured, if the difference exceeds 600Hz in a very short time it is considered glide (there can be more than one in a crying unit). A *vibrato* is defined as a series of waves with at least four movements of ascent and descent in the fundamental frequency

III. CLASSIFICATION METHODS:

With the purpose of provide the interface of the system with the best classifiers available several hybrid classification methods were tested, all of them implemented with a combination of fuzzy systems with evolutionary algorithms or neural networks. Some of them were tested for the quantitative features and some for the qualitative ones. They are described next.

3.1. The Quantitative Experiments

For classifying the quantitative features, in this case, we experimented with a method that first reduces the size of features and instances from the data base by means of a selection process based a hybrid fuzzy-genetic algorithm as presented in [16]. The hybrid fuzzy-genetic algorithm for feature and instance subset selection combines a Hybrid Meta-Heuristic (HMH) algorithm and a Fuzzy Self-Adaptive Genetic Algorithm whose crossover operator is a Rotary Circular Crossover (RCC) based on Half Uniform Crossover (HUX). The best individual in the initial population is used as the initial solution of the HMH with the purpose of improving its fitness. This method –also proposed in this work- is a combination of simulated annealing, taboo search and hill-climbers algorithms which allows us to speed up the convergence of the initial population. The genetic algorithm adjusts its own control parameters while running the algorithm by means of two fuzzy inference systems. For a complete description of the functioning of the selection process the reader is refered to the work presented by Leon-Barranco *et al* in [16].

3.2. The Pattern Classification Stage.

We first note that the amount of data is very large, so it will be useful to select variables or features just as to select instances with the intention of reducing computational resources and, to improve the accuracy of the classification process. We have 1386 instances or vectors in the infant cry data base, and each infant cry vector is represented by a vector of 304 features. All these 1386 instances belong to 2 classes (normal and hypo-acousic infant cry). What we want to do is to find a smaller feature and instance subset that can better represent the infant cry dataset avoiding the need of using all data, without degrading the classification accuracy of the system.

The results of evaluating the Hybrid Fuzzy-Genetic Algorithm are presented in Table 1. The baby's crying dataset has 1386 instance with 304 features each. Ten training, evaluation and test sets were formed, i.e., for each experiment 826, 280 and 280 instances for training, evaluation and test respectively were randomly selected from the baby's crying dataset. The presented results are from experiments with a maximum population size Mp = 100. The stop criterion was to reach 8,000 fitness function evaluations.

Table 1. Results of generating solutions with the Hybrid Fuzzy-Genetic Algorithm for the classes of Normal vs Hipoacusic (deaf) infant cry, where (E.A) is Evaluation Accuracy, (T.A) is the Test Accuracy and Storage is the percentage of data kept after the reduction/selection stage.

Method	E.A. (%)	T.A. (%)	Storage (%)
Hybrid Fuzzy-GA	100	97.35	0.08

Table 2. Results of FRNN and PCA for the classes of Normal vs Hipoacusic (deaf) infant cry, where (E.A) is Evaluation Accuracy and (T.A) is the Test Accuracy. PCA refers to the percentage of information we want to preserve

	PCA 70%		PCA 80%	
# MF	E.A. (%)	T.A. (%)	E.A. (%)	T.A. (%)
3 MF	87.74	86.68	85.02	84.85
5 MF	88.58	87.93	86.76	86.81
Average	**88.16**	**87.30**	**85.89**	**85.83**

Additionally, the results of an average of 10 experiments with other method that combines vector reduction with PCA and a Fuzzy Relational Neural Network (FRNN) are presented in Table 2. The FRNN receives as input a training set and returns as output a fuzzy relational matrix. Before presenting the training data to the FRNN, PCA is applied to the dataset and the experiments are performed by using 2, 7, 26 and 65 features respectively from the transformed matrix. The training, evaluation and test sets used in the experiments were the same than those used to obtain the results presented in Table 1. Experiments with 3 and 5 linguistic properties are reported, the input membership values were obtained with the Trapezoidal membership function, with fd = 6 and fe = 1.5. Three epochs were completed to train the Neural Network. The classification was performed with the max-min composition.

This process was carried out once each input feature of the training samples was transformed in membership values to each of the assigned linguistic properties, i.e., once a vector containing m features was transformed in a $3m$, or $5m$-dimensional vector, and the fuzzy relational matrix was returned by the FRNN. If we consider that the 826 instances with the 304 features are the 100% of data, then 0.66%, 8.55%, 16.45% and 49.34% was the storage size used for 2, 7, 26 and 65 principal components, respectively.

3.3 The Qualitative Experiments

For the classification experiments when using the qualitative features four adaptive neuro-fuzzy classifiers were implemented and adapted [17-20], which are; Neuro-fuzzy adaptive with linguistic modifiers classifier (ANFCLH), neuro- fuzzy classifier with linguistic modifiers and selected characteristics (LHNFCSF), neuro-fuzzy conjugate gradient classifier (SCGNFC) and accelerated conjugated gradient neuro- fuzzy classifier (SSCGNFC). The SCGNFC and SSCGNFC systems are optimized by scaled conjugate gradient algorithms. In these two systems, the k-means algorithm is used to initialize fuzzy rules. Also, the Gaussian membe rship function is only used for descriptions of fuzzy sets. The other two systems are based on linguistic modifiers (LH) tuned by scaled conjugate gradient.

3.4 Experimental Results

The system was tested with two sets of samples, one obtained from the Chillanto database of the National Institute of Optical Astrophysics and Electronics of Mexican infants, in which it was tested with a base of Normal cries against that of Hipoacusic (deaf) cells collected at the National Rehabilitation Institute in Mexico. In addition, for the purposes of the international project mentioned in the Aknowledgements, we tested the classifers in another environment to differentiate between premature and term baby crying registered in the University Hospital of Liège, in Belgium to assess the versatility of the system.

Table 3. Classification results with each of the classifiers for the classes of Normal vs Hipoacusic (deaf) infant cry. The results of different iterations will be displayed together with the average accuracy and the standard deviation.

NFC	f NFC_LH	NFC_accelerated	NFC_LH_FS
61.54	61.54	69.23	76.92
100	100	100	100
78.57	78.57	78.57	85.71
69.23	69.23	61.54	46.15
85.71	92.86	85.71	92.86
71.43	78.57	71.43	78.57
61.54	61.54	61.54	76.92
76.92	76.92	76.92	92.31
76.92	76.92	76.92	76.92
69.23	61.54	69.23	46.15
Average Accuracy			
75.1099	75.7692	75.1099	77.2527
Standard Deviation			
8.7061	8.6885	8.8066	11.334

Table 4. Classification results with each of the classifiers for the Premature vs. Term classes. The results of different iterations will be displayed together with the average accuracy and the standard deviation

NFC	f NFC_LH	NFC_accelerated	NFC_LH_FS
100	100	66.67	100
100	100	100	83.33
100	100	100	85.71
83.33	83.33	83.33	83.33
83.33	83.33	83.33	50
85.71	42.86	71.43	71.43
100	83.33	100	66.67
85.71	71.43	100	85.71
Average Accuracy			
84.52	81.90	84.52	72.14
Standard Deviation			
17.96	17.43	15.64	20.65

IV. DISCUSION

As can be observed, the classification results can be more accurate when using the quantitative features. Besides the high efficiency shown by the hybrid fuzzy-genetic algorithm since, after the process of instance/features selection/reduction, with only a small fraction of the data base it was able to obtain a test accuracy of 97.35%. This result was higher than the highest obtained with the FRNN and PCA method when keeping the 70% of information whose Test Accuracy was of 87.3 % for the same set of samples.

Regarding the accuracy in the recognition when using qualitative features, it can be seen that in the case of Normal vs Hipoacusic (deaf) infant cry the NFC_LH_FS recognizer was the one that obtained better results with 77.25% and in the case of Premature vs. Term two classifiers, NFC and NFC_accelerated, obtained the highest precision that is 84.52%.. Although the results seem not to be very high in any way, they are very encouraging, since they were obtained using only the qualitative characteristics.

All four classifiers are trained through a series of iterations following a 10-fold Cross Validation scheme. Tables 3 and 4 show the partial results that each classifier obtains at the end of each iteration (fold).

Although better results have been obtained using quantitative characteristics, the presented study is relevant since the extraction of qualitative characteristics is important because besides allowing a description very close to that used by the medical specialists when doing a perceptive study of the waves of baby crying acceptable classification results can be obtained through which an initial diagnosis can be made and the right advise of first care be provided.

V. CONCLUSION

By the results obtained we are supported to

Perceptually, there is a great similarity between the crying waves of healthy and deaf babies, so, obtaining a recognition accuracy of 97.35% when using the quantitative features encourages us to continue with our work to finish the non invasive diagnostic tool.

The accurate results obtained with our developed models for the processing of quantitative features reinforced with the descriptive properties of the qualitative characteristics reflects a high potential for its later application in the development of non-invasive diagnostic systems.

In the same way, in the case of the classification of premature and term cries in which 84.52% was obtained, it allows to see a greater possibility of application of the qualitative characteristics in domains other than the diagnosis. In this case the greater precision can be explained by the difference between the two types of crying caused by the improved maturity of the phonatory apparatus of babies born at term.

The system and the models still require more studies and experiments with various modifications, such as the use of a combination of both qualitative and quantitative characteristics tanking the advantages shown by each one, including other methods that we have tested for the same purpose.

With the results obtained it is possible to ensure that the use of both quantitative along qualitative characteristics for their classification for various purposes opens a large window of application opportunities limited only by the imagination of the developers. The first proposed task in progress is to finish a robust, easy to use and friendly interface that can be handled and the results interpreted by any non-specialized health personal. An application is thought to be installed in smart phones with the possibility to be integrated to telemedicine systems and which could be particularly useful in rural environments where the where the specialists do not arrive and the diagnostic devises are extremely limited.

ACKNOWLEDGMENT

The authors are grateful to the project MX14MO06 "Techniques of analysis and classification of voice and facial expressions: application to neurological diseases in newborns and adults" of the Mexico-Italy scientific and technological cooperation executive program financed by AMEXID of the SRE and the Ministry of Health. Foreign Affairs of Italy.

REFERENCES

[1]. Michelsson K, Michelsson O. Phonation in the newborn, infant cry. Int J Pediatr Otorhinolaryngol 1999;49 (Suppl 1):S297-S30.

[2]. Fuller BF. Acoustic discrimination of three types of infant cries. Nurs Res 1991; 40(3):156-160.

[3]. Laufer MZ, Horii Y. Fundamental frequency characteristics of infant non-distress vocalization during the first twenty-four weeks. J Child Lang 1977; 4(02):171-184.

[4]. Reggiannini B, Sheinkopf SJ, Silverman HF, Li X, Lester BM. A Flexible Analysis Tool for the Quantitative Acoustic Assessment of Infant Cry. J Speech Lang Hear Res 2013; 56(5):1416-1428.

[5]. Sirviö P, Michelsson K. Sound-spectrographic cry analysis of normal and abnormal newborn infants. Folia Phoniatrica et Logopaedica 1976, 28(3):161-173.Hevia-Montiel N, Molino-Minero-Re E, Carrillo-Bermejo AJ. Tortuosidad discreta como medida morfométrica en tumores cerebrales. Rev Mex Ing Biom. 38(1):188–98. 2017 DOI: 10.17488/RMIB.38.1.13

[6]. A. Fort, C. Manfredi, "Acoustic analysis of newborn infant cry signals", Med. Eng. Phys, vol. 20, pp. 432-442, 1998.

[7]. Kheddache Y, Tadj C. Resonance Frequencies Behavior in Pathologic Cries of Newborns. Journal of Voice 2014. http://dx.doi.org/10.1016/j.jvoice.2014.04.007

[8]. Reyes-Galaviz, O. F., Tirado, E. A., & Reyes-Garcia, C. A. (2004).Classification of infant crying to identify pathologies in recently born babies with ANFIS (pp. 408-415). in Lecture Notes in Computer Science 3118, Springer Berlin Heidelberg. Computers Helping People with Special Needs, edited by Klaus Miesenberger et. al., Springer, Berlin, 2004, pp 408-415.

[9]. Cano, O.S.D., Escobedo, D.I., Suaste, I., Ekkel, T., Reyes Garcia, C.A.: A Combined Classifier of Cry Units with New Acoustic Attributes. In: Martínez-Trinidad, J.F.,Carrasco Ochoa, J.A., Kittler, J. (eds.) CIARP 2006. LNCS, vol. 4225, pp. 416–425. Springer, Heidelberg (2006).

[10]. Amaro-Camargo, E., & Reyes-García, C. A. (2007). Applying statistical vectors of acoustic characteristics for the automatic classification of infant cry. InAdvanced Intelligent Computing Theories and Applications. With Aspects of Theoretical and Methodological Issues (pp. 1078-1085). Springer Berlin Heidelberg.

[11]. Poel M., Ekkel T. Analyzing infant cries using a committee of neural networks in order to detect hypoxia related disorder 2006 International Journal on Artificial Intelligence Tools 15 3 397-410.

[12]. K. Michelsson, K Eklund, P. Leepänen, H. Lyytinen, "Cry characteristics of 172 Healthy 1- to 7-Day-Old Infants", International Journal of Phoniatrics Speech Therapy and Communication Pathology, vol. 54, 2002

[13]. K. Wermke, W. Mende, C. Manfredi, and P. Bruscaglioni, "Developmental aspects of infant's cry melody and formants", Med Eng Phys, vol. 24(7-8), pp. 501-514, 2002.

[14]. I Saratxaga, I Luengo, E Navas, I Hernández, J Sánchez, I Sainz, "Detección de PITCH en condiciones adversas", IV Jornada en Tecnología del Habla, Universidad del País Vasco – Euskal Herriko Unibertsitatea pp. 13-18,2006.

[15]. Benyó Z. Várallyay G Jr., Illényi A. Melody analysis in the newborn infant cries. pages 11-14, Diciembre 2009.

[16]. Agustin Leon-Barranco, Carlos A. Reyes-Garcia, and Ramon Zatarain-Cabada, "A Hybrid Fuzzy-Genetic Algorithm", in Lecture Notes in Computer Sciences (LNCS) 4113, edited by D.-S. Huang, K. Li, and G.W. Irwin, Springer, Berlin, 2006, pp. 500-510, ISBN: 3-540-37271-7, ISSN: 0302-9743.

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

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

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

[20]. Schönweiler R, Kaese S, Möller S, Rinscheid A, Ptok M. Neuronal networks and self-organizing maps: new computer techniques in the acoustic evaluation of the infant cry, Int. J. Pediatr. Otorhinolaryngol. 38 (1) (1996) 1–11.