Abstract: The paper assumes the implementation of a cry-based classifier for neonatal diagnosis. The main contribution is concerned with the articulated processing of cry signals, which includes two kinds of approaches: a threshold-based classification and ANN-based classification. Every one of those approaches makes its own contributions to the cry classification, both are adequately combined in a classifier of two-class (pathological and normal). Moreover the use of cry unit as a primary data was also an interesting aspect held by the authors. This articulated cry processing is the main body of a new cry-based methodology for neonatal diagnosis, which will be presented in a few months by the Group of Speech Processing in Cuba.

Keywords: cry analysis, neural networks

I. INTRODUCTION

Since the use of new approaches like ANN’s have been applied for cry classification the possibility to make a cry-based diagnosis in newborns has become in reality [1-3] [17]. In this paper the state-of-art in cry analysis and new focus of soft computing have been properly combined, leading up to a suitable articulated processing of the cry signals oriented for a neonatal diagnosis. As it is explained in the main body of paper five specific forms of processing are articulated in one: (1) a digital signal processing (acoustic cry parameters extraction and Mel frequency cepstral coefficient (MFCCs) estimation), (2) data management (BDLlanto: a Cuban corpus of cry signals), (3) principal component analysis (PCA), (4) neural network –based classification and (5) a threshold-based decision.

II. METHODS

The basis of the research work was based on the physioacoustic model for cry production and the Golub’s muscle control model. As it was mentioned above two classification approaches are properly articulated 2-in-1:

(1) Threshold-based classifier: the threshold values of four cry features for normality are considered [5-9] [16]:

Voicedness: the ratio of the amount of periodic sound versus the amount of noise. (the higher the voicedness, the weaker the noise component in comparison to the periodic sound).

Melody: the performance of fundamental frequency over time, within one cry unit.

Stridor: a rapid increase in air pressure causes the vocal cords to enter a turbulent state resulting in a sudden loss of pitch.

Shift: a sudden large change in pitch

The procedures for the computation of those attributes are the same suggested by Cano et al [17] in 2006.

Cry Corpus.

The cry samples were taken from a Cuban cry corpus named BDLlanto database (32 cases: 16 healthy children and 16 pathological children). The database includes a friendly user interface, which let the user manage acoustical and clinical information of newborns in an efficient manner. It also incorporates some features of Web technologies for Internet facilities.

(2) ANN-based classifier: it consists on a feed-forward network using the method of scale gradient conjugate (MSGC) as learning algorithm. The input vector is composed by the Mel frequency cepstral coefficients (MFCCs) [4] [11-13]

Mel Frequency Cepstral Coefficients.

The low order cepstral coefficients are sensitive as overall spectral slope and the high-order cepstral coefficients are susceptible to noise. This property of the speech spectrum is captured by the Mel frequency cepstral coefficients (MFCCs) [4] [11-13]

Many speech recognition systems are based on the MFCC approach and its first and second order derivative. The derivative normally approximate through an adjustment in the line of linear regression towards an adjustable size segment of consecutive information frames. The resolution of time and the smoothness of the estimated derivative depend on the size of the segment.
The computation of MFCCs follows the steps:
- Converting the signal in small segments
- Computing the Discrete Fourier Transform
- The spectrum converts into a logarithmic scale
- The scale is transformed into a soft MEL spectrum
- The discrete cosine transform (DCT) is computed

The above mentioned algorithm is illustrated in Fig. 2.

The Artificial Neural Network (ANN)

The use of ANN has been a great impact in the development of several research areas like computer vision, autonomous vehicle, pattern recognition, connected-speech synthesis and more recently into the classification of cry units [1-2] [4]. In the paper the ANN structure used is shown in Fig. 3. It corresponds to a Feed-Forward network in which x1, x2, x3, ..., xn represent the acoustic features of signals and y1, t2, y3, ... ym the m classes to be identified. This kind of supervised ANN has been also used in cry classification with succeed [11].

In order to increase the efficiency of the learning process the Method of Scaled Conjugate Gradient (MSCG) is chosen. [13]. The MSCG algorithm shows a linear convergence accentuated in most of the problems.

III. RESULTS

Starting from the primary information in BDLlanto, a segmentation process was developed to generate the cry units being obtained 73 healthy cry units and 68 pathological cry units (relative to hypoxia). 58 cry units were chosen (for class) for training and 10 for classification. The segmentation stage was semi-automatic combining a begin/end detection (based on function energy and zero-crossing rate) and a manual correction to reduce the negative effect of considering inappropriate sections within the cry unit. In the Figure 4 the scheme of the combined classifier is presented.

Starting from the cry units obtained from database a parameter estimation for every cry unit is done, following two possible ways:

(a) estimation of 4-acoustic features for the threshold-based classifier: the estimated feature is then compared with the normal threshold values associated to each one of the 4 selected parameters, generating to the exit an index FN1 with the following gradation:

\[\begin{align*}
&FN1: 0.25 \text{ for 1 parameter altered} \\
&0.5 \text{ for 2 parameters altered} \\
&0.75 \text{ for 3 parameters altered} \\
&1.0 \text{ for 4 parameters altered} \\
&0 \text{ for no one parameter altered (normality index)}
\end{align*}\]

(b) estimation of MFCC’s for the ANN-based classifier.

500 MFCC’s were computed for each generated cry unit (because of the differences in time duration among the cry units it was necessary to normalize and to adjust the vector of coefficients).

After the initial vector of characteristics was passed through the analysis of principal components (PCA) the dimension of the vector was definitively reduced to 50

Fig. 1 The Mel Filter Bank

Fig. 2 The MFCC’s computation from a cry signal

Fig. 3 A Feed-Forward architecture

Fig. 4. Block diagram of the combined cry-based classifier
principal components. Then the input vector to the ANN was presented, with the following structure: 50 nodes in the input layer, 15 nodes for the hidden layer and finally 2 nodes for the output layer. To detect the cry type in the newborn the output values of the net are analyzed. The output values of the net are coded between 0 and 1. If the value of the output node 1 is bigger than the value of the output node 2 the sample is assigned to the class "normal" (N) generating a FN1 index equal to 0, otherwise it is assigned to the class "pathologic" (P) generating a FN2 index equal to 1.

Finally both FN1 and FN2 indexes are processed in a classes-based decision with 3 qualitative levels:
- Normal: $D \leq 0.5$
- Moderately-pathologic: $0.5 < D \leq 0.75$
- Pathologic: $D = 1.0$

IV. DISCUSSION

The following table shows results from the Combined Classifier.

<table>
<thead>
<tr>
<th>Class</th>
<th>Confusion Matrix</th>
<th>D index</th>
<th>% of Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>10</td>
<td>&lt;=0.5</td>
<td>100</td>
</tr>
<tr>
<td>Pathology</td>
<td>2</td>
<td>0.5</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>20</td>
<td>0,75</td>
<td>90</td>
</tr>
</tbody>
</table>

The gradation in the D index lets physicians to use properly the output of the cry classifier in order to compare and to evaluate its "possible meaning" in front of the results from the neurophysiological evaluation of the newborn (how much abnormal the infant cry is from the acoustical point of view and its "weight" for diagnostic purpose). The need to include more acoustic features in cry classifier for better classification rates proposed and argued by Schonweiller in 1996, is well demonstrated here. An interesting aspect that deserve to be commented is the fact that the only two cry units misclassified as normal obtained a FN1 equal to 0.75 (significative abnormal for the threshold-based classifier), so both outputs from the classifiers also offer valuable information to be considered by the specialists.

The soft tools used in the experience were: BDLlanto database with 12 seconds-cry recordings of Cuban children, BPVOZ soft-package, PCVOX and praat software for the acoustic signal processing. The ANN implementation (including the MSGC algorithm) was done with Neural Network Toolbox from Matlab v. 6.0.

V. CONCLUSION

The articulated processing of cry signals was well implemented in order to improve the effectiveness of a N/P cry classification, obtaining satisfactory results. Both output indexes FN1 and FN2 offer also valuable information for specialists when they analyze them together or in separate environment. The cry unit as a basic element for signal processing displayed also a positive performance during the research experience. The use of this articulated-signal processing will be the keystone for a new cry-based methodology for newborn diagnosis with CNS disorders (based on hypoxia) to be issued by the Group of Speech Processing.

REFERENCES


