ESANN'2003 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 23-25 April 2003, d-side publi., ISBN 2-930307-03-X, pp. 349-354

Detecting Pathologies from Infant Cry Applying Scaled Conjugate Gradient Neural Networks

José Orozco, Carlos A. Reyes García

Instituto Nacional de Astrofisica Optica y Electronica (INAOE)
Luis Enrique Erro # 1,
Tonantzintla, Puebla, México,
jose@ccc.inaoep.mx, kargaxxi@inaoep.mx

Abstract. This work presents the development of an automatic recognition system of infant cry, with the objective to classify two types of cry: normal and pathological cry from deaf babies. In this study, we used acoustic characteristics obtained by the Linear Prediction technique and as a classifier a neural network that was trained with the scaled conjugate gradient algorithm. Preliminary results are shown, which, up to the moment, are very encouraging.

1. Introduction

The infant's crying is a communication way, although more limited, it is similar to adult's speech. Parents and specialists in the area of child care learn to distinguish among the different kinds of baby cries, making use of individual perception for the auditive differentiation and interpretation of the several ways an infant cries. Both, differentiation and interpretation are totally subjective, and their only support comes from training and experience of each person. According to the specialists, babies' crying wave carries useful information, to determine the physical state and/or state of mind of the baby, the same as to detect possible physical pathologies, mainly cerebral, from very early stages. In previous works on the acoustical analysis of baby's crying, it has been shown that there exists significant differences among the various types of crying, like healthy infant cry, pain cry and pathological infant cry, using classification methodologies based on Self-Organizing Maps [1], neural networks [2] and spectral analysis [3]. Still, there is not a concrete and effective research technique on baby crying with clinical and diagnosis's purposes.

2. Automatic Infant Cry Recognition Process

The Automatic Infant Cry Recognition process (fig. 1) is basically a problem of pattern processing,. The goal is to take the crying wave as the input pattern, and finally obtain the type of cry or pathology detected in the baby. Generally, Automatic Cry Recognition is done in two steps. The first step is known as signal processing, while the second is known as pattern classification.

In the acoustical analysis, the crying signal is analyzed to extract the more important features in the time domain. The crying wave is filtered to eliminate irrelevant or undesirable information like noise, channel distortion, and other particular signal's characteristics. Although data are reduced when removing repetitive components, the relevant information for patterns classification is preserved in an optimal way. Some

of the more usual simple techniques for signal processing are: linear prediction coding, cepstral coefficients, pitch, intensity, among others.
The set of obtained characteristics can be represented by a vector, and each vector may represent a pattern. This pattern is then compared with the knowledge the computer has on pattern classification.
Traditionally, four popular pattern recognition approaches have been used: pattern comparison, stochastic models, knowledge based systems and connectionist models.

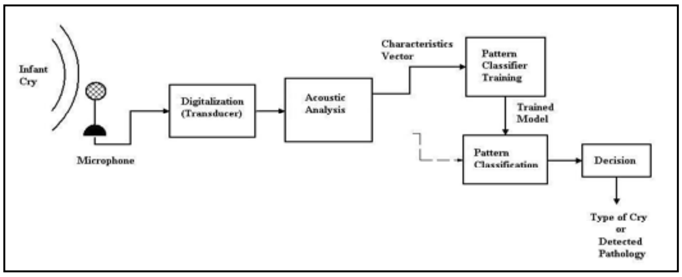

Figure 1 - Automatic Infant Cry Recognition Process

3. Linear Prediction Coefficients

The objective of the application of these techniques is to describe the signal in terms of its fundamental components. Linear Prediction (LP) analysis has been one of the time domain analysis techniques more used during the last years. LP analysis attempts to predict "as well as possible" a speech sample through a linear combination of several previous signal samples. Thus, the spectral envelope can be efficiently represented by a small number of parameters, in this cases LP coefficients.
The particular way in which data are segmented determines whether the covariance method, the autocorrelation method, or any of the so-called lattice methods of LP analysis is used. The first method that we are using is the autocorrelation LP technique, where the autocorrelation function, for the finite-length signal s(n) is defined [4]:

$$\mathrm{r}_j = \sum_{n=0}^{N-1-l} s(n)s(n+l) \qquad (l \geq 0) \qquad (1)$$

where, l is the delay between the signal and his delayed version, N is the total number of examples. The prediction error function between the data segment s(n) and the prediction $\hat{s}$(n) is defined as [4]:

$$e(n) \quad = \quad \sum_{i=0}^{M} a_i s(n-i) \qquad (a_0 = 1) \qquad for \quad n \quad = \quad 0,1,\ldots,N+M-1 \qquad (2)$$

As the order of the LP model increases, more details of the power spectrum of the signal can be approximated.

4. Neural Networks

Neural Networks are among the more used methodologies for classification and patterns recognition. Among the more utilized neural network models, there are the feed-forward networks which use some version of the back-propagation training method. In general, a neural network is a set of nodes and a set of links. The nodes correspond to neurons and the links represent the connections and the data flow between neurons. Connections are quantified by weights, which are dynamically adjusted during training. During training, a set of training instances is given. Each

ESANN'2003 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 23-25 April 2003, d-side publi., ISBN 2-930307-03-X, pp. 349-354

training instance is typically described by a feature vector (called an input vector). It should be associated with a desired output, which is encoded as another vector, called the desired output vector.
The back-propagation training method uses the following technique [5]: given an input pattern to the network, its output is compared with the desired output, and a distance or error between them is calculated. Next, all relevant weights are adjusted in such a way that next time the same instance is processed, the real output is closer to the desired one, which means an error decrease. This process continues until a minimum error is reached or until a given number of training epochs is completed. To evaluate whether the system can perform at an acceptable level, in terms of accuracy and efficiency, we used the 10-fold cross validation technique [6].

5. Scaled Conjugate Gradient Method

From an optimization point of view learning in a neural network is equivalent to minimizing a global error function, which is a multivariate function that depends on the weights in the network. Many of the training algorithms are based on the gradient descent algorithm.
Minimization is a local iterative process in which an approximation to the function, in a neighborhood of the current point in the weight space, is minimized. Most of the optimization methods used to minimize functions are based on the same strategy. The Scaled Conjugate Gradient (SCG) algorithm [7] denotes the quadratic approximation to the error E in a neighborhood of a point w by:

$$E_{\mathrm{qw}}(y) = E(w) + E'(w)^T y + \frac{1}{2} y^T E''(w) y \qquad (3)$$

In order to determine the minimum to Eqw(y)the critical points for Eqw(y) must be found. The critical points are the solution to the linear system defined by Moller in [9]

$$E'_{\mathrm{qw}}(y) = E''(w)\, y + E'(w) = 0 \qquad (4)$$

SCG belongs to the class of Conjugate Gradient Methods, which show superlinear convergence on most problems. By using a step size scaling mechanism SCG avoids a time consuming line-search per learning iteration, which makes the algorithm faster than other second order algorithms. And also we got better results than with other training methods and neural networks tested, as standard back-propagation and cascade neural network.

6. Data Set

The infant cry corpus has been collected from 31 babies. A set of 78 samples have been directly recorded from 31 babies by pediatricians, with digital ICD-67 Sony digital recorders, and then sampled at 8000 Hertz. The babies selected for recording are from just born up to 6 month old, regardless of gender. For our experiments, we took two kinds of crying: normal and pathological crying. Pathological cry is a cry from a deaf baby and a cry from a healthy baby is considered normal cry. 78 crying records, from both categories, were segmented in signals of one second length. 1036 segmented samples were obtained, 157 of them belong to normal cry, and 789 to pathological cry.
Infant cry shows significant differences between the several kinds of crying, which can be perceptually distinguished by a trained person. The general acoustical features

ESANN'2003 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 23-25 April 2003, d-side publi., ISBN 2-930307-03-X, pp. 349-354

for normal crying show, raising-falling pitch pattern, ascending-descending melody, high intensity as shown in figure 2. Pathological crying (fig. 3) shows different acoustical characteristics like: intensity lower than normal, rapid pitch shifts, generally glottal plosives, weak phonations and silence.

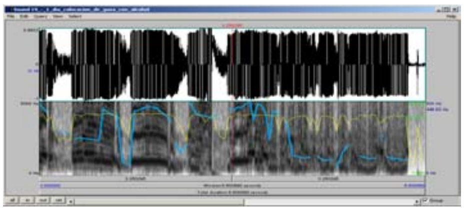

Figure 2 - Waveform and spectrogram of normal crying

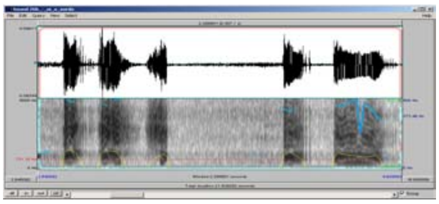

Figure 3 - Waveform and spectrogram of pathological crying

7. Parametric Performance

In the baby cry we do not have a basic unit for analysis, as phonemes in speech. That fact allows us to focus in the analysis and feature extraction of long period patterns.
For the LP analysis, each one second sample was segmented in windows of 50 ms for different experiments. We extracted 16 LP coefficients per window, consequently, the feature vectors contain 320 parameters, corresponding to 320 data inputs to the neural network. In this situations, the dimension of the input vector is large, but the components of the vectors are highly correlated. It is useful in this situation to reduce the dimension of the input vectors. An effective procedure for performing this operation is principal component analysis (PCA). This technique has three effects: it orthogonalizes the components of the input vectors; it orders the resulting orthogonal components (Principal Components or PC) so that those with the largest variation come first; and it eliminates those components that contribute the least to the variation in the data set. After several test, we chose 30 parameters by vector.

8. System Implementation

For the acoustic processing of the cry waves, we used Praat 4.0.2 [8] to obtain the LP coefficients. To perform pattern recognition, a 30 input nodes-- 15 nodes in the hidden layer- 2 output nodes, feed-forward network was developed for training and testing LPC samples. The number of nodes in the hidden layer were heuristically established. The implementation of the neural network and the training method were done with the Neural Networks ToolBox of Matlab 6.0.0.88 [9]. The same Matlab version was used to implement the PCA algorithm.

9. Experimental Results

The neural networks were trained to classify the cries into normal and pathological classes. For training in the first case, with 10-fold cross validation technique, the sample set is divided into 10 subsets (4 groups with 32 samples and 6 groups with 31 samples), each time leaving one set for testing and the remaining for training. This process is repeated until all sets have been used once for testing. In this way, we applied a total of 1413 samples of each class for training and 157 samples for testing.

The classification accuracy was calculated by taking the number of correctly classified samples by the network, and divided by the total number of samples into the test data set. The classification results from each network for every cry category, are shown in Figure 4.

Kind of crying	Samples	Confusion Matrix		Accuracy
		Normal	Pathological	
Normal	157	**149**	8	
Pathological	157	20	**137**	
Total	314			**91.08%**

Figure 4 - Infant cry classification using a SCG neural network with 15 hidden nodes, and 30 Principal Components out of 320 LPC coefficients from 314 samples.

In a second experiment, for testing purposes, we used the total number of samples. In this case, as in the previous experiment, the training was done through the same total number of samples of each class, which is 1413. Then, given that we have more pathological samples available, we used 157 and 879, from normal and pathological crying respectively, for testing as shown in Figure 5. The value under last column represents the accuracy in the classification, which is 86.20 %.

Kind of crying	Samples	Confusion Matrix		Accuracy
		Normal	Pathological	
Normal	157	**142**	15	
Pathological	879	128	**751**	
Total	1036			**86.20%**

Figure 5 - Infant cry classification using a SCG neural network with 15 hidden nodes, and 30 Principal Components out of 320 LPC coefficients from 1036 samples.

Recent work on cry analysis, by Ekkel [10], attempted to expand the set of useful sound features, and tried to find a robust way of classifying these characteristics based on a Radial Basis Function network. Ekkel tried to distinguish between healthy infants (‘normal cases’) and infants diagnosed by a medical expert as ‘abnormal cases’ with 38 records. In his investigation he used two sets of patterns. His results showed that the most successful feature set yields a correct classification rate of 85%(±5), using a set that contains information from All Cry Units (ACU), except the first aspiration, resulting in 183 patterns. We make a comparison considering that the infant cry number and the type of cry are different.

Actual class	Predicted class		Accuracy
	Normal	Hypoxia	
Normal	**79%**	21%	
Hypoxia	12%	**88%**	
Total			85%

Figure 6 - Confusion matrix for multiclassifier system on ACU data set by Ekkel

We want to remark that show Ekkel’s results just to illustrate similar work, even when they were obtained under different conditions.

ESANN'2003 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 23-25 April 2003, d-side publi., ISBN 2-930307-03-X, pp. 349-354

10. Conclusions and Future Work

The SCG Method avoids a time consuming line-search per learning iteration, which makes the algorithm faster than other second order Conjugate Gradient algorithms. This work has shown good results, with the LP technique, using a neural network architecture. We still intent to experiment with other techniques, as well as with the length of frames to further improve our results. During the progress of the project, besides consistently improving the results, we have gathered useful acoustical information on the infant cry. We think that in one moment, this information will be very helpful to pediatricians, and doctors in general. Our main problem has been the collection of samples, not only in number, but with a good labeling, and with an even amount among classes. As future work we consider to collect enough samples to train the classifiers appropriately. Also, we plan to experiment with other acoustical features like, cepstrum and Mel coefficients, pitch, intensity, etc, and different combinations of them, as well as other neural network models, combination of them. Finally, after tuning up our recognizers we will attempt to develop a real time system that recognize the infant cry type and emit a synthesized voice message.

Acknowledgment

This work is part of a project that is being financed by CONACYT (37914-A number). We like to thank Dr. Edgar M. Garcia-Tamayo and Dr. Emilio Arch-Tirado for their invaluable collaboration in helping us to collect the cry samples.

References

[1] Cano, S., Escobedo, D., Coello, E. *El Uso de los Mapas Auto-Organizados de Kohonen en la Clasificación de Unidades de Llanto Infantil*, 1er Taller AIRENE, Universidad Catolica del Norte, Chile, 1999, pp 24-29.
[2] Petroni, M., Malowany, A., Johnston, C., Stevens, B. *Identification of Pain from Infant Cry Vocalizations Using Artificial Neural Networks* (ANNs), The International Society for Optical Engineering. Volume 2492. 1995. pp.729-738.
[3] Wasz-Hockert, O., Lind, J., Vuorenkoski, V., Partanen, E. *El Llanto en el Lactante y su Significación Diagnóstica*, Científico-Médica, Barcelona, 1970.
[4] John D., Gray, Augustine H., *Linear Prediction of Speech*. New York: Springer-Verlag, 1976.
[5] LiMin Fu, *Neural Networks in Computer Intelligence*, Ed-McGraw-Hill. 1994.
[6] Haykin, Simon S. *Neural Networks*: *A Comprehensive Foundation*. New York: Macmillan College Publishing Company, Inc., 1994.
[7] Moller, A *Scaled Conjugate Gradient Algorithm for Fast Supervised Learning*, Neural Networks, 6 (4), 1993, pp.525-533.
[8] Boersma, P., Weenink, D. Praat v. 4.0.8. *A System for Doing Phonetics by Computer*. Institute of Phonetic Sciences of the University of Amsterdam. Netherlands. February, 2002.
[9] MathWoks, Inc., Matlab. *Neural Networks Toolbox*, V.6.0.8
[10] Ekkel, T. "*Neural Network-Based Classification of Cries from Infants Suffering from Hypoxia-Related CNS Damage*", Tesis de Maestría. University of Twente, 2002. The Netherlands.