A Study on the Recognition of Patterns of Infant Cry for the Identification of Deafness in Just Born Babies with Neural Networks

José Orozco-García and Carlos A. Reyes-García

Instituto Nacional de Astrofísica Óptica y Electrónica (INAOE)
Luis Enrique Erro # 1,
Tonantzintla, Puebla, México,
jose@ccc.inaoep.mx, kargaxxi@inaoep.mx

Abstract. In this paper we present the methodologies and experiments followed for the implementation of a system used for the automatic recognition and classification of patterns of infant cry. We show the different stages through which the system is trained to identify normal and hypo acoustic (deaf) cry. The cry patterns are represented by acoustic features obtained by the Mel-Frequency Cepstrum and Lineal Prediction Coding techniques. For the classification we used a feed-forward neural network. Results from the different methodologies and experiments are shown, as well as the best results obtained up to the moment, which are up to 96.9% of accuracy.

1 Introduction

The infant crying is a communication way, although more limited, it is similar to adult's speech. Through crying, the baby shows his or her physical and psychological state. Based on human and animal studies, it is known that the cry is related to the neuropsychological status of the infant [1]. According to the specialists, the crying wave carries useful information, as to determine the physical and psychological state of the baby, as well as to detect possible physical pathologies, from very early stages. In previous works on the acoustical analysis of baby crying, it has been shown that there exist significant differences among the several types of crying, like healthy, pain and pathological infant cry. Using classification methodologies based on Self-Organizing Maps, Cano [2] attempted to classify cry units from normal and pathological infants. In another study, Petroni used Neural Networks [3] to differentiate between pain and no-pain crying. Previously, in the seminal work done by Wasz-Hockert spectral analysis was used to identify several types of crying [4]. In a recent investigation, Taco Ekkel [5] attempted to expand a set of useful sound characteristics, and find a robust way of classifying these features. The goal of Ekkel was to classify neonate crying sound into categories called normal or abnormal (hypoxia). However, up to this moment, there is not a concrete and effective automatic technique, on baby crying, useful for clinical and diagnosis purposes.

A. Sanfeliu and J. Ruiz-Shulcloper (Eds.): CIARP 2003, LNCS 2905, pp. 342–349, 2003.

2 Infant Cry

Crying is the only communication mean that the baby has in the first months of life, before the use of signs or words. The Crying wave is generated in the Central Nervous System, that's why the cry is thought to reflect the neuropsychological integrity of the infant, and may be useful in the early detection of the infants at risk for adverse developmental outcome. In this work, two kinds of crying are considered: normal and pathological (hypo acoustical) crying. 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. First, we have to take a sample set, apply acoustical analysis and principal component analysis to get a reduced vector, which is used to train the recognizer. Second, we take a test sample set, and also apply acoustical analysis and principal component analysis to reduce the vector's dimension. Then the reduced unknown vector is passed by the pattern classifier, which, at the end, classifies the crying sample.

In the acoustical analysis, the crying signal is analyzed to extract the more important features in time domain. Some of the more usual simple techniques for signal processing are: Linear Prediction Coding, Cepstral Coefficients, Pitch, Intensity, among others. The extracted features from each sample are kept in a vector, and each vector represents a pattern.

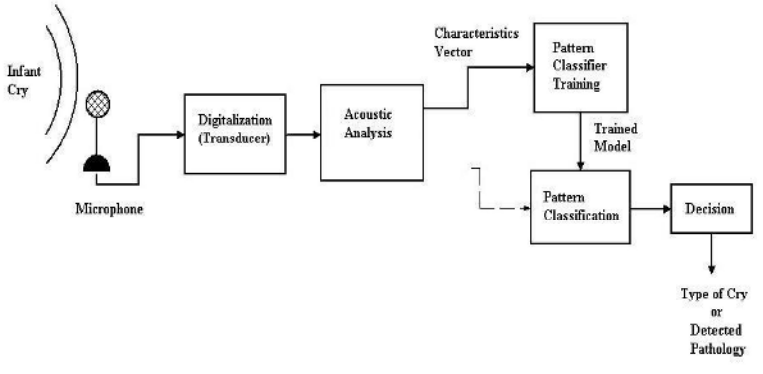

Fig. 1. Automatic Infant Cry Recognition Process

Infant cry shows significant differences between the several kinds of crying, which can be perceptually distinguished by a trained person. The general acoustical features for normal crying show, raising-falling pitch pattern, ascending-descending melody, high intensity as shown in Fig. 2. Pathological crying (Fig. 3) shows acoustical characteristics like: intensity lower than normal, rapid pitch shifts, generally glottal plosives, weak phonations and silences during the crying.

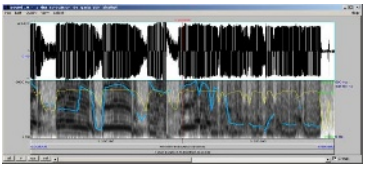

Fig. 2. Waveform and spectrogram of normal crying

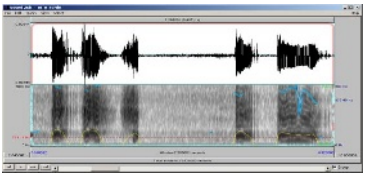

Fig. 3. Waveform and spectrogram of pathological crying

3 Mel Frequency Cepstral Coefficients

The digitized sound signal contains irrelevant information and requires large amounts of storage space. To simplify the subsequent processing of the signal, useful features must be extracted and the data compressed. The power spectrum of the speech signal is the most often used method of encoding. Mel Frequency Cepstral coefficients (MFCCs) [6] are used to encode the speech signal. Cepstral analysis calculates the inverse Fourier transform of the logarithm of the power spectrum of the speech signal. For each utterance, the Cepstral coefficients are calculated for all samples with successive frames. The energy values in 20 overlapping Mel spaced frequency bands are calculated. This results in each frame being represented by 16 and 21 MFCCs.

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

5 Neural Networks

Neural Networks are one of 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 among neurons. Connections are quantified by weights, which are dynamically adjusted during train-

ing. The required training can be done through the back-propagation technique. During training (or learning), a set of training instances is given. Each training instance is typically described by a feature vector (called an input vector). It should be associated with a desired output (a concept, a class), which is encoded as another vector, called the desired output vector. In our study, several methods were tested to train the feed-forward neural networks.

5.1 Training with Scaled Conjugate Gradient Method

After analyzing the performance of the algorithms [13], we chose the one with high classification accuracy and low training time. Under these conditions we selected SCG to continue with our experiments. 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. SCG belongs to the class of Conjugate Gradient Methods, which show super-linear 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 when using other training methods and neural networks tested, as standard back-propagation and cascade neural network.

6 Training Process and Experimentation

We made two kinds of experiments, one with Linear Prediction Coefficients (LPCs) and the other with Mel-Frequency Cepstral Coefficients (MFCCs). The selection of samples for training and testing was done at random. Training stops when the maximum number of epochs is reached, or when the maximum quantity of time has been exceeded, or when the performance error has been minimized. To be sure the performance is at an acceptable level, in terms of accuracy and efficiency, we used the 10-fold cross validation technique [10]. The sample set was randomly divided into 10 disjoint subsets, each time leaving one subset out for testing and the others for training. After each training and testing process, the classification scores were collected, and a new complete process started. In this way, we performed 10 different classification cycles, until all data subsets were used once for testing. All the experiments are done without ever using the same training data for testing. Once the 10 experiments were done, the overall scores were calculated from the average of all the individual ones. For the MFCCs and LPCs analysis, the samples were segmented in windows of 50 ms and 100 ms for different experiments. We extracted 16 and 21 MFCCs per window. Depending on coefficients number and window length, we got different parameters number for each sample. For example, with 16 coefficients for a window length of 50ms for a one second sample, the features vector contains 320 parameters, corresponding to 320 data inputs to the neural network. In this situation, the dimension of the input vector is large, but the components of the vectors are highly correlated

(redundant). It is useful in this situation to reduce the dimension of the input vectors. An effective procedure for performing this operation is the Principal Component Analysis (PCA). After several tests, we got good results with 50 parameters by each vector or pattern [13].

7 Data Set

A set of 116 samples have been directly recorded from 53 babies by pediatricians, with digital ICD-67 Sony digital recorders, and then sampled at 8000 Hertz. The same pediatricians, at the end of each recorded sample, do the labeling. The collection of pathological samples is done by a group of doctors specialized in communication disorders, from babies already diagnosed as deaf by them. The babies selected for recording are from just born up to 6 month old, regardless of gender. The corpus collection is still in its initial stage, and will continue for a while. 116 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 879 to pathological cry. For the reported experiment, we took the same number of samples for each class, 157.

8 System Implementation

For the acoustic processing of the cry waves, we used Praat 4.0.2 [11] to obtain the LPCs and MFCCs. To perform pattern recognition, a 50 input nodes – 15 hidden layer nodes – 2 output nodes, feed-forward network was developed for training and testing with LPC and MFCC samples. The number of nodes in the hidden layer was heuristically established. The implementation of the neural network and the training methods were done with the Neural Networks Tool Box of Matlab 6.0.0.88 [12]. The same Matlab version was used to implement the PCA algorithm.

9 Experimental Results

To establish the adequate number of principal components, we made an analysis on the information each number preserves. The Fig. 4 shows the preserved information by principal components from 1 to 928. For example, the 10 first components keep 90.89%, while the 30 first components keep 93.08% from the original features. The original LPC features vector was reduced to different number of principal components. Their performance was evaluated by measuring the precision of the classifier with vectors containing between 10 to 110 principal components (Fig. 5). As can be observed, up to 50 components, the recognition accuracy increases as the number of principal components also increases. From 50 components on, the precision slightly

decreases (Fig. 5). Based on this analysis we selected 50 to be the size of the input vector.

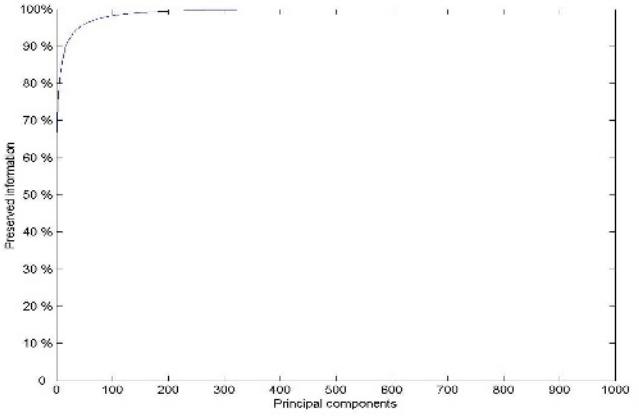

Fig. 4. Preserved information to different number of principal components

We did not analyzed more than 110 components because our goal was to reduce the original features vector to a manageable size. These results were obtained with the SCG network configured by 50, 15 and 2 nodes, in the input, hidden, and output layers, respectively.

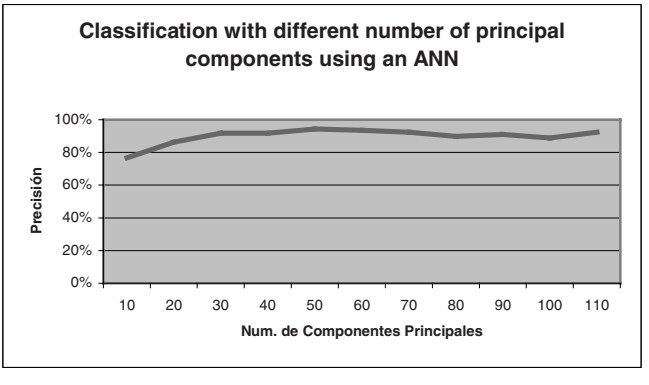

Fig. 5. Accuracy achieved by using different number of principal components.

9.1 Training and Classification Results

The neural networks were trained to classify the cries into normal and pathological classes. For training 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. 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. Some of the best results obtained for both types of features, LPC and MFCC are shown in the following confusion matrices in Table 1 and Table 2 respectively. In both cases, the results were produced by the net from a 50 principal components input vector. The reduced vectors come from the original feature vectors, which are one second length samples divided in 21 coefficients per 100 ms window. Table 1 shows the results obtained with LPC features, and Table 2 shows the corresponding to the MFCC features.

Table 1. Infant Cry classification for LP Coefficients

Type of Cry	# of Samples	Confusion Matrix		Classification
		Normal	Deaf	
Normal	157	**150**	7	
Deaf	157	11	**146**	
Total	314			**94.3 %**

Table 2. Infant Cry classification for MFC Coefficients.

Type of Cry	# of Samples	Confusion Matrix		Classification
		Normal	Deaf	
Normal	157	**149**	8	
Deaf	157	2	**155**	
Total	314			**96.80 %**

10 Conclusion and Future Work

This work has shown that the results obtained when using the MFCCs features are better than LPCs features in the test. Besides observing the neural network's performance, we have gathered useful acoustical information on the infant cry. We hope this information could be helpful to pediatricians and doctors in general. As can be noticed ,in the confusion matrices, still many samples form one class can be confused as belonging to the other. We are working to explain why that happens, in order to avoid the problem and to improve the classification accuracy. At this moment we are starting new experiments to also identify asphyxia in new born babies, by their crying. As

future work we consider to collect enough samples to train the classifiers appropriately and to have some other classes to classify. We still intent to experiment with mixed features, as well as with hybrid intelligent classification models. Moreover, we will intent to identify the degree of deafness.

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] Bosma, J. F., Truby, H. M., & Antolop, W. *Cry Motions of the Newborn Infant*. Acta Paediatrica Scandinavica (Suppl.), 163, 61–92. 1965.

[2] Sergio D. Cano, Daniel I. Escobedo y Eddy Coello, *El Uso de los Mapas Auto-Organizados de Kohonen en la Clasificación de Unidades de Llanto Infantil*, Grupo de Procesamiento de Voz, 1er Taller AIRENE, Universidad Catolica del Norte, Chile, 1999, pp 24–29.

[3] Marco Petroni, Alfred S. Malowany, C. Celeste Johnston, Bonnie J. Stevens. *Identification of pain from infant cry vocalizations using artificial neural networks* (ANNs), The International Society for Optical Engineering. Volume 2492. Part two of two. Paper #: 2492-79. 1995. pp.729–738.

[4] O. Wasz-Hockert, J. Lind, V. Vuorenkoski, T. Partanen y E. Valanne, El Llanto en el Lactante y su Significación Diagnóstica, Cientifico-Medica, Barcelona, 1970.

[5] Ekkel, T. "Neural Network-Based Classification of Cries from Infants Suffering from Hypoxia-Related CNS Damage", Master Thesis. University of Twente, 2002. The Netherlands.

[6] Huang, X., Acero, A., Hon, H. "Spoken Language Processing: A Guide to Theory, Algorithm, and System Development", Prentice-Hall, Inc., USA, 2001.

[7] Markel, John D., Gray, Augustine H., *Linear prediction of speech*. New York: Springer-Verlag, 1976.

[8] LiMin Fu, Neural Networks in Computer Intelligence, Ed-McGraw-Hill Inc., 1994.

[9] Moller, A Scaled Conjugate Gradient Algorithm for Fast Supervised Learning, Neural Networks, 6 (4), 1993, 525–533.

[10] Haykin, Simon S. *Neural Networks*: *A Comprehensive Foundation*. New York: Macmillan College Publishing Company, Inc., 1994.

[11] 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. February, 2002.

[12] Manual *Neural Network Toolbox*, Matlab V.6.0.8, Development by MathWoks, Inc.

[13] Orozco-García, J., Reyes-Garcia, C. A. "Acoustic Features Analysis for Recognition of Normal and Hypoacoustic Infant Cry Based on Neural Networks", in Lecture Notes in Computer Science 2687: IWANN 03, Artificial Neural Nets Problem Solving Methods, Springer, Berlin, pp. 615–622, ISBN 3-540-40211, ISSN 0302-9743.