

# A Radial Basis Function Network Oriented for Infant Cry Classification

Sergio D. Cano Ortiz<sup>1</sup>, Daniel I. Escobedo Beceiro<sup>1</sup>, and Taco Ekkel<sup>2</sup>

<sup>1</sup> Group of Voice Processing, Faculty of Electrical Engineering, University of Oriente,  
Ave.Las Américas s/n, 90900 Santiago de Cuba, Cuba  
{scano,daniele}@fie.uo.edu.cu

<sup>2</sup> Dept. of Computer Science, University of Twente, The Netherlands  
taco@utwente.nl

**Abstract.** Several investigations around the world have been postulated that the infant cry can be utilized to assess the infant's status and the use of artificial neural networks (ANN) has been one of the recent alternatives to classify cry signals. A radial basis function (RBF) network is implemented for infant cry classification in order to find out relevant aspects concerned with the presence of CNS diseases. First, an intelligent searching algorithm combined with a fast non-linear classification procedure is implemented, establishing the cry parameters which better match the physiological status previously defined for the six control groups used as input data. Finally the optimal acoustic parameter set is chosen in order to implement a new non-linear classifier based on a radial basis function network, an ANN-based procedure which classifies the cry units into a 2 categories, normal-or abnormal case. All the experiments were based on the physioacoustic model for cry production and the Golub's muscle control model.

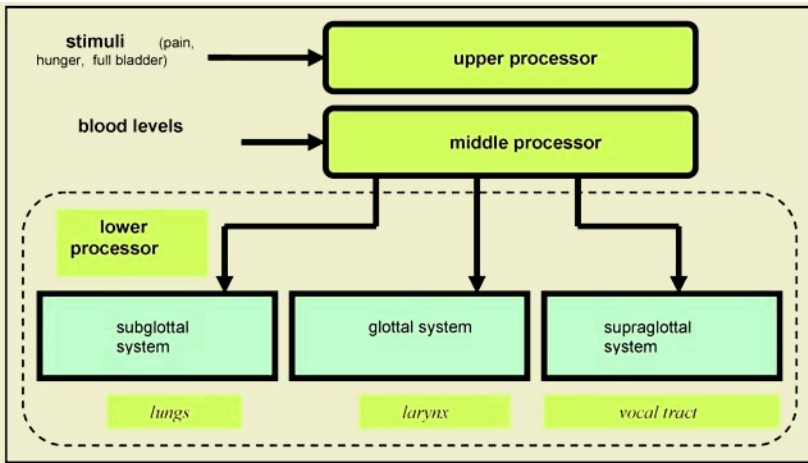
## 1 Introduction

Last research has reported the effectiveness of cry analysis for the study of infant's neurological status. Since this early work, there have been a number of advances that have broadened the potential application of cry analysis not only for cry generation area but also for cry perception. Some of them are: a better understanding of the mechanisms of cry production, getting factors that place the infant at risk for later handicap, such as prematurity, low birth weight, obstetrical complications [8], the cry as an evolutionary function of helping to ensure the survival of the infant [1], cry as a developmental function that make easy the development of social, perceptual, and cognitive skills [1,7], and so on. Recently the use of ANN has been one of the clue in the road mainly in the field of the Infant cry analysis oriented for diagnosis [4,9]. In this work, a RBF network is implemented and tested. Firstly the data acquisition and the structure of control groups are explained (see Section 2). Then the cry parameters which better matches the physiological status of the child are determined leading to the optimal acoustic parameter set (see Section 2). The RBF network performance and results are commented in Section 3. Finally, some concluding remarks and suggestions are also given in Section 4.

## 2 Materials and Methods

In this paper we address some general aspects concerned with the Golub's theory [6]. The Golub's hypothesis assumes that muscle control is accomplished within three

levels of central nervous system processing as we see in Fig. 1. The *upper processor* is involved in choosing and modulating the state of action of the child and is probably where the more complex feedback from external and internal factors is collected and acted upon. Of course, during the neonatal period, this higher processor may be relatively immature and precise control infrequent. As a result, at this stage of maturation, many activities occur in a reflexive manner. The *middle processor* encircles all vegetative states such as: swallowing, coughing, respiration, bowel movements, crying, etc. The stimuli that help the upper processor to choose the appropriate vegetative state include hunger, pain, hypoxemia, or hypercapnia, and a full bladder. The initiation of each of the previously mentioned vegetative states, in response to a stimulus, must result in the control of a large number of muscles.



**Fig. 1.** The Golub’s Muscle Control Model

Most likely, there exists some sort of “coordinative structure” that makes it unnecessary for higher processors to exert control of each individual muscle within a muscle group. For example, swallowing involves a well-defined control sequence that is kept relatively constant from one swallow to the next. Upper processor control of each muscle involved in swallowing would be that following the stimulus for swallowing the upper processor triggers the “swallow box” in the middle processor and the lower processor acts to control the muscular movements involved in swallowing. The newborn cry is envisioned as resembling the same type of process as the vegetative status mentioned above. Following the cry stimulus, the upper and/or middle processors for any production trigger the lower processor control of the relevant muscle groups.

Based on these muscle control hypothesis, we assume that each of the three muscle groups important for cry production are controlled independently. Then parameters that each are responsible for are likely to vary independently. Secondly, if we can pinpoint differences in the cry as caused by subglottal (respiratory), glottal (laryngeal), or supraglottal malfunctions, then we will be able to correlate the acoustic abnormality with specific physiological or anatomical abnormalities [6].

The cry recordings were made at the Eastern Maternity Hospital of Santiago de Cuba. A pain cry was induced by a standardized stimulus: a “heelstick”, the infants

were positional decubito supine and flat in open cribs and were not crying at the time of the cry stimulus. The digitization of the cry signals was made by an A/D acquisition system connected to an IBM-PC. The data set consists of a successive cry episodes in which the average values for one cry unit is considered as a single pattern, resulting in 187 patterns to be evaluated. The cry episodes were obtained from 35 neonates, nineteen healthy cases and sixteen sick neonates with hypoxia-based CNS diseases, considering six control groups of data (under the specialized criterion of physicians) as follow in Table 1:

**Table 1.** The control groups

Control Groups	
Normal Group 1:	normal transpelvian delivery
Normal Group 2:	cesarean dysthociac
Pathologic Group 3:	hypoxemia
Pathologic Group 4:	retarded intra-uterine growth
Pathologic Group 5:	hypoxemia with another risk
Pathologic Group 6:	hyperbilirrubin

Every pattern consists of 25 values as shown in Table 2,

**Table 2.** Example results from the feature extraction process

Feature	Parameters
$F_0$ (mean, min., max., std dev , variability)	5
$F_1$ (mean, min., max., standard deviation , variability)	5
Voicedness (mean, min., max., std dev , variability)	5
Energy (mean, min., max., std dev, variability)	5
First latency	1
Rising melody type	1
Stridor occurrences	1
Shift occurrences	1
Class	1

The goal of this research is to use crying features to automatically classify infant cries into those produced by children suffering from hypoxia-induced damage to the central nervous system, and those produced by children not suffering from this damage.

## 2.1 Pre-processing

A set of features of 25 different variables that could serve as input are quite available. The goal of *pre-processing* in general is to present and /or transform the input data in such a way that it can be optimally processed by a neural network (to reduce dimensionality). A process called *feature selection* was developed: to try every possible feature that can be constructed from the whole feature set and then selecting the subset yielding the best results. The feature selection process was then optimized in order to make it manageable in terms of available resources.

## 2.2 Feature Subset Testing Algorithm

Given the nonlinearity in the data set, we are forced to discard linear discriminant analysis as a viable option for producing a performance criterion for a data set, and must turn to non-linear methods such as neural networks [3,5]. When choosing a neural network model for testing subsets, it is important to choose one that has little manual parameters to be tuned. For example, some networks need a *learning rate* to be chosen when they are trained. This value can not or hardly be determined automatically, and therefore a number of different learning rates have to be tried to find the one yielding the best results. Obviously, this multiplies the total search time needed by the number of trials needed to find this optimal parameter. To counter this, we are looking for a fast, non-parametric, supervised network model with little or no parameters to tune.

This leads us to a class of neural networks known as *radial basis function networks*. These networks have their origin in a procedure known as the *exact interpolation procedure*.

Exact interpolation works as follows: consider a mapping from a d-dimensional input space  $x$  (consisting of  $N$  points) to a one-dimensional target space. An exact interpolation procedure introduces a set of *basis functions*, one for each data point  $x^n$ . The basis functions take the form of

$$\phi (\|x - x^n\|) \tag{1}$$

where  $\phi$  is some non-linear function depending on the distance between  $x$  and  $x^n$ . Usually, the Euclidean distance metric is used. Then, the functional mapping  $h(x)$  that is performed by the procedure is constructed by linearly combining the basis functions:

$$h ( x ) = \sum_n w_n \phi (\|x - x^n\|) \tag{2}$$

Usually, the basis function  $\phi$  is chosen to be the Gaussian

$$\phi(x) = \exp\left[-\frac{x^n}{2\sigma^2}\right] \tag{3}$$

where  $\sigma$  is the 'width' of the basis function, a parameter controlling the smoothness of the total approximation  $h(x)$ .

Many basis functions, similar to the Gaussian, will in fact work more or less satisfactory. The Gaussian is often chosen because of its analogy to the normal distribution, on which the probability density functions of many natural processes are based [5]. The described exact interpolation procedure is known as a Probabilistic Neural Network (PNN) [3]. As you see in this algorithm the only parameter to be tuned is the width  $\sigma$  of the basis functions. In MATLAB, an algorithm was constructed for performing the search, using the classification rate of the accelerated PNN as criterion. The Table 3 shows the features that will be used according the feature search results:

**Table 3.** The best features

Feature	Appears in best feature set
First latency	71%
Voicedness mean	60%
$F_1$ minimum	59%
$F_1$ mean	52%
$F_0$ mean	51%
Voicedness variability	41%
Energy maximum	30%
Energy maximum	30%
$F_0$ maximum	30%
Energy mean	26%

$F_1$ : First formant     $F_0$  Fundamental frequency (pitch)

### 3 Results and Discussion

#### 3.1 Building the Classifier

Normally, after the pre-processing stage, one would make the step to actual classification. In this case, classification has already been performed in the selection of the best feature set. However, as we optimized the classification algorithm for speed, there are probably some improvements that can be made in the classification stage. Apart from that, we should end up with an 'out-of-the-box'-classifier that can be used on new data, so this is the place to construct it. We already discussed exact interpolation as a powerful classification algorithm which has a flaw, namely that it tends to over-fit the data set. The interpolation function yielding the best generalization is the one which is typically much smoother than the one that is the result of exact interpolation. To achieve this smoothness, the algorithm is adjusted so that the complexity of the underlying function, rather than the size of the data set, determines the number of basis functions.

Using MATLAB neural network toolbox, a RBF network was constructed to classify the data set resulting from the feature search. The data sets maintain their respective sizes in terms of number of patterns; however, the dimensionality of the sets has been reduced to 10.

First a simple classification algorithm based on a single RBF network is constructed. As with the feature selector however, the resulting network is quite unstable, so, just as in the feature selection process, a kind of averaging mechanism is devised. This time however, it is not enough to simply average the classification rate since the output itself is of critical importance in this version of the classifier (as opposed to the feature search algorithm where it is the classification rate that is the only important parameter – the actual results did not yet). Instead, a mechanism called bagging- shorthand for bootstrap aggregation- is introduced [ 2].

#### 3.2 Bootstrap Aggregation

Bootstrap aggregation is a form of *arcing*: adaptive re-weighting and combining. This general term refers to reusing or selecting data in order to improve classification [5].

In bagging, several versions of a training set, each created by drawing samples with replacement from the data set, are created. Each training set is then used to train a classifier (RBF network). Then, in the classification stage, the outputs of all those networks are combined to form the final classification based on the vote of each individual network (also called component classifier).

In this multiclassifier system, all component classifiers are of the same type and share the same output format. The output of one component classifier may simply be seen as the probability that some case belongs in a certain class (or the probability that the case does not belong in the other class). We may therefore simply linearly add the probabilities provided by the component classifiers and average them. After this, rounding the average value directly provides us with the classification result.

A multiclassifier system of 10 networks was constructed. The performance on the data set was 85%. A confusion matrix is shown in Table 4, it shows a slight inclination towards type 1 errors ('false alarms') which is a 'good' thing (in medical terms, a false alarm could be considered less damaging than unjustly *not* reporting an abnormality). The resulting networks are 'fixed' by writing them to a MATLAB variable store so they can be used in real classification tasks.

The classifier's performance (correct classification rate) was 85% (with +/-5% on the confidence interval).

**Table 4.** Confusion matrix for the multiclassifier system

Actual class	Group size	Predicted class	
		0	1
0	15	79%	21%
1	21	12%	88%

(0: normal, 1: abnormal)

## 4 Conclusions

The extraction and classification system that was devised was capable of distinguishing between normal and abnormal cases with up to 85% (with +/-5% on the confidence interval). This result can be regarded as statistically significant when you compare with recent works [4,9]. The resulting cry feature selection is not a *generalized paradigm* that is just the result from a specific approach into the cry analysis: *the use of ANN for a cry unit classification*. However another approaches concerned with the best parameter selection should be developed in the future. The classification performances that were found quantify the network's ability to relate some examples of measured values to some examples of diagnosed classes.

Although the results presented here are preliminary in nature, more tests are currently ongoing. Different features for input into the neural network and other network architectures are currently being investigated so that it can be determined which features are the most suitable for correct classification of cry units.

## References

1. Bell, R.Q.: Contributions of human infants to caregiving and social interaction. In M.Lewis & L.Rosenblum (Eds), *The effect of the infant on its caregiver*. New York: Wiley, (1974) pp 1-19

2. Breiman L., *Bagging Predictors*, Journal of Machine Learning, vol. 24 no. 2 (1996) pp123-140
3. Bishop C., *Neural Networks for Pattern Recognition*, Oxford University Press, (1995) , ISBN 0198538642
4. Cano S. D., Escobedo D. I., *El uso de los mapas auto-organizados de Kohonen en la clasificación de unidades de llanto infantil*, Proceedings of the CYTED-AIRENE Project Meeting, Universidad Católica del Norte, Antofagasta, Chile (1999) pp 24-29
5. Duda R., Hart Po, Stork D., *Pattern Classification, 2 nd edition*, John Wiley & Sons, Inc., ISBN 0-471-05669-3, (2001).
6. Golub H., Corwin M., *Infant cry: a clue to diagnosis*, Pediatrics, vol. 69 (1982), pp. 197-201.
7. Gustafson, G.E. & Green, J.A: On the importance of fundamental frequency in cry perception and infant development. Child Development, 60, Aug. (1989).
8. Lester, B.M.: A biosocial model of infant crying. In Leipsitt, L. & Rovee, C. (Eds.). Advances in Infancy Research. Norwood. N.Y. Ablex, (1984) pp 167-207
9. Schonweiler R., Kaese So, Moller S., Rinscheid A., Ptok M., *Neuronal networks and self-organizing maps: new computer techniques in the acoustic evaluation of the infant cry*, International Journal of Pediatric Otorhinolaryngology 38 (1996) pp. 1-11
10. Wasz-Hockert O. Et al: *The infant cry: a spectrographic and auditory analysis*, Clin. Devo Med. 29 (1968) pp. 1-42.