

# MULTI-STAGE NONLINEAR CLASSIFICATION OF RESPIRATORY SOUNDS

*E. Çağatay Güler<sup>†</sup>, Bülent Sankur<sup>‡</sup>, Yasemin P. Kahya<sup>‡</sup>, and Sarunas Raudys<sup>§</sup>*

<sup>†</sup>Biomedical Engineering Institute, <sup>‡</sup>Electrical Engineering Department,  
Boğaziçi University, Bebek-80815, İstanbul, Turkey

<sup>§</sup>Institute of Mathematics and Informatics, Akademijos 4, 2600, Vilnius, Lithuania  
e-mail: {gulerc, sankur, kahyay}@boun.edu.tr, Sarunas.Raudys@DAS.MII.lt

## ABSTRACT

The three-class recognition problem of respiratory sounds based on multi-stage decisions is addressed. The method consists of dividing respiratory cycles of patients into phases, and classifying each phase with a separate multilayer perceptron, called the “phase expert”. Each phase information consists of several time segments and their parametric representation. Expert decisions on phase segments are then combined by a decision fusion scheme, simulating a consultation session. Thus in the first stage of hierarchy one uses signal features to reach segment decisions, while in the second stage one uses decision votes themselves as features inputted into a second classifier. Furthermore a new regularization scheme is applied to the data to stabilize training and consultation.

## 1 INTRODUCTION

Automatic classification of respiratory sounds is significant in that it provides a computer-aided tool to auscultation and increases its potential diagnostic value. The difficulties encountered in developing such a tool are diversities in lung sounds of different subjects due to age, weight, sex, and hence, the need to form a very large database. The use of conventional classification algorithms have been found to be inadequate for such a problem and, consequently, these methods had to be enhanced to combat such problems as the small sample size, diversity of sounds, and the cyclic behaviour of signals.

The classification problem of respiratory sound signals has been addressed by taking into account their cyclic nature. From cycle to cycle, respiratory waveforms can be assumed to be statistically identical, but the waveforms are

not, of course, periodic. The statistical characteristics of the process evolve in a cycle. In other words, respiratory signals are nonstationary due to changes in lung volume and flow rate during a cycle. In addition respiratory sounds have a large inter-subject variability and differences due to the evolution stage of various pathologies. Consequently, respiratory sounds are difficult signals to classify, especially, since their feature distributions are much overlapping.

In our work, these signals are first partitioned into segments that are short enough to guarantee their stationarity but large enough to allow for reliable parameter estimation. A number of these segments are joined together to form a phase of the respiration cycle, e.g, the initial inspiratory phase, when lung volume is still small but air flow rate is maximum. Finally six phases form a whole respiratory cycle. This partitioning of the signal (a) reduces the dimensionality of the feature space to a manageable level where “small sample size-high dimensionality” trade off is overcome to some extent, and (b) avails us of the possibility to zoom in the different sound production mechanisms governing each phase. The class decisions are first taken at the phase level, which are then combined into a final decision for the patient. The block diagram of the proposed classification method is given in Fig. 1.

The novelties brought to the classification problem of nonstationary cyclic data can be listed as follows: (i) Data from an evolutionary cyclic process have been re-organized into phases, and a separate phase expert (multilayer perceptron) is designed for each phase, (ii) Phase expert opinions are fused via a “consultation process”, (iii) Judiciously controlled noise injection has been implemented to stabilize the feature space and to construct regularized training sets.

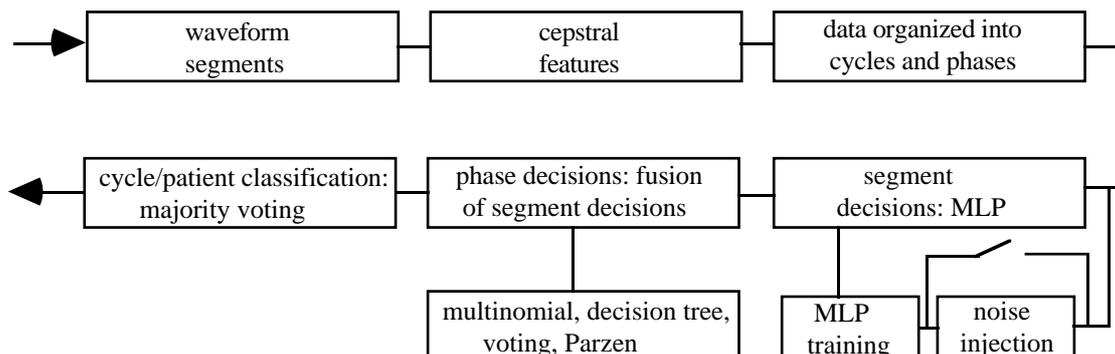


Figure 1. Block diagram of the classification method.

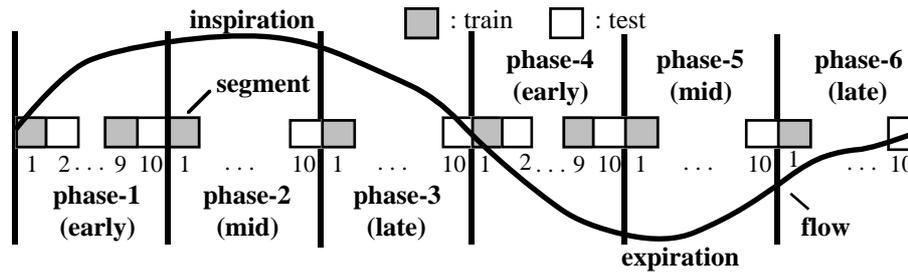


Figure 2. Division of a respiratory cycle into segments and phases, and the choice of training and test segments from each phase. These operations take place on the respiratory sound signal synchronized with the above flow signal.

The proposed classification method can be equally well applied to other cyclic biological data, such as blood flow, ECG, or to industrial sounds, like shaft, drill sounds, or to data obtained from seasonal climatic, oceanic measurements.

## 2 METHODOLOGY

Patient recordings were each 12.8 seconds in duration, a period covering 3-4 respiration cycles. Measurement records from 18 chronic obstructive patients (class-1), 19 restrictive lung disease patients (class-2) and 20 healthy subjects (class-3) were analyzed (total of 57 patients). The measurement setup for lung sound recording and processing operations are detailed in [1]. Recorded signal in a respiration cycle was divided into a fixed number of segments as illustrated in Fig. 2. The number of segments was chosen to be 60 for a whole respiration cycle, and the segments overlapped by 25%. Each segment was characterized by 6 cepstral coefficients forming a feature set. These segments in each cycle were further partitioned into 6 phases, namely, early, mid, late inspiration/expiration phases, each group consisting of 10 consecutive segments. A separate classifier was designed for each phase feature set.

The segments belonging to each phase were further split into a training and a test set. More explicitly, every other segment of a phase was assigned to the training set, while the remaining segments of the same phase were used in the test stage. The choice of such interleaving test and training sets was made purely as a proving ground for any improvements that our proposed scheme could bring.

Since feature data were sparse with respect to the

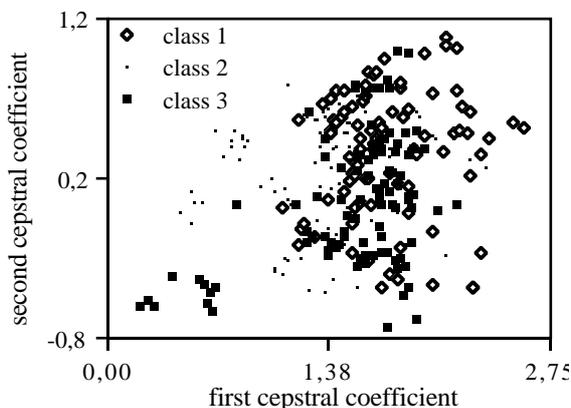


Figure 3. Scatter diagram of first vs. second cepstral coefficients from mid expiration training data.

dimensionality of the feature space, data had to be enriched. To this purpose, several runs of independent and identically distributed Gaussian noise, with zero mean and variance  $\lambda$  was added to each training set vector. If  $\mathbf{f}$  denotes a feature vector, it is replicated  $K$  times ( $K$  was taken 20) with the addition of noise to its components. A classifier designed with a sparse set may adapt itself too strongly to the training vectors due to the gaps. The added noise fills the gaps between the training vectors, and thus, in a way, regularizes and stabilizes the feature space.

Each of the 6 phases had its own classifier, referred to as the “phase expert”. In each phase there resulted 5 segment decisions for each subject (since the other 5 alternating segments are reserved for testing). The decisions on segments in different phases but with the same sequence number, i.e., segment- $i$  of phase-1, segment- $i$  of phase-2 etc., were fused via consultation. This reduced the votes for a cycle from  $5 \times 6 = 30$  to just 5. Finally, the classification of a cycle was based on majority voting of the consultation decisions on each of the 5 “fused” segments. The consultation actually takes place via a classifier operating on the pattern of class decisions from each segment.

In order to decide for an appropriate classification algorithm, an exploratory data analysis was performed based on scatter diagrams of features. In Fig. 3, a scatter diagram of a pair of cepstral features from the mid-expiration data is presented. This scatter diagram more or less is typical for all the data analyzed. It was observed that the classes overlapped substantially, while classes themselves consisted of more than one cluster in the space. Thus there was strong evidence that a classifier is needed with nonlinear decision boundaries. To this purpose a Parzen classifier and an artificial neural network were compared. Although both techniques resulted approximately in the same classification accuracy, the Parzen classifier was computationally more complicated. Therefore it was decided on the use of multilayer perceptrons (MLP) as phase experts, and for decision fusion [2].

After several preliminary classification experiments, six hidden units were chosen for each expert. In order to find the weights, a standard backpropagation MLP training algorithm [3] was used. Results of the segment classification of the training, and test sets, using cepstral coefficients are presented in Table 1.

## 3 CONSULTATION

The classification performances of phase experts on the individual segments were rather mediocre (63% correct

Table 1. Segment misclassification probabilities of the six phase training sets, and test sets using cepstral features.

Phase 1		Phase 2		Phase 3		Phase 4		Phase 5		Phase 6	
TRAIN	TEST	TRAIN	TEST	TRAIN	TEST	TRAIN	TEST	TRAIN	TEST	TRAIN	TEST
0.31	<b>0.37</b>	0.23	<b>0.28</b>	0.27	<b>0.33</b>	0.32	<b>0.43</b>	0.31	<b>0.43</b>	0.34	<b>0.41</b>

classification on the average), as documented in Table 1. Their performance was boosted up by the strategy of using a second stage decision on the pattern of segment decisions, which can be thought as “a consultation of experts”. The decisions for each segment, say  $k$ th, from each phase were grouped together into a decision vector, i.e.,  $\mathbf{D}^k = [d_1^k \ d_2^k \ d_3^k \ d_4^k \ d_5^k \ d_6^k]$ .

Obviously, if there are  $\phi$  phases/experts, and the decisions are “ $d_j^k = r_j$ ” valued, there will be overall “ $m = \prod_{j=1}^{\phi} r_j$ ” possible states. In our case  $\phi$  was 6 and segment decisions were three valued, namely class-1, class-2, class-3. For any respiratory cycle, the vector  $\mathbf{D}^k$  for segments with the same index from each phase can assume a value in one of the  $m$  bins. One way to combine the expert decisions is to use majority logic. However we conjectured that the final segment decisions could be improved by Bayesian treatment of the expert voting patterns. The conditional distribution of vectors  $\mathbf{D}^k$  given that it is from class  $c$ , is denoted by the  $m$  probabilities,  $P_1^c, \dots, P_m^c$ . The expert consultation in which a decision is made for  $\mathbf{D}^k$  according to the maximum of a posteriori probability, i.e.,  $\max_{c,j} \{P_j^c(c|\mathbf{D})\}$ , is called the multinomial classifier [4].

In our case, with 6 phase experts, there are  $3^6=729$  combinations. Such a large number of bins make the multinomial decision fusion impractical, and simpler schemes must be invoked for. Simpler schemes for the fusion of expert class decisions are the decision tree, Parzen window or  $k$ -NN classifiers. Two principle assumptions used in these algorithms are: (i) many states belonging to one class can be lumped together and described by smaller number of features, (ii) bins with small probabilities can be neglected.

A decision tree classifier [4] would consider one expert opinion at a time, i.e., one component of  $\mathbf{D}^k$ . Decisions on the test set are made by remembering class membership of the leaves and the architecture of the tree. An alternate scheme to simplify the multinomial problem would be the classifier based on the nonparametric Parzen window estimate of the multivariate distributions [4].

#### 4 RESULTS AND DISCUSSION

Recall that each decision vector,  $\mathbf{D}^k$ , has 6 elements of which each can take the value either 1 or 2 or 3, i.e., the class membership decision given by a phase expert, and a class decision is reached by one of the consultation schemes for each  $\mathbf{D}^k$ . This procedure is applied five times to each of the 5  $\mathbf{D}^k$ 's resulting from every cycle of respiration.

The four decision fusion algorithms, namely, the multinomial classifier (M), the decision tree classifier (DT), Parzen window classifier (P), voting (V) schemes have been

compared. The improvement in the classification performance of segments (decision patterns,  $\mathbf{D}^k$ ) by fusing expert decisions are shown in Table 2. It can be observed that all fusion schemes bring significant improvements but to varying degrees. The segment classification error of phase experts on the test sets is at an average of 37.5% (Table 1), while after the consultation, segment classification error drops down to 34%, 24.2%, 17.9%, 21.8% for the M, DT, P, and V schemes, respectively.

However one notices, also, a large discrepancy between training and test performances of consultation schemes. The discrepancy is most severe for the M case and can be explained on the basis of the sparsity of the data. In fact the M scheme has 729 states, but is trained with a set of 285 vectors overall. As can be noticed from Table 2, P and DT schemes give error rates of 17.89% and 24.21% on the test set, respectively, while this value is 34.04% for the M scheme. In other words, as such, the multinomial decision fusion turns out to be useless. Thus simplification of the M type of decision fusion is justified. Simpler consultation schemes are more robust in that they do not get overtrained. In Table 2, the DT had 49 final leaves (instead of 729 bins), while the P window parameter is  $\lambda=1.6$ .

A second alternative for the small sample size problem is the enrichment of the feature space by noise injection to the training data, i.e., the construction of a regularized training set. Noise injection to the feature space is similar to Parzen kernels, in that each measured feature points is interpreted as signaling other potential vectors in its neighborhood. We added noise to features in each phase, with an empirical standard deviation,  $\sigma$ . The “regularized training decision set” constructed from twenty runs of independent Gaussian noises with  $\sigma=0.0645$  injected into the features of six phase training sets resulted in the best M performance for the “test decision set”. In Table 3, “test decision set” segment classification errors of M and DT designed with the regularized training set are presented. Note from Tables 2 and 3 that the use of the regularization set instead of the actual training set results 12.2% and 4.6% improvements in the performances of M and DT consultations, respectively. Thus with noise injection both the multinomial and decision tree classifiers improve to a level comparable to that of the Parzen scheme.

The third stage of classification is in fact to reach a class decision for the whole cycle of a subject. This is achieved by merging the 5 consultation decisions resulting from the 5  $\mathbf{D}^k$ 's of a cycle via majority logic. The misclassification probabilities of subjects based on majority voting of the M, DT, P and V consultation decisions on  $\mathbf{D}^k$ 's are presented in Table 4. Subject misclassification probabilities of the M and DT schemes designed with the regularized training set are also given in Table 4. Table 4 states that the performance of the P scheme is the best with 14% error rate, and the DT classifier with 15.8% error follows P.

Table 2. Average segment (decision vector) classification errors using various consultation schemes.

Multinomial		Decision Tree		Parzen Window		Voting	
TRAIN	TEST	TRAIN	TEST	TRAIN	TEST	TRAIN	TEST
0.028	<b>0.340</b>	0.077	<b>0.242</b>	0.147	<b>0.179</b>	0.165	<b>0.218</b>

Table 3. "Test decision set" segment classification errors for the multinomial and decision tree schemes designed from the regularized training set.

Multinomial	Decision Tree
0.218	0.196

The performance of the DT is 1.7% better than V scheme in subject classification, while V was 2.4% better compared to DT in segment classification (Table 2). On the other hand, the use of the regularized training set, i.e., noise injection to the training space, in the design of M and DT schemes is once more justified: Subject classification performances of M and DT are improved 19.3% and 1.8%, respectively.

Table 4.(a) Subject classification errors based on majority voting of various consultation decisions.

Multinomial	Decision Tree	Parzen	Voting
0.333	0.158	0.140	0.175

Table 4.(b) Subject classification errors (again with majority logic) of the multinomial and the decision tree classifiers designed using the regularized training set.

Multinomial	Decision Tree
0.140	0.140

Finally considering again the "small sample size" situation, the number of experts, hence the complexity of the decision mechanism, can be reduced. Thus, the above schemes were run again but with smaller number of experts. Preliminary experiments showed that the subject classification error of DT designed with 3 of the best experts was improved to 0.105. Therefore the simplification of the multinomial scheme by the use of the decision tree classifier with smaller number of experts and by noise injection exploits the decision vectors of a subject in a more reasonable way.

## 5 CONCLUSIONS

In this work, hierarchical decision fusion schemes based on the cooperation of neural networks to classify cyclic respiratory sound patterns were investigated. The small sample size problem was addressed in the form of feature space stabilization by calculated noise injection, that is construction of regularized sets from training sets.

The average correct classification performance of phase experts on individual segments was at 60s% (Table 1). It

was also noticed that the duration of segments did not affect the classification performance much. For example, a whole cycle could be divided into 60 segments or the whole cycle could be considered as one big segment. In either case the average correct classification performance was approximately the same. What brought in significant improvements was the nonlinear treatment, that is decisions on the "decision patterns". Thus, the use of the consultation schemes, i.e., fusion of class decisions given by phase experts, improved the correct classification performance to 80s% (Tables 2, 3 and 4).

The multinomial classifier, the decision tree classifier, the Parzen window classifier, and simple voting were used as alternate decision combination algorithms. Parzen window approach was the best decision fusion mechanism. It was shown, however, that once the small sample size problem is overcome with noise injection both the multinomial and decision tree approaches became comparable with Parzen. In fact the "small learning set size-complexity of the multinomial classifier tradeoff" can be overcome to some extent with the use of (1) the Parzen window approach (2) the decision tree classifier, (3) regularization, and (4) smaller number of phase experts. We also showed that the design of a decision tree classifier with smaller number of phase experts and with the regularized training set may be helpful in expert consultations.

The results obtained via a multistage decision scheme are promising in developing a cheap, noninvasive, auscultation based, assisting diagnostic device for physicians which can be implemented on a PC.

## Acknowledgment

This work was supported by TÜBİTAK under contract EEEAG-127.

## REFERENCES

- [1] B. Sankur, Y. P. Kahya, E. Ç. Güler, T. Engin, Comparison of AR based algorithms for respiratory sounds classification, *Comp. Biol. Med.*, vol. 23, no. 4, pp. 67-76, 1994.
- [2] F. Fogelman-Soulie, E. Vinnet, B. Lamy, Multimodular neural network architectures: application in optical character and human phase recognition, *Int. J. on Pattern Recognition and Artificial Intelligence*, vol. 7, pp. 721-755, 1993.
- [3] R. Hecht-Nielsen, *Neurocomputing*, Addison-Wesley Publishing Comp., 1989.
- [4] S. Raudys, *Fundamentals of Statistical Pattern Recognition for Connectionist Students*, Submitted to Publication, 1995.