CLASSIFICATION OF VOICELESS PLOSIVES USING WAVELET PACKET BASED APPROACHES

Ewa Lukasik
Poznań University of Technology, Institute of Computing Science,
ul. Piotrowo 3a, 60-965 Poznań, POLAND
tel: +48 61 665 2373, fax: +48 61 877 1525
email: lukasik@put.poznan.pl

ABSTRACT

There are contradictory reports on the usefulness of the Wavelet Packet Transform for feature extraction. In this paper we continue the investigation of this subject with reference to non-stationary speech signals, namely unvoiced plosive consonants /p/, /t/, /k/. We concentrate on the influence of the feature reduction method on the classification rate. Two strategies have been applied: feature selection, performed using the Local Discrimination Basis and feature projection performed using Primary Components Analysis (Singular Value Decomposition). Classification has been performed by cluster analysis and neural network. The classification results obtained for PCA outperform those for LDB and other methods examined earlier.

1. Introduction

Signals possessing non-stationary character are not well suited for detection and classification by traditional Fourier methods. An alternate means of analysis is sought, so that valuable time-frequency information is not lost. The Wavelet Packet Transform (WPT) is one of such time-frequency analysis tools.

Relatively little attention has been paid to WPT as a basis for pattern recognition as compared to compression and denoising tasks. Examples of using wavelet transform and wavelet packet transform for feature selection come from biological signals: ECG [1], myoelectrical signals [2], underwater acoustic signals [3] and in musical acoustics [4].

In the paper we examine the feasibility of using the WPT in automatic classification of context independent voiceless plosives /p/, /t/, /k/. These are speech signals of non-stationary character. However performing the wavelet packet transform on the source data set is only the first step in the processing stages of classification, namely feature extraction. The essential task is to perform the features dimensionality reduction to yield the minimized data set to a classifier. There is a multitude of methods of dimensionality reduction [11,12]. However two main strategies are to be identified: feature selection and feature projection. The first strategy may be performed e.g. through the Local Discriminant Basis (LDB) algorithm described in [9] and recalled in [2]. The feature projection method is performed using the Principal Components Method (PCA) usually computed by means of Singular Value Decomposition (SVD) [10]. The input data matrices are composed of the entropy bins of wavelet packet transform [3, 2]. Other possible approach to reducing the features set using wavelet packet transform is through the best basis search [5]. Classification has been performed by cluster analysis and using neural network.

2. Data characteristics

The particular group of speech signals under consideration is a category of voiceless plosives /p/, /t/, /k/. Therefore three classes have to be distinguished. The data for experiments have been taken from speech database for Polish - CORPORA [8]. The set of 334 context independent utterances from 2 male speakers have been analysed (129 /t/, 111 /k/, 94 /p/, sampling frequency 16 kHz). It should be noted that each speech sample came form different, unrepeatable utterance. The segment length varied from 150 up to 1024 samples.

3. Wavelet Packet Transform

Wavelet Packet Transform (WPT) [5] can be viewed as a generalized version of the wavelet transform providing level by level transformation of a signal from the time domain into the frequency domain. It is calculated using a recursion of filter-decimation operations leading to the decrease in time resolution and increase in frequency resolution. The frequency bins, unlike in wavelet transform, are of equal width, since the WPT divides not only the low, but also the high frequency subband.

4. Feature extraction and reduction

4.1. Introductory remarks

Figure 1 represents the processing stages leading to final data classification.

The feature extraction stage is the transformation upon the measured "raw" signals, producing an original feature set. In our case this is the full wavelet packet decomposition of measured data.

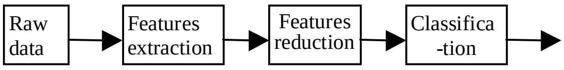

Fig.1. Processing stages in data classification

This set is the subject to the dimensionality reduction yielding a smaller feature set, which is more suitable for the presentation to a classifier. Dimensionality reduction strategies may be characterised either as feature selection or feature projection. The feature selection approach attempts to reduce the number of variables by selecting the best subset of the original feature set, according to some criterion. It is usually based upon the class separability measure, e.g. mean energy, entropy or Euclidean Distance.

Feature projection is performed by principal components analysis (PCA) that provides a linear map with the minimum mean square criterion. PCA effectiveness in pattern recognition is due to its ability to eliminate linear dependencies and uncorrelated noise in the data. It computes a set of orthonormal vectors or "components" such that the sample variances of the elements are maximized. So principal components analysis finds a set of vectors such that when the training data is projected onto these vectors, maximum variance is obtained. Principal Components analysis can be completed by finding eigenvectors that have largest eigenvalues. For numerical reasons, singular value decomposition is usually used for this purpose.

4.2. Features selection using Local Discriminant Basis

Feature selection attempts to select the minimally sized subset of features according to the following criteria [11]:

- the classification accuracy does not significantly decrease; and
- the resulting class distribution, given only the values for the selected features, is as close as possible to the original class distribution, given all features.

Ideally, feature selection methods search through the subsets of features, and try to find the best ones among the competing 2^N candidate subsets according to some evaluation function. However this procedure is exhaustive as it tries to find only the best one. It may be too costly and practically prohibitive, even for a medium-sized feature set size (*N*). Other methods based on heuristic or random search methods attempt to reduce computational complexity by compromising performance. These methods need a stopping criterion to prevent an exhaustive search of subsets. Feature selection in signal processing applications is rather specific task and specially devoted methods have been proposed. One of them is Local Discriminant Basis search algorithm [9,2]

The basic idea of Local Basis discrimination can be described as the best basis search [5] algorithm over the calculated discriminant measure D between classes. It represents the measure of class separability. The input parameters to D are the time-frequency energy maps of each class calculated by accumulating the squares of the WPT coefficients for each entry in the binary packet tree and normalized by the total energy of the signal belonging to given class. Then the distance measure (cost function) has to be introduced. In our case it is or relative entropy:

$$D_{s,q} = s_j \, log(s_j/q_j) \qquad (1)$$

or Euclidean Distance:

$$D_{s,q} = | s_j - q_j | \qquad (2)$$

where s_j and q_j are the features characterizing elements of two classes, j=1,....,n.

To compute the discrepancy between the distributions of the three classes of plosives under consideration, one must sum up $\binom{3}{2}$=3 pairwise combinations of D:

$$D = D_{p,t} + D_{t,k} + D_{k,p} \qquad (3)$$

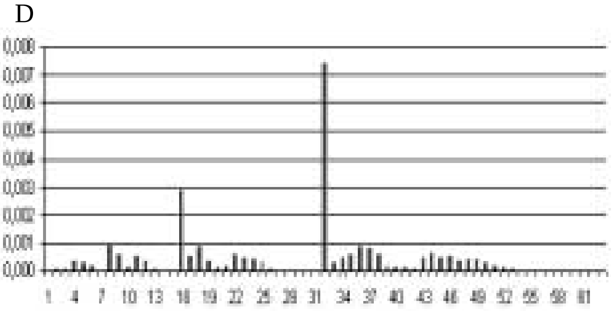

Fig.2. Distribution of entropy discriminant measure in LDB

Figure 2. shows the values of discriminant measure, relative entropy (1) for all bins in WPT decomposition of the set of plosive consonants under investigation. Since the decomposition depth J=5, number of bins B=63. It should be noted that the choice of discriminant measure is not crucial for features selection. In every case the separability is not high.

4.3. Feature projection using Singular Value Decomposition

The method consists in calculating the full WPT of each signal in the training set, then creating the energy (entropy) map for each signal from its WPT, organize it into energy (entropy) matrices, one for each class and calculate the singular vector for each class [3]. Next step is to determine the parsimonious set of features from the most significant singular vectors indicating satisfactory class separation. As the tests showed that using entropy function cost gives better results than the energy, the entropy function (4) was used for further tests:

$$e(s) = -\sum_j s_j^2 \, log \, s_j^2 \qquad (4)$$

s - denotes feature under consideration.

For each representative signal and each of its packets we calculate the entropy and create the energy (entropy) map. For easier manipulation the map is represented as a column vector using lexicographic order of the bins $\mathbf{e}_{r,i}$., where r denotes the class and i - index of signal realization. In our case r stands for one of three classes: /p/, /t/, /k/, the depth of the wavelet decomposition J=5, therefore number of bins B=63. Number of elements in $\mathbf{e}_{r,i}$ is equal to B. For each signal class the entropy matrix $\mathbf{E}_r$ is created by aligning the column vectors of the same class:

$$\mathbf{E}_r = [\ \mathbf{e}_{r1}\ \mathbf{e}_{r,2}\ \ \mathbf{e}_{r,3}\ \ldots\ldots\ \mathbf{e}_{r,Mr}\ \} \qquad (5)$$

$\mathbf{E}_r$ is a BxM$_r$ matrix, M$_t$ being the number of sample signals in the training set for a given class.

Singular Value Decomposition (SVD) of the matrix $\mathbf{E}_r$ is denoted as:

$$\mathbf{E}_r = \mathbf{U}\,\boldsymbol{\Sigma}\,\mathbf{V}^T \qquad (6)$$

The B-element singular vectors $\mathbf{u}_i$ make the columns of the BxB orthogonal matrix **U**

$$\mathbf{U} = [\ \mathbf{u}_1\ \mathbf{u}_2\ \mathbf{u}_3\ \ldots\ \mathbf{u}_B\ ] \qquad (7)$$

The BxM$_r$ singular value matrix, $\boldsymbol{\Sigma}$ reveals the rank of $\mathbf{E}_r$ in first M$_r$ diagonal elements. The effective rank of $\mathbf{E}_r$ is equal to the number of non-zero or non-negligible singular values.

These singular vectors identify the dominant entropy patterns for each signal class. If the difference ratio between the largest and second largest singular values for each class is close to one, then we can assume that the representative vector for a given class is the first of the singular vectors.

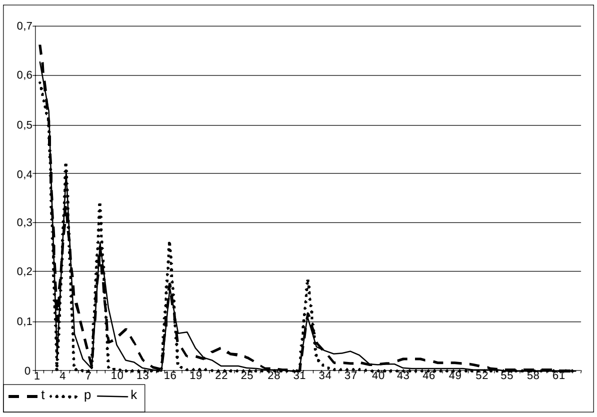

Fig. 3. Components of the 63-element primary singular vectors for three classes of voiceless plosives using Db6 wavelet mother function (entropy)

However for the purpose of classification not only the information about the highest entropy patterns is important, but rather choice of those features that are distinctly different between classes. Interclass difference Δp[b] for each bin b is defined as a sum of entropy differences between classes:

$$\Delta\mathbf{p}[b] = |u_{1,p}[b] - u_{1,t}[b]| + |u_{1,t}[b] - u_{1,k}[b]| + |u_{1,k}[b] - u_{1,p}[b]| \quad (8)$$

The highest values of Δ**p**[b] indicate the bins, that carry the most distinctive features between classes. Fig. 4. represents values of Δ**p**[b] for each bin in the decomposition.

In our case the candidates for the reduced data set can be found using the thresholding method: for the classification these bins b are taken into account, for which Δ**p**[b] has the value higher then above given threshold. This threshold in practice is usually being set heuristically, however in our case we decided to take into account 20 bins.

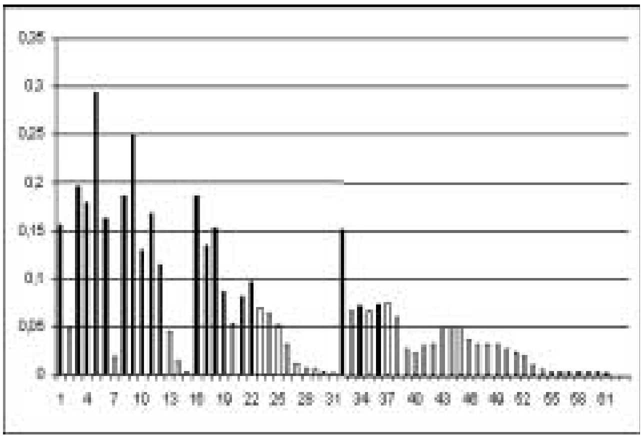

Fig.4. Values of Δ**p**[b] for all WPT bins b

6. Classification results

In the Table 1. we present exemplary classification results for two methods of feature reduction from the Wavelet Packet Transform bins. The classifier was the multilayer perceptron neural network with backpropagation training algorithm (one hidden layer). In training phase 70% of input vectors have been used, the rest being used for testing. Division into training and testing sets has been performed randomly.

Table 1. Average classification rate in training and testing phases for two methods of WPT features reduction (20 features)

Method	Discrimination measure/ mother wavelet	Classification rate in training phase [%]	Classification rate in testing phase [%]
LDB	Euclidean measure Db14	91,25	72,50
	Entropy measure Db14	91,78	78,24
SVD	Entropy Db 6	98,07	81,68
	Entropy Coiflet 3	98,69	82,27
	Entropy Vaidyanathan	99,73	86,72

The SVD method gives much better results than the LDB. Such a result could be predicted from the values of elements of primary singular vectors in comparison with discriminant measures for LDB, no matter what mother wavelet has been used for analysis.

For comparison k-means clustering classification technique has been applied for 10 maximum valued elements of primary singular vectors using entropy (4) and Db14 mother wavelet. The division into three clusters was quite satisfactory (see fig.5). The classification rate was around 75%.

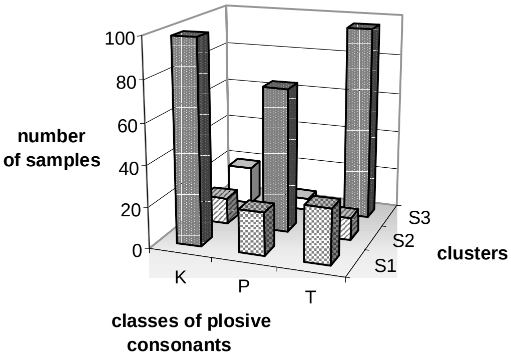

Fig.5. Results of cluster analysis of three classes of plosive consonats represented by primary singular elements

7. Conclusions

Two different strategies have been used for finding the reduced set of features: feature selection and feature projection. Feature set has been composed of Wavelet Packet Transform coefficients of non-stationary speech signals - unvoiced plosive consonants. Feature projection performed by Singular Value Decomposition outperformed feature selection method based on the Local Discriminant Basis. This is in a concordance with conclusions presented in [4] for classification of time-varying transient acoustical signals, i.e. piano attack sounds. On the other hand the classification results of WPT-SVD method are the best for several method exercised on given data set [7]. The overall outcome resulted from the research devoted to the use of Wavelet Packet Transform for context independent classification of plosive consonants /p/, /t/, /k/ is successful and encourages to further experiments with its application to the classification of non stationary signals.

Acknowledgements
This work has been supported by Poznan University of Technology grant 43-342/00-DS

References

[1] H. Krimm, D.H. Brooks, Feature-Based Segmentation of ECG Signals, *Proc. of the IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis*, Paris 1996, pp. 97-100.

[2] K Englehart, Signal Representation for Classification of the Transient Myoelectric Signal, *Ph.D. Thesis*, University of New Brunswick, Fredericton, New Brunswick, 1998.

[3] R.E. Learned, A.S. Willsky, A Wavelet Packet Approach to Transient Signal Classification, *Applied and Computational Harmonic Analysis, Academic Press*, Vol. 2, No. 3, July 1995, pp. 265-278.

[4] Ch.M. Delfs, F.M. Jondral, Classification of Transient Time-Varying Signals Using DFT and Wavelet Packet Based Methods, *Proc. of Int. Conf. on Acoustics, Speech and Signal Processing, 1998*, pp.1569-1572.

[5] R. Coifman, M.Wickerhauser, *Entropy - based algorithms for best-basis selection*, IEEE Trans. Information Theory, vol.38,No2, March 1992.

[6] E.Łukasik, S.Grocholewski, Comparison of Some Time-Frequency Analysis Methods for Classification of Plosives, *Signal Processing IX, Theories and Applications*, Typorama Publications, Greece, 1998, pp. 709-712.

[7] E. Łukasik, *Wavelet Packets Based Features Selection for Voiceless Plosives Classification*, accepted for ICASSP 2000.

[8] S. Grocholewski, *CORPORA – Speech Database for Polish Difones*, *Proc. EUROSPEECH'97*, Rhodes, Greece, 1997, pp. 1735-1738.

[9] N.Saito, *Local Feature Extraction and Its Applications Using a Library of Bases*, PhD thesis, Yale University, December 1994.

[10] E.Deprettre (ed.), *SVD and Signal Processing*, Elsevier Science Publ., Amsterdam 1989.

[11] M.Dash and H.Liu, Feature Selection for Classification, *Intelligent Data Analysis, Vol. 1, no. 3, http:llwww.elsevier.com/locate/ida*) ©1997 Elsevier Science Inc.

[12] H.Liu, H.Motoda (ed.), Feature Extraction Construction and Selection, a Data Mining Perspective, *Kluwer Academic Publishers, 1998.*