A Method of Wavelet Selection in Phone Recognition

Hugo L. Rufiner,
Lab. de Cibernética, Facultad de Ingeniería
Universidad Nacional de Entre Ríos, Argentina , lrufiner@arcride.edu.ar
John Goddard C.,
Depto. Ingeniería Eléctrica, UAM-I, México, jgc@xanum.uam.mx

Abstract- In this paper a method is proposed for choosing between different wavelets, and their corresponding parameters, and it is applied to TIMIT phoneme classification. The method involves the use of a Kohonen network to extract prototypes for each class. Distance measures between them are used as criteria for choosing the best wavelet. In the case of phoneme classification, a time delay neural network is used as the classifier. A high correlation is found between results of the proposed method and the classification accuracy obtained by the network.

I. INTRODUCTION

Wavelets have come to play an important role in problems involving signal-processing [1]. Their advantages over the traditional Fourier Transform have been shown in a variety of studies given their natural ability to analyze transitory components. Because there are an increasingly large number of wavelet families available, an important question is that of which one to choose for a particular application. A wavelet which works well for one application may perform poorly for another, even though the same signals are involved.

One way is to perform an exhaustive experimentation with different families of wavelets and their possible parameters. This is usually not very practical.

Another is to choose ones own favourite wavelet and stick with it, although there is no guarantee that it is the best for that particular application. Among the popular choices, for example, are those of Meyer, Daubechies, and Splines [2].

Finally the introduction of Wavelet Packets [3] and algorithms such as Matching Pursuit [4] try either to choose a "better base" for the chosen wavelet, or to avoid the decision of which wavelet to choose and to consider a dictionary of different wavelets and let the algorithm pick the best functions available to make the required approximation.

In this paper a method is proposed, as a first intent at attacking the problem, for choosing a wavelet in the problem of signal classification. It is illustrated in the case of phoneme classification. This area of application is one which requires a decision on the choice of wavelet from the outset as the time involved in the classification process is computationally very expensive.

II. WAVELET TRANSFORM

The type of analysis applied to the processing of voice signals has usually been that of the Short Time Fourier Transform (STFT). One could say that it has dominated the field of signal analysis [5]. Recently the Wavelet Transform (WT) has been developed wich conducts the analysis of non stationary signals in more "efficient" way.

One of the limitations of the STFT is that it analyzes signals with transitory components using fixed resolution. The WT provides an alternative to this technique. The main difference is that the WT uses small windows at high frequencies and large windows at low frequencies. This behavior is similar to the analysis that the human hearing system carries out.

In case of the digital signals the Discret Wavelet Transform (DWT) has several attractive characteristics that have contributed to its recent increase in popularity in mathematics and signal processing:

1. Its hierarchical decomposition allows the characterization of the signal at different scales (multiresolution analysis).
2. It gives in essence a decomposition of the signal in sub-bands, which is related to a wide variety of multifrequency decomposition techniques.
3. A fast algorithm exists in order to calculate it, of simple digital implementation.

All of these characteristics, plus a series of recent successful applications [6], [7], [8], [9], allow it to be considered as an interesting alternative to the classical methods of voice analysis.

III. WAVELETS FAMILIES

Many different wavelets exist that form ortogonal bases, and a problem exists in terms of choosing an appropriate wavelet for a particular application. In this respect wavelets possess different characteristics that could guide the selection: Compact Support, Symmetry, Regularity, and Time-frequency Localization. Recently some families of wavelets have appeared based on models of hearing [10],

[11]. These require further investigation in order to understand their possible relevance.

In this paper the wavelets of Meyer, Daubechies, Haar and Splines [12] are used. The following table sumarizes their principal characteristics:

TABLE I
WAVELETS CHARACTERISTICS

Family	Compact Support	Simetry	Regularity	Freq. Localiz.	Coments
Haar	Y	Y	N	bad	the simplest
Meyer	N	Y	Y	good	wide use
Daubechies	Y	N	variable	variable	smooth
Splines	Y	Y	variable	variable	biorthogonal

IV. The Data

For the phoneme classification, the speech data was taken from the TIMIT database [13], which was especially designed for the development of automatic speech recognition systems. A subset of 5 phonemes of different speakers from one region of TIMIT was selected in order to maintain a high variability amongst the data whilst ensuring that the experimentation time was not excessive. The average signals for the data varied from 2 to 11 frames, where each frame had a duration of 8 ms.

V. The Classifier

The classifier used for the recognition task was a time delay neural network (TDNN) [14], [15] with 128 input units, representing the processed frame sizes, another 128 for the delayed input, 150 units in the hidden layer, and 5 output nodes corresponding to the 5 classes. This architecture was chosen after experimenting with the same data, but processed using the STFT (128 coefficients).

A training file, representing almost 80% of the data set, was generated with the frames of the unprocessed signals and their corresponding labels, the rest was used for testing purposes with the TDNN. The training file was divided into the five categories of phonemes chosen for this study. The training was conducted at least twice for each family, and the best value chosen.

VI. Wavelet Selection Method

A Kohonen network [16] was trained using this data with three prototypes for each class. The decision to take three prototypes was made in order to reflect the range of small, medium, and high energy frames for each of the classes. This was in fact born out by the results. The Discrete Wavelet Transform [2] was applied to each of the fifteen prototypes for each wavelet family chosen, giving vectors with 128 coefficients. The vectors for each class were normalized and combined, giving 5 different vectors. The sum of the euclidean distances between them were calculated and the minimum distance also obtained. This is schematically shown in figure 1.

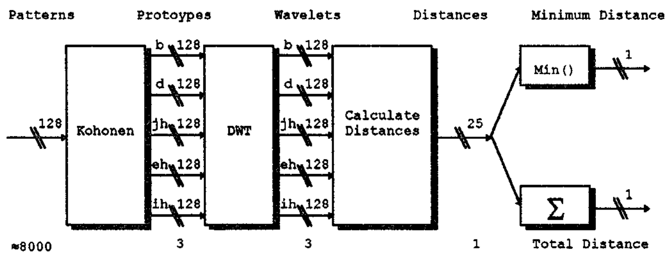

Figure 1: Process of Selection

These distances provided the criteria for choosing the parameters of a particular family of wavelets and the wavelets themselves. The idea was to first use the total distance as a criterion for class separability, selecting the largest. In the case of "closeness" between the total distance values, the minimum distance served as a further criterion for tie-breaking.

The different wavelets, with the chosen parameters, were then trained with the TDNN and classification results were obtained for the phonemes.

VII. Experiments

First the method, explained in the previous section, was applied to the selection of the parameters required by the Daubechies and Spline families. This method was applied to each family separately, varying the corresponding parameters (number of moments, regularity, etc.). An outline of the process can be appreciated in Figure 2.

Once the optimal values of the parameters were obtained, the process was repeated to determine the best family of wavelets. This process is schematized in Figure 3.

VIII. Results

The results obtained for the parameters of the Daubechies family are shown in Figures 4 and 5. Here the selected value was 8 moments (m=16).

The distances obtained for the different families are shown in Table II.

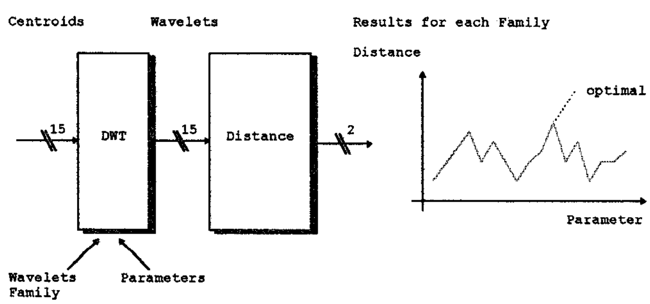

Figure 2: Optimal Parameter Selection for each Family

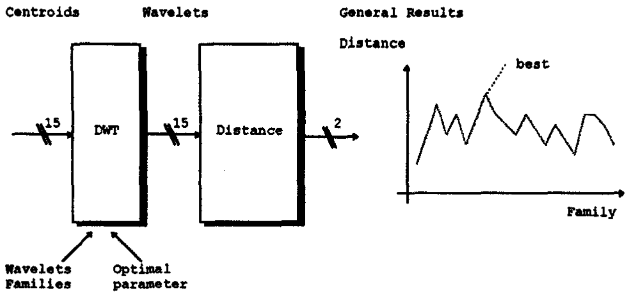

Figure 3: Best Family Selection

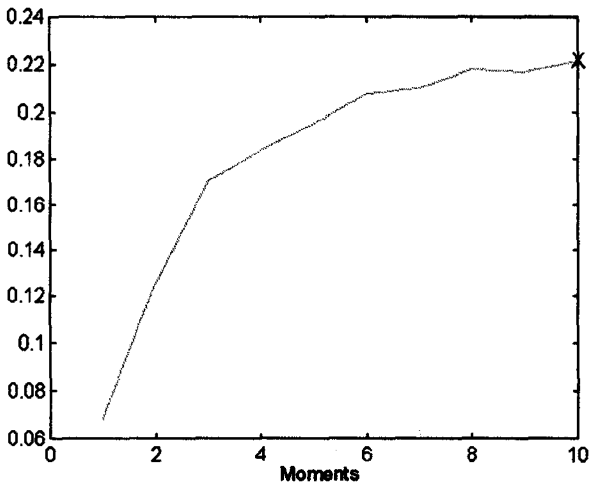

Figure 4: Comparison of Daubechies Wavelets (Total Distance)

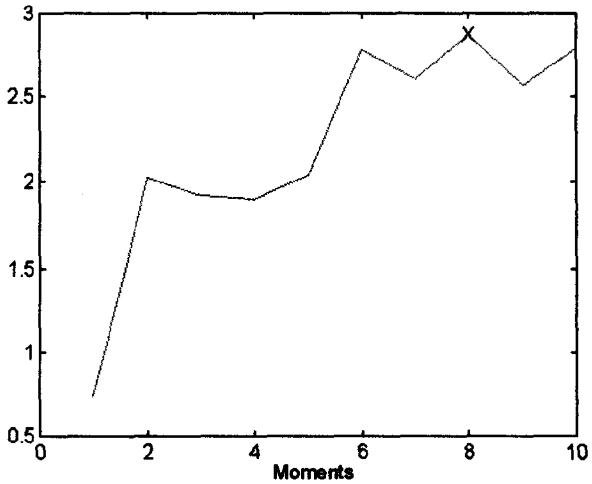

Figure 5: Comparison of Daubechies Wavelets (Minimun Distance)

Finally, the classification results corresponding to the training and test sets for the TDNN with the fixed architecture of 128+128/150/5, are given in Table III.

TABLE II

CENTROIDS DISCRIMINATION WITH DIFERENT WAVELETS

Family	Total Distance	Minimun Distance
Splines (9,37)	0.5152	0.0150
Meyer	0.3946	0.0058
Daubechies (16)	0.2179	0.0028
Haar	0.0685	0.0007

TABLE III

NEURAL NETS RECOGNITION RATE

Description	Estructure	Train (%)	Test (%)	Epochs
Splines (9,37)	128+128/150/5	70.43	70.92	75
Meyer	128+128/150/5	65.46	67.58	50
Daubechies (16)	128+128/150/5	63.76	63.11	163
Haar	128+128/150/5	53.15	50.00	178

IX. CONCLUSIONS

In this paper a natural method for wavelet selection has been proposed. The results obtained for the centroid distances and the training and test classification results, using a TDNN, are presented. A high correlation was observed between them. It is interesting to observe that the time taken to apply the method to each wavelet and its parameters represented approximately 10% of the time taken to train the TDNN once. These results motivate further study in order to improve the method and apply it to other signal classification problems.

REFERENCES

[1] Olivier Rioul, Martin Vetterli *"Wavelets and Signal Processing"*, IEEE Magazine on Signal Processing, pp. 14-38, October 1991.

[2] I. Daubechies, *"Ten Lectures on Wavelets"*, Rutgers University and AT&T Bell Laboratories, Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania, 1992.

[3] M.A. Cody, *"The Wavelet Packet Transform : Extending the Wavelet Transform"*, Dr. Dobb's Journal, April 1994.

[4] S. Mallat, Z. Zhang, *"Matching Pursuit With Time-Frequency Dictionaries"*, IEEE Trans. in Signal Proc., December 1993.

[5] J.B.J. Fourier, "Thèorie Analytique de la Chaleur", Oeuvres de Fourier, Tome Premier, G. Darboux Eds., Paris, Gathlers-V. Vilars, 1888.

[6] Richard F. Favero and Robin W. King "Wavelet Parameterisations for Speech Recognition", ICSPAT93.

[7] Richard F. Favero and Robin W. King "Wavelet Parameterisations for Speech Recognition: Variations of scale and translation parameters", ISSIPNN94.

[8] E. Wesfreid, M.V. Wickerhauser, "Adapted Local Trigonometric Transforms and Speech Processing", Université Paris IX-Dauphine, 1992.

[9] M.D. Wickerhauser, "Acoustic Signal Compression with Wave Packets", Yale University, 1991.

[10] M.N. Souza, L.P. Caloba, "An Analytical Auditory Wavelet", Annals of the 13th Brazilian Telecommunications Symposium, Águas de Sindorá, 1995.

[11] J. Benedetto, A. Teolis, "Wavelets Auditory Model and Data Compresion", Applied and Computational Harmonic Analysis, Vol.1 Nº1, December 1993.

[12] I. Daubechies, "Ten Lectures on Wavelets", Rutgers University and AT&T Bell Laboratories, Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania, 1992.

[13] Garofolo, Lamel, Fisher, Fiscus, Pallett, Dahlgren, *DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus Documentation*, National Institute of Standards and Technology, February 1993.

[14] A. Waibel, T. Hanazawa, G. Hinton, K. Shikano, K. Lang; *"Phoneme Recognition Using Time-Delay Neural Networks"*. IEEE Trans. ASSP Vol. 37, No 3 (1989).

[15] W.H. Wilson, *"A comparison of Architectural Alternatives for Recurrent Networks"*, Proceedings of the Fourth Australian Conference on Neural Networks, pp.189-192, 1993.

[16] T. Kohonen, *"The 'Neural' Phonetic Typewriter"*, Computer, March 1988.