Selection of the Most Indicative Wavelets for the Multiresolution of the Vowels

I.Traljic, B.Perunicic, M.Ahic-Dokic
Faculty of Electrical Engineering University of Sarajevo

ABSTRACT

The paper describes one approach of the selection of the most indicative wavelets for each of the vowels in the author's native language. Analysis is performed on the correct and incorrect vowels. On each of the sample multiresolution decomposition is applied. For each of the detail and approximation the most indicative wavelet is selected using value of the variance as the criteria. Some interesting results are obtained and biorthogonal wavelets have been select as the most appropriate for the multiresolution of the vowels. Using this criterion, any further analysis of the samples can be done using only coefficients of the discrete wavelet transformation on the level of approximation or any level of the detail, with enough guarantees that they are most appropriate for each vowel.

Keywords: Multiresolution Decomposition, Speech Therapy, Voiced Sounds, Biorthogonal Wavelets

1. INTRODUCTION

Because speech is the most natural form of human communication, speech processing has been one of the most exciting areas of signal processing [1]. The phonemes of speech encompass a wide variety of characteristics, in time and frequency domain. The two major types of excitation in speech are voiced and unvoiced. Voiced sounds are produced when air flows between the vocal cords and causes them to vibrate. The sound produced by this process is periodic in nature [2]. Voiced sounds in author's native language are A, E, I, O and U. In structure of Bosnian language vowels are appearing with 41.89%. The most frequent is appearance of vowel 'A', than 'I", 'E', 'O' and 'U'. That explains why is the correct pronunciation of the vowels most important during the speech therapy. Voiced sounds, e.g. vowels are typically lower in frequency for a longer time duration. High-frequency localization gives poor time resolution, and high time resolution gives poor frequency localization. Because of that, it is important to analyze signal from a multiresolution perspective. During the process of the multiresolution, different type of wavelets can be used. Wavelet transformations have beautiful and deep mathematical properties, making them a well-adapted tool for a wide range of different type of data [3]. Among all wavelets, biorthogonal wavelets are the one which are used in the most of practical application, from FBI fingerprints archive to JPEG2000 standard. For biorthoganal wavelet, in order to gain flexibility in the construction of wavelet basis, nonorthogonal wavelet bases are allowed. Since the symmetry of the filter coefficients results in liner phase, it is more often used in applications. Also, it is possible to have different scaling functions and wavelet for the analysis and synthesis, and they can be nonsmooth and smooth function respectively. It will be shown that biorthogonal wavelet, particularly bior3.1 wavelet, is the most indicative for analysis of the speech signals analyzed in this paper. Analysis is done on 40 different samples from children in the age from 12 to 14, without capability to correctly pronounce vowels. Also 40 samples of correctly pronounce vowels are taken. The next section describes process of the samples analysis. Samples are taken with frequency 44.1 kHz on the personal computer during speech therapy. During this process it was necessary to determine pitch period and prepare signal for the multiresolution analysis only using that part of the signal. The third section describes process of the selection of the most indicative wavelet. The most indicative wavelet for each level of the detail and for approximation is selected, and the criterion was the biggest value of the variance for each level. In the conclusion some further directions and possible use of this results are presented.

2. ANALYSIS OF THE SAMPLES

The samples of the speech signals for the children with incapability to correctly pronounce vowels have been taken during the speech therapy. Usually during this therapy, patient tries to repeat particular vowel after speech therapist's instructions and supervision. One vowel is pronounced constantly, without any accent or emotions, and there is no any other consonant around it. This signal can be considered as the stationary, because there is no change in structure that crate that signal, e.g. there is no change in the vocal tract. Signal has waveform as it is shown on the figure 1. It is obvious that signal is almost periodic, and that is typical for all vowels. For the further analysis the signal has been considered only on the pitch period. Pitch period corresponds to the frequency F0.

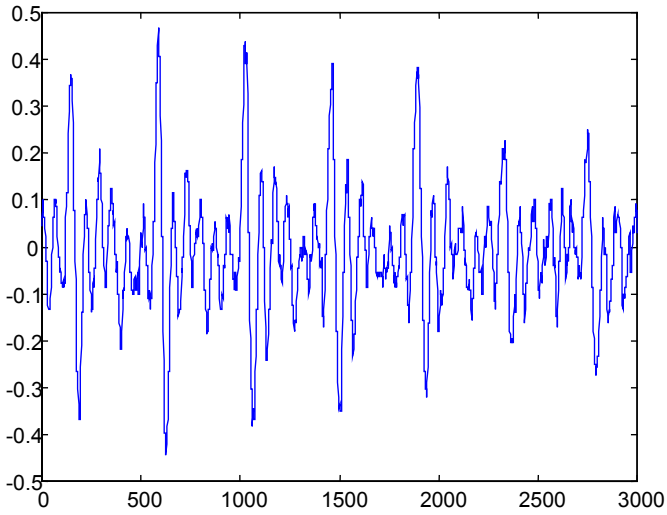

Fig. 1. Vowel A on the period of 68 ms

To discover that period, autocorrelation function has been used, and it is shown on the figure 2. for one of the vowels using only first 2000 samples. Mean value of the distances between maximums of the function is calculated, and that presents pitch period.

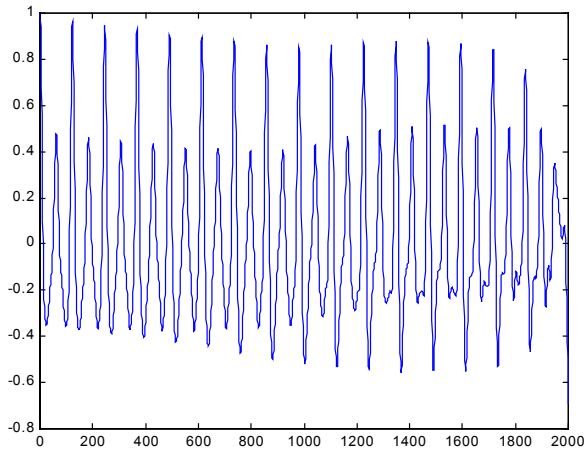

Fig. 2. Autocorrelation function for the vowel shown on fig. 1

Once signal has been taken on the pitch period it is non-stationary signal, and wavelet analysis is becoming very powerful tool for the further analysis. Two important transformations of this signal have been done:

-Each signal has been rescaled to the same period, and it is 600 samples for each signal, which is still equal to 2,33 ms. This has to be done, because it will be irrelevant value of the laryngeal tone F0, which can be vary from person to person. For the woman this frequency is 18% higher, and for children it is 25%.
-Each signal is normalized that energy is equal 1. In that case higher value of the coefficient in the multiresolution is not result of the loudness of the pronounced vowel.

3. SELECTION OF THE WAVELETS

Multiresolution decomposition is performed using following 37 wavelets [4]:
Daubechies wavelets: db1, db2, …db10
Symlet wavelets: sym2, sym3, … sym8
Coiflet wavelets: coif1, coif2, …coif5
Biorthogonal wavelets: bior1.1, bior1.3, bior1.5, bior2.2, bior2.4, bior2.6, bir2.8, bior3.1, bior3.3, bior3.5, bior3.7, bior3.9, bior4.4, bior5.5, bior6.8

Multiresolution has been done till fifth level. That means, result of the decomposition is five details and one approximation. Before the process of selection is done, results of the decomposition are prepared in the following way:
For each wavelet six matrices are prepared, for each result of the decomposition. Columns are samples, and rows are coefficients. As example for multiresolution using Daubechies wavelet db5, from the signal with 600 samples, the results are matrices with following dimensions:

Approximation A5 27rowsX40columns
Detail D5 27rowsX40columns
Detail D4 45rowsX40columns
Detail D3 82rowsX40columns
Detail D2 156rowsX40columns
Detail D1 304rowsX40columns

For each row in the matrices, variance has been calculated, and means value of that vector column has been found. Mean values of the variances have been compared for approximation and for each detail separately. As example, for approximation obtained result has been shown in table 1 for the first five most indicative wavelets. A criterion for selection e.g. the most indicative wavelet is one that has the biggest value of the variance. Higher variance mean that coefficients obtained using that wavelet are most indicative for that signal and should be considered as the optimal or "the best" coefficient for the analysis of the signal.

Applying multiresolution decomposition and using all this wavelets mean value for the variance for each wavelet has been calculated.

Wavelet	Variance's mean value
bior3.1	0.04938047
bior3.3	0.04359715
bior2.2	0.03914922
bior3.5	0.03875904
bior3.7	0.03497560

Table 1. First five wavelets for the multiresolution decomposition of the incorrect vowel 'A' for the fifth detail level

The same process is performed for the 40 samples obtained from the children that don't have problems with pronunciation of the vowels.

The final result for vowels A, E and I is presented in the table 2. It is interesting that even for correct and incorrect samples, for each vowel, on the approximation level the best wavelet is bior3.1 wavelet. Biorthogonal wavelets are generally the most indicative for any level of the multiresolution. Also, it can be noticed that Daubechies wavelet db1, e.g. Haar wavelets are the most indicative for the first level of the detail.

cA5

	A		E		I	
	Incorr.	Corr.	Incorr.	Corr.	Incorr.	Corr.
1	bior3.1	bior3.1	bior3.1	bior3.1	bior3.1	bior3.1
2	bior3.3	bior3.3	bior3.3	bior3.3	bior3.3	bior3.3
3	bior2.2	bior3.5	bior3.5	bior3.5	db1	bior3.5
4	bior3.5	bior2.2	bior2.2	bior3.7	bior1.1	db1
5	bior3.7	bior3.7	db1	bior2.2	bior3.5	bior1.1

cD5

	A		E		I	
	Incorr.	Corr.	Incorr.	Corr.	Incorr.	Corr.
1	bior1.3	db1	bior3.1	bior1.3	bior3.1	bior3.1
2	db1	bior1.1	bior1.3	bior3.1	bior1.3	db2
3	bior1.1	bior1.3	db1	db1	db1	sym2
4	bior1.5	bior1.5	bior1.1	bior1.1	bior1.1	coif1
5	bior3.1	bior3.1	bior1.5	bior1.5	bior1.5	db3

cD4

	A		E		I	
	Incorr.	Corr.	Incorr.	Corr.	Incorr.	Corr.
1	db1	bior3.1	bior1.3	bior1.3	db1	bior3.1
2	bior1.1	db1	bior1.5	db1	bior1.1	bior1.3
3	bior1.3	bior1.1	db1	bior1.1	bior1.3	db1
4	bior1.5	bior1.3	bior1.1	bior1.5	bior3.1	bior1.1
5	bior3.1	bior1.5	bior3.1	bior3.1	bior1.5	bior1.5

cD3

	A		E		I	
	Incorr.	Corr.	Incorr.	Corr.	Incorr.	Corr.
1	bior1.5	bior3.5	bior1.5	bior3.7	bior1.5	bior3.5
2	bior1.3	bior1.5	bior1.3	bior3.3	bior1.3	bior1.5
3	db1	bior3.9	db1	bior1.5	db1	bior3.9
4	bior1.1	bior3.7	bior1.1	bior3.5	bior1.1	bior3.7
5	bior3.5	bior3.3	bior3.3	bior3.9	bior3.3	bior3.3

cD2

	A		E		I	
	Incorr.	Corr.	Incorr.	Corr.	Incorr.	Corr.
1	bior1.3	bior1.3	bior1.3	bior1.3	bior1.3	bior1.3
2	bior1.5	bior1.5	bior1.5	bior1.5	bior1.5	bior1.5
3	db1	db1	db1	db1	db1	db1
4	bior1.1	bior1.1	bior1.1	bior1.1	bior1.1	bior1.1
5	db2	db2	bior3.9	coif1	db2	db2

cD1

	A		E		I	
	Incorr.	Corr.	Incorr.	Corr.	Incorr.	Corr.
1	db1	db1	db1	db1	db1	db1
2	bior1.1	bior1.1	bior1.1	bior1.1	bior1.1	bior1.1
3	bior1.3	bior1.3	bior1.3	bior1.3	bior1.3	bior1.3
4	bior1.5	bior1.5	bior1.5	bior1.5	bior1.5	bior1.5
5	db4	coif1	coif1	db3	db4	db4

Table 2. The most indicative wavelets for the first three vowels

4. CONCLUSION

It is shown that using mean value of the variance the most indicative wavelets for each level of the decomposition can be obtained. The biorthogonal wavelets became the most indicative, and specially bior3.1 wavelets for the level of the approximation. Result of this work can be used for algorithms of the compression of the voice signal. Also, final result of the multiresolution, e.g. the coefficients of the most indicative wavelets can be used like well prepared input for any classification, such as input for neural network to classify incorrect and correct vowels.

5. REFERENCES

[1] The Speech Processing Technical Committee, B. H. Juang Ed., "The past, present, and future of speech processing", IEEE Signal Processing Magazine, 24-48, May 1998.

[2] Louis R. Litwin, Jr., "Speech coding with wavelets", IEEE Potentials, 38-41, April/May 1998.

[3] I. Daubechies, "Recent results in wavelet applications", Journal of Electronic Imaging, 7 (no. 4), pp. 719-724, 1998.

[4] Wavelet Toolbox User's Guide, Mathworks Inc., 1997