

Hybrid Method Using Wavelets and Predictive Method for Compression of Speech Signal

Karima Siham Aoubid, and Mohamed Boulemden

Abstract—The development of the signal compression algorithms is having compressive progress. These algorithms are continuously improved by new tools and aim to reduce, on average, the number of bits necessary to the signal representation by means of minimizing the reconstruction error. The following article proposes the compression of Arabic speech signal by a hybrid method combining the wavelet transform and the linear prediction. The adopted approach rests, on one hand, on the original signal decomposition by ways of analysis filters, which is followed by the compression stage, and on the other hand, on the application of the order 5, as well as, the compression signal coefficients. The aim of this approach is the estimation of the predicted error, which will be coded and transmitted. The decoding operation is then used to reconstitute the original signal. Thus, the adequate choice of the bench of filters is useful to the transform in necessary to increase the compression rate and induce an imperceptible distortion from an auditive point of view.

Keywords—Compression, linear prediction analysis, multiresolution analysis, speech signal.

I. INTRODUCTION

THE subject of data compression is the object of a particular development, which led a variety of algorithms. The analysis by ways of wavelets is one of the new approaches used in the signal-processing field; the method allows the decomposition of a given signal on concentrated functions both in time and frequency. In coding by linear prediction, the value of each signal coefficient is predicted based on the preceding coded coefficients. The difference between the predicted value and the real one, e.g. the given error, is quantified, then coded and transmitted. In this article, we present an hybrid method for the Arabic speech signal compression; this method consists in combining the coding by wavelets [1], [2], [3] and the linear prediction [4] in order to improve the compression rate which an acceptable auditive quality of the reconstructed signal.

The article is organized in the following manner:

In the second and third sections, we describe the elementary notions concerning the analysis by wavelets and the linear prediction method as well as an introduction of the adopted approach in the case of the signal compression. The fourth section will be devoted to the implantation of the compression algorithms of speech signal and the analysis of the obtained

results is discussed in the fifth section. Finally, in section six, a conclusion is made concerning the evaluation of the performances of the adopted approach.

II. ANALYSIS BY THE WAVELET TRANSFORM

The wavelet theory has known a great development since S. Mallat's proposition of a fast algorithm to compute the wavelet coefficients. In this algorithm, S. Mallat ties the orthogonal wavelets to the bench of filters used in signal processing. The fundamental idea is building the signal by successive refinements i.e. by means of adding details to a approximation following an iteration process [2], [5]. From a mathematical point of view, the multiresolution analysis is based on writing every function f from $L^2(\mathbb{R})$ as a discrete series of $f(x)$ function where each one is an approximation of the desired function for a given resolution. This function is defined by an increasing series $\{V_j\}$, $j \in \mathbb{Z}$ of closed vectorial under spaces of $L^2(\mathbb{R})$. Every V_j , is related to one 2^j resolution and to the approximation of the given signal.

This series is obtained by projection over the corresponding spaces, such that :

- if P_j are the orthogonal projection over the spaces V_j , then, for every function f in $L^2(\mathbb{R})$:

$P_j : f \in L^2(\mathbb{R}) \rightarrow P_j(f) \in V_j$ and $P_j(f)$ is an approximation of f for a 2^j resolution.

- Since $V_j \subset V_{j+1}$, then approximation $P_{j+1}(f)$ is "a better" approximation than $P_j(f)$.

- From the initial value V_0 , we can deduce, the base V_j by building the family of expanded and translated functions:

$$\phi_{j,k}(x) = 2^{-j/2} \phi(2^j x - k) \quad (1)$$

obtained from the scale function $\phi(x)$.

- The information difference, $P_{j+1}(f) - P_j(f)$, which exists between two consecutive approximations represents the additional detail necessary to pass from the approximation $2^j(P_j(f))$ to a finer approximation with scale $2^{j+1}(P_{j+1}(f))$. According to the theory of Daubechies-Mallat, a set of functions $\{\psi_{k,j}\}$ is then introduced; these functions from the wavelet $\psi(x)$ and form an orthogonal base of the space V_j .

Finally, the decomposition of the signal to be analysed S , presumably approximated $j=0$, is carried out according to the following steps:

- The approximation of the signal S^{-1} to the $j=-1$ resolution is obtained by the convolution of the S^0 signal by the analysis filter $h(x)$ and by decimation of order 2, such as:

$$\tilde{h}(x) = h(-x) \quad (2)$$

where h is a low pass filter whose impulse response:

$$h(k) = \frac{1}{\sqrt{2}} \langle \phi_0^{-1}, \phi_k^0 \rangle \quad (3)$$

- The D_j details are obtained starting from S^{j+1} by convolution by the filter $g(x)$ and by retaining only a sample over 2 ():

$$\tilde{g}(x) = g(-x) \quad (4)$$

Where g is a high pass filter, of impulse response:

$$g(k) = \frac{1}{\sqrt{2}} \langle \phi_0^{-1}, \phi_k^0 \rangle \quad (5)$$

Thus, the signal to be analysed will be represented by the approximated signal S^j of low resolution and the details signals with the $2j$ resolutions for $J < j < -1$. Then the schemes of discrete wavelet transform decomposition and reconstruction of a signal x is shown in Fig. 1 and Fig. 2 respectively.

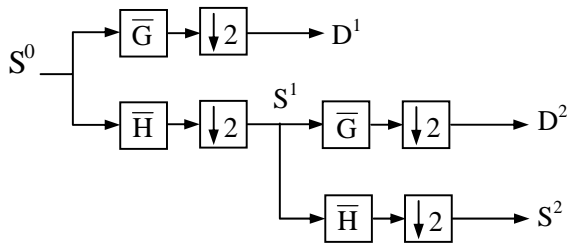


Fig. 1 Banc of analysis at two levels

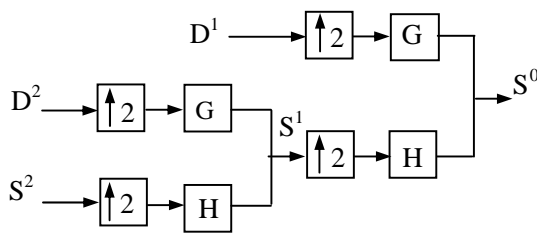


Fig. 2 Reconstruction scheme

The adopted strategy in compression consists in coding the last signal with the coefficients of the details signals, with various resolutions, which exceed a certain fixed threshold [6], [7].

III. ANALYSIS BY THE LINEAR PREDICTION

The linear prediction is another manner to reduce the redundancies between the close coefficients; it is based on the assumption that each sample of the original signal $x(n)$ can be approximated by linear combination of samples which precede it [5], such as:

$$\hat{x}(n) = - \sum_{i=1}^p a(i) x(n-i) \quad (6)$$

$$\hat{x}(n) = - a(1) x(n-1) - a(2) x(n-2) - \dots - a(p) x(n-p) \quad (7)$$

where $a(i)$ represents the coefficients of the prediction of order p, $\hat{x}(n)$ is the predicted value.

The prediction error of order p is given by:

$$e(n) = x(n) - \hat{x}(n) \quad (8)$$

$$e(n) = \sum_{i=0}^p a(i) x(n-i); \quad a(0)=1 \quad (9)$$

The remaining prediction energy is:

$$E = \sum_n e^2(n) \quad (10)$$

Also the calculation of the gain is often used to measure the performance of the predictive modelling:

$$G_p = \frac{\sigma_x^2}{\sigma_e^2} \quad (11)$$

where σ_x^2 is the variance of the predictor input signal and σ_e^2 is the variance of the prediction error.

The samples of residue of prediction $e(n)$ have a level generally lower; they require less bits to be represented. The compressed signal is then rebuilt starting from the parameters of the predictor and the samples of the error signal. The optimal predictor is that which minimizes the power of the prediction error, $E\{e(n)^2\}$, which leads to an adequate choice of the coefficients $a(i)$, by the resolution of the equations of Yule-Walker. These latter can, particular, be solved by the Levinson-Durbin algorithm and that of Schur. Let us recall that the coefficients $a(i)$ are those of the denominator of the filter all poles whose transmittance is the spectral envelope of the signal.

IV. IMPLEMENTATION OF THE COMPRESSION OF SPEECH SIGNAL BY THE HYBRID METHOD

The type of signals considered as the object to be treated in our study is the Arabic speech data base ("sin"), illustrated in Fig. 3, presented on 3842 samples at a rate of 16 KHz. The S. Mallat algorithm is applied to this configuration in order to break it up into two signals approximated S_j with D_j details obtained respectively by low pass filtering (the h filter) and high pass filtering (the g filter) according to the values of the analysis filters provided by the works of I. Daubechies [2]. A factor 2 decimation is then produced on the signals S_j and D_j in order to preserve the same information quantity. The compression stage rests on the omission of the coefficients of the signal of detail and the conservation of the obtained approximated signal, Fig. 4, which will be treated thereafter by the linear prediction algorithm.

This algorithm will allow the calculation of the parameters of the order $p=5$ predictor and of the samples of the error signal $e(n)$, Fig. 5. The parameters and the samples of error must be quantified and coded perfectly for their transmission or their storage.

The reconstruction of the speech signal is realized by the follow stages:

- i) the calculate, of the data, from the parameters and the samples of the error.
- ii) and then the application of the inverse discrete wavelet transform.

The reconstructed signal is shown in Fig. 6.

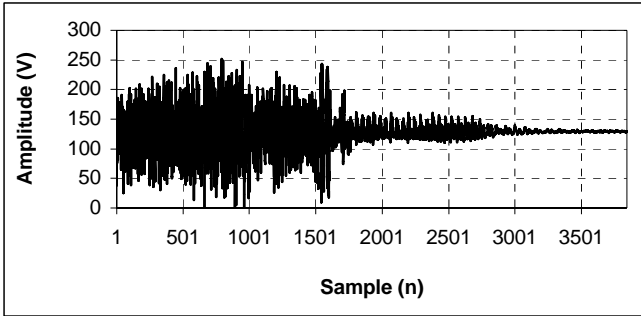


Fig.3 Arabic letter 'sin'

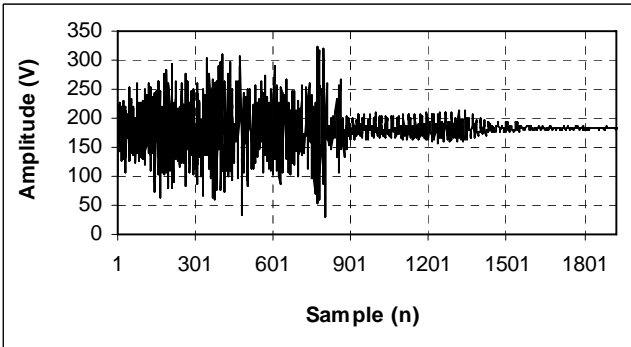


Fig. 4 Coefficients of approximation at level 1

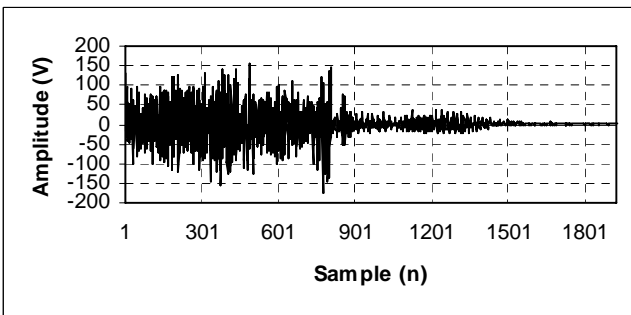


Fig. 5 Residue of prediction

The performance of the compression process is quantify by the measure of several parameters [8]:

- Compression rate :

$$CR = \frac{\text{number of bits to represent original signal}}{\text{number of bits to represent compressed signal}} \quad (12)$$

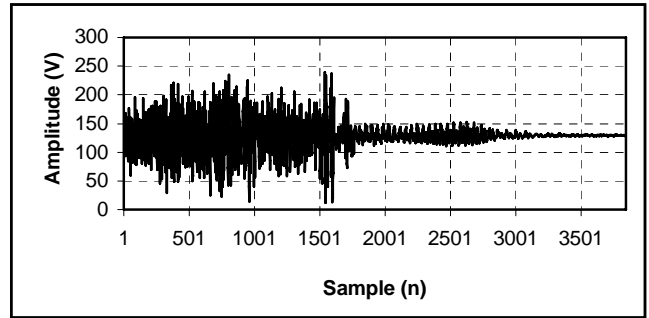


Fig. 6 Reconstructed signal

- Mean square error:

$$MSE = \frac{1}{N} \sum_{i=0}^{N-1} [S(i) - S'(i)]^2 \quad (13)$$

where $S(i)$ is the original speech data and $S'(i)$ is the compressed signal.

- Root mean square error :

$$RMSE = \left[\frac{1}{N} \sum_{i=0}^{N-1} [S(i) - S'(i)]^2 \right]^{1/2} \quad (14)$$

- Peak signal to noise ratio :

$$PSNR = 10 \log \left(\frac{(255)}{RMSE} \right) \quad (15)$$

V. RESULTS AND DISCUSSION

For computing the hybrid method the algorithm are developed on the one hand by choosing the analysing function such as Daubechies2, Daubechies4 and Daubechies8 from levels 1 and 2; on the other hand by varying the prediction order.

The compressed Arabic speech signal were obtained using Daubechies2 and decomposed up to level 2. It is reconstructed only from the approximated coefficients and for this reason the decomposition at a level superior leads at an unacceptable distortion of the compressed signal.

The results of this analysis given in Table I illustrate that as compression ratio and the MSE increase where as the PSNR will decreases and the compression rate $CR = 2$ were obtained with $MSE = 275.89$ distributes on 3842 samples.

For calculating the compression rate in linear analysis the parameters and the samples of error must be coded perfectly.

Fig. 7 shows the compression rate versus prediction order. It is observed that compression rate decreases lightly when the order of prediction increases. For that predictor order of 5 provided the sufficient result in this study.

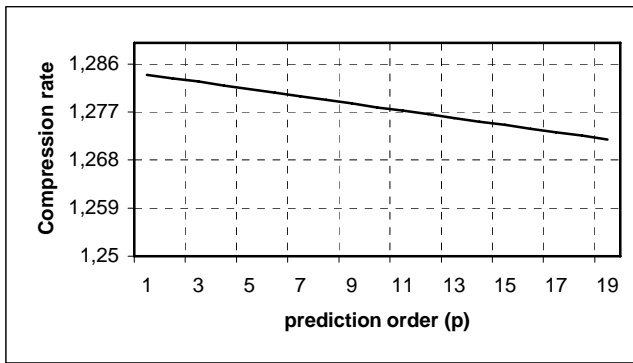


Fig. 7 Compression rate for various prediction order

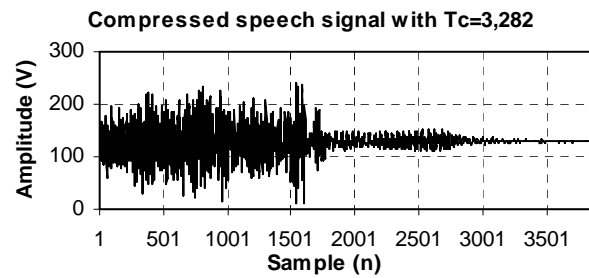
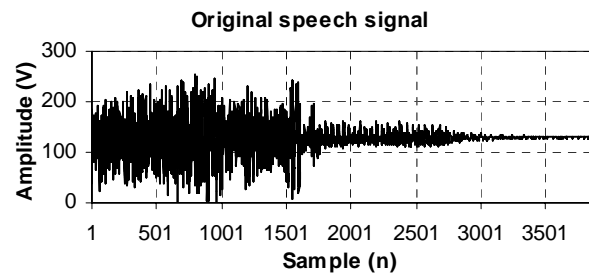


Fig. 8 Arabic speech signal Synthesis

TABLE I
 ANALYSIS USING HYBRID SCHEME FOR COMPRESSED ARABIC SPEECH SIGNAL

CR	DWT	
	Level 1	Level 2
	2	4
RMSE	16.61	21.80
PSNR	27.30	24.5913
Linear prediction		
CR	P=5	P=10
	1.2820	1.2786
E	$2.4806 \cdot 10^6$	$2.3153 \cdot 10^6$

The compression ratio obtained combining wavelet and linear prediction attains the value 3,282 with the good speech signal quality like illustrates in Fig. 8, also some of the information is lost. For speech signals this loss is acceptable since it is interested only in recognizing the compressed signal.

VI. CONCLUSION

In this paper the hybrid scheme combining discrete wavelet transform and linear prediction method was proposed for the compression of the Arabic speech signal. The associated both methods are proved its efficiency by an increase of the compression rate with preservation of the quality of speech signal.

The applying a linear prediction only on the approximated coefficients signal with out detail signal was allowed the evaluation of the speech-compressed signal.

It is also observed that the efficient of this technique can be varies with the coding of the prediction error and the adopted method for the choice of the coefficients to truncate.

REFERENCES

- [1] Frédéric Truchetet, "Ondelettes pour le signal numérique," collection traitement du signal, 1998, chap.4, 25-51.
- [2] I. Daubechies, "Ten lectures on wavelets," society for industrial and applied matematics 1992.
- [3] Ankil Patel, Mark Tonkelowitz and Mike Vernal, "Lossless sound compression using the discrete wavelet transform," 2002.
- [4] René Boite, Hervé Bourlard, Thierry Dutoit, Joël Hancq et Henri Leich "Traitement de la parole," 2000 Presses plytechniques et universitaires romandes, Collection électricité, chap. 2, pp 27-64.
- [5] Stephane Mallat and Sifen Zhong , "Caracterization of signals from multiscale edges", IEEE Transactions on pattern and machine intelligence, Vol. 14, N°7, July 1992.
- [6] Jalal Karam, Raed Saad "The Effect of Different compression schemes on speech signals," International Journal of Biomedical Sciences Volume 1 Number 4, 230-234.
- [7] Jalal Karam, 'Various Speech Processing Techniques for Speech Compression and Recognition', Proceeding of world academy of science, engineering and technology volume 26 décembre 2007 issn 1307-6884.
- [8] Amhamed Saffor, Abdul Rahmen Ramli and Kwan-Hoong Ng "A comparative study of image compression between JPEC and wavelet," Malaysian Journal of computer science, volume 14 No. 1, June 2001, pp 30-45.