Design of two blocks of a speech coding system to be implemented on an FPGA based card

T. Benmakhlouf[1], S. Mekaoui[1] & K. Ghoumid[2]
[1]*Département des Télécommunications, L.C.P.T.S, USTHB, Alger, Algérie*
[2]*ENSAO, Laboratoire d'Electronique et Télécommunication, Complexe Universitaire, Oujda, Maroc*

Abstract

The main aim of this paper is concerned with modern telecommunications systems which involve modern methods of coding, encryption and decryption of speech signals. For a long time, for the transmission of a speech signal analog telecommunications systems have been used. Because of unexpected and unavoidable interference, wave fading perturbations and different kinds of noise occurring in the channel, it was not possible to detect and receive the same transmitted speech signal. Consequently, digital systems have steadily replaced the former. Here, we have simulated two blocks of such systems, namely the source coding block and the encryption/decryption block. We tested them by listening to the synthesized signals via headphones and using a simulation operated using Simulink of the source software Matlab. Although metallic in their tonalities, results were found to be acceptable.
Keywords: digital speech signal processing, digital data transmission, Simulink, LPC algorithm, AES algorithm, time-frequency analysis, STFT transform, analysis/synthesis of speech signal.

1 Introduction

One specific need when transmitting information through data communication systems is the increasing flow rate and the occupied bandwidth in its spectrum, especially for the transmission of speech signals through a digital coding chain. A second requirement is the constraint of the protection and the security of the transmitted information insuring the confidentiality of the exchanged messages

WIT Transactions on Information and Communication Technologies, Vol 45, © 2013 WIT Press
www.witpress.com, ISSN 1743-3517 (on-line)
doi:10.2495/DATA130071

and data. In this way, many encoding and encrypting algorithms have been developed. However, an encrypted signal requires a long time for treatment which leads to large time delays in the transmission of the coded signal. So, it appears as a necessity to add to the coding block of the speech coding chain a block of compression which then considerably reduces the transmission time. This fact induced the design of particular speech coding algorithms such as the well known LPC (linear predictive coding) algorithm [1], and the LDPC algorithm [2]. Thanks to these kinds of algorithm many speech coding applications have been implemented on DSP based hardware cards or on FPGA (Field Programmable Gates Array) based hardware cards [3]. Since the breaking and the hacking of the DES encryption algorithm, the NIST organization has launched an international offer to the market to design a new product that can best replace the DES algorithm having the advantage of a secret key to insure the confidentiality of the transmitted information with high reliability and precision [4]. Then, in 2000, a new standard of algorithm was designed by J. Daemen and V. Rijmen [4, 5], which became the new Advanced Encryption Standard, simply known as the AES algorithm, which insured a high level of security and confidentiality. Since that time many implementations have been performed on FPGA hardware based cards [6–8], which processed well the Encryption/Decryption of the speech signal before being transmitted, and recovered and correctly recognized the signal at the receiver. Our work focuses on the simulation of these two chaining blocks (Source coding, Encryption/Decryption) of a speech coding chain to be implemented on a FPGA hardware base card. This second step will be processed later. So, in this paper we are much concerned with the study of these two blocks. The system comprises a block of compression followed by a ciphering unit in order to implement both operations in a S.O.C. chip based system. Then, the LPC coding (analysis/synthesis) consisting of the encryption and decryption using the AES algorithm is described. Simulink software was used to implement and display the simulation of the digital coding chain. Time Frequency Analysis (TFA) in its STFT (Short Time Fourier Transform) tool and its well known spectrogram have been performed to validate the results. Work is in progress to generate, with the help of the HDL coder of Simulink, the VHDL code necessary for the implementation of the whole digital chain on an FPGA based hardware card.

2 Digital transmitting chain

A digital telecommunication chain transmits the speech information from a source to a receiver via a transmission channel. The source (transmitter) and the receiver can be at short or at long distances. Usually, this information is private and requires security and confidentiality. A digital transmitting chain frequently presents interesting advantages in terms of noise protection and error correction as it has a coding channel stage. Therefore, most digital communication systems are built on the global block diagram as illustrated in Figure 1.

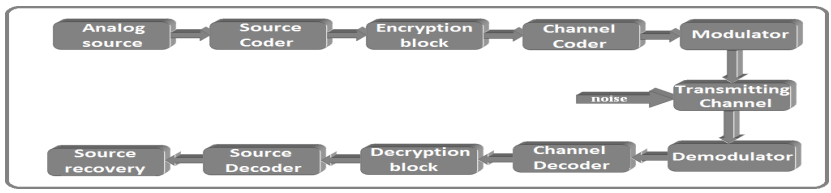

Figure 1: Block diagram of a digital transmitting chain.

3 Source coding block

Speech analysis is an important prior treatment to the coding, the synthesis or to the recognition of the speech signal. This analysis relies on a particular model which consists of a set of digital parameters whose variations define the different signals covered and accepted by the model.

3.1 Source modeling

Figure 2 shows the most common and accepted biological model illustrating the physical production of a speech signal. This well known model, inspired from human nature, can also be seen as an adaptation between the biological nature and the mathematical modeling of the voice tract to produce the speech signal. Thus, it is generally called the "source-filter-model".

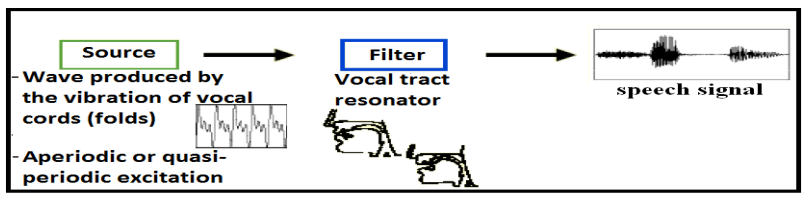

Figure 2: Source-filter-model.

Consequently, the idea was the modeling of the vocal tract by a recursive filter of type (1/A(z)), the air flow-rate from the lungs by an excitation signal u(n) and finally the air volume by a gain parameter denoted G.

3.2 LPC synthesizer

In 1960, Fant [10] proposed a pattern that specified that a voiced signal can be modeled as a pulse train u(n) passing through a recursive filter of all poles type. This assertion was shown to be still valid for unvoiced signals unless u(n) is white noise. The final model is illustrated in Figure 3. This model is also called an auto-regressive (AR) model as it corresponds to a linear regression in the time domain which has the following expression:

$$X(n) = G \cdot u(n) + \sum_{i=1}^{P} -a_i X(n-i), \quad (1)$$

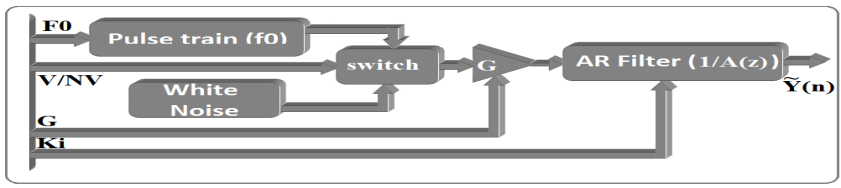

Figure 3: Block diagram of the LPC synthesizer.

3.3 LPC analysis

After the implementation of the first speech synthesizer which involved the use of the K_i, F_0, and G parameters, the question was how could we extract these parameters from the speech signals. Among the various methods proposed at that time was LPC analysis [11, 12]. Figure 4 summarizes the specific steps performed in an LPC analyzer. So, in this analysis, the first step is to determine the predictive coefficients and calculate the gain. Then, the second step is concerned with the pitch extraction. The later operation is slightly more complicated as the human ear is more sensitive to pitch variations.

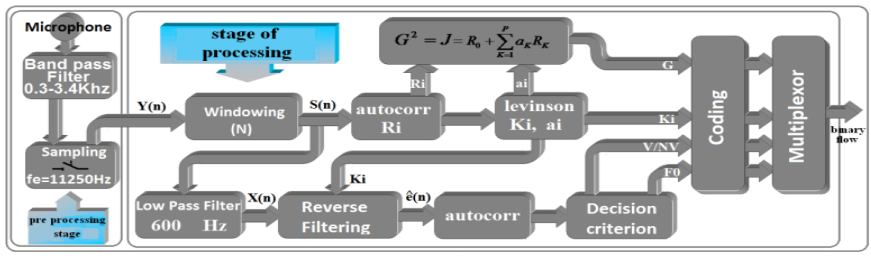

Figure 4: Block diagram of the LPC analyzer.

3.3.1 LPC coefficient determination

The LPC coding consists of estimating the value of the future samples in terms of a few previous ones noted S[n-k], [11, 12]. Then, equations (2) and (3) express respectively the estimated sample value and its prediction error:

$$S'[n] = \sum_{k=1}^{P} a_k S(n-k) \quad (2)$$

$$e(n) = s(n) - \sum_{k=1}^{P} a_k s(n-k) \quad (3)$$

Then, to obtain the LPC coefficients, we apply the least mean square criterion by minimizing the energy error following equation (4):

$$J = \sum (s(n) - \sum_{k=1}^{P} a_k s(n-k))^2 \quad (4)$$

This criterion will be satisfied unless the following derivation will be made equal to zero and solved:

For: $1 \leq i \leq P$;

$$\frac{\delta J}{\delta a_k} = \frac{\delta\left[\sum\left(s(n)-\sum_{k=1}^{P} a_k s(n-k)\right)^2\right]}{\delta a_k} \tag{5}$$

Solving (5) leads to equation (6) which assumes that the signal is stationary on an interval of 15 to 25 milliseconds:

$$\sum_{n1}^{n2}\sum_{k=1}^{P} a_k s(n-k)s(n-i) = \sum_{n1}^{n2} s(n)s(n-i) \tag{6}$$

Among various methods that can be applied to solve equation (6) are two reputed techniques known as the autocorrelation function and the covariance methods. We personally have chosen in this paper the autocorrelation function method which consists of calculating the short-term autocorrelation function of the signal defined in (7) by:

$$R(i) = \sum_{n=0}^{N-1} s(n)s(n-i) \tag{7}$$

Substituting (7) into (6) yields the following system of equations:

$$R(i) = \sum_{k=1}^{P} a_k R(i-k) \tag{8}$$

Then (8) should be solved by encountering the number of calculations to perform. The classical algebraic methods require P^3 operations whereas the Levinson algorithm only P^2. The well known Levinson-Durbin algorithm allows the solution of the system of equations given in (8) by operating recursive iterations of the order P and hence revealing three interesting sets of parameters, namely, the predictive coefficients *(a_i)*, the energy prediction coefficients *(E_i)* and the reflection coefficients *(k_i).* For more details on this method one should refer to references [9, 10].

3.3.2 Pitch determination (F_0)

As it is very difficult to estimate or calculate the fundamental frequency F_0 (pitch) of the speech signal, many techniques have been proposed to extract the pitch. One very reputed method is the SIFT algorithm [13]. Markel [13] found that the pitch can be extracted by using a technique based on a reverse filter whose transfer function is given by:

$$A(z) = 1 - \sum_{k=1}^{P} a_k z^{-k} = \frac{1}{H(z)} \tag{9}$$

This technique also uses the observation of the autocorrelation function of the LPC residue e(n) and is built around the following structural steps :

a) Signal filtering

We know that the frequency range of the pitch from the speech spectrum analysis of the human kind is in the range [80–600] Hz. So, it is necessary to use a low pass filter whose cut-off frequency is about 600 Hz [13].

b) Research of the excitation signal e(n)

The residual signal e(n) from the linear prediction process is considered as the excitation signal that can generate the s(n) signal through a recursive filter:

$$H(z) = \frac{S(z)}{E(z)} = \frac{1}{A(z)} \tag{10}$$

In our case s(n) is known and we can then obtain the residual e(n) by reverse filtering, that is to say:

$$E(z) = S(z) \cdot A(z) \tag{11}$$

This last operation is equivalent to the convolution of the coefficients a(n) with the signal s(n):

$$e(n) = a(n) * s(n) \tag{12}$$

c) Autocorrelation function

As the signal is corrupted by noise, the determination of the period will be made easier if we calculate the autocorrelation function of e(n) [6]. Hence, the later will result in a vector of length 2N-1, where N is the total number of samples and where the function will be a maximum at exactly the middle of this vector length.

d) Decision criterion (interpolation)

The autocorrelation of the residue e(n) can reveal several peaks, one at the origin and another one at a second position on the axis if the signal corresponds to a voiced signal. To consider the second peak as significant, its amplitude should be 40% (per cent) that of the first peak at the origin. Then the estimated distance between both peaks is the excitation frequency (F_0). If the second peak does not exist, then the signal is considered to be an unvoiced signal. Figure 5 depicts a tested example for a voiced frame of the autocorrelation of the e(n) signal whereas Figure 6 shows a typical test for an unvoiced frame of autocorrelation of the e(n) signal. The difference is obvious.

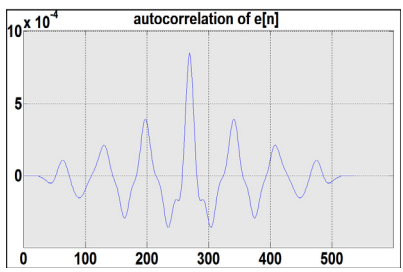

Figure 5: Autocorrelation of a voiced frame.

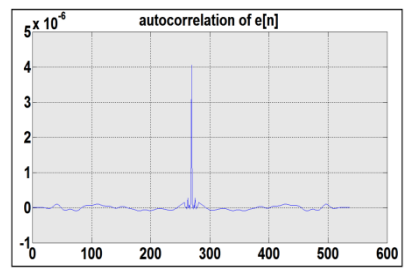

Figure 6: Autocorrelation of an unvoiced frame.

3.3.3 Gain calculus (G)

The overall energy contained in the synthesis sequence of the signal should be the same as that of the corresponding analysis sequence whatever the signal (voiced or unvoiced) and is determined by:

$$G^2 = E_P = R_0 - \sum_{i=1}^{P} a_i R_i \quad (13)$$

Subsequently, the gain G is the root square of the total minimal quadratic error.

4 Encryption/decryption blocks

For some reason, telecommunications operators have always been interested in protecting and securing the transmission of information, specifically for speech systems insuring in this process the confidentiality of the transmitted speech information. This is known as the cryptography process today and requires sophisticated algorithms. In our case, we have chosen the above cited AES algorithm of ciphering, which proceeds by symmetric blocks and iterative operations called "rounds" with variable block size and key size. Indeed, the AES algorithm can bear block sizes and key sizes up to 128, 192 or 256 bits, independently. Obviously, it is impossible to detail all the processes of encryption and decryption here. To learn more about it, one should refer to [14] and [15].

5 Simulation results and discussion

5.1 Blocks simulation

For economic reasons, a simulation is obviously performed prior to any hardware implementation. This caution also allows the optimization and validation of the results. In this application, we realized using Simulink the LPC analysis/synthesis with the help of embedded function blocks of Matlab. Figure 7 presents the different blocks involved in a speech coding chain.

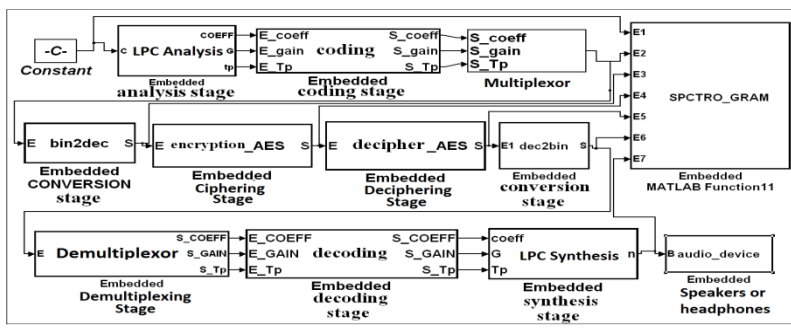

Figure 7: Simulation flow chart under Simulink of the LPC blocks.

To read the energies in the time frequency plane and observe the variations of the speech signal in the time and frequency planes, we have implemented the Short Time Fourier Transform (STFT) and its spectrogram on the speech signal. The results obtained are illustrated in Figures 8, 9 and 10. It can be noticed that in each figure a spectrogram is displayed in the time-frequency plane.

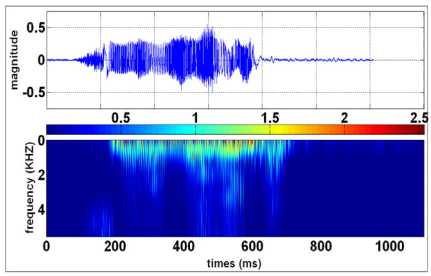

Figure 8: Original signal and its spectrogram.

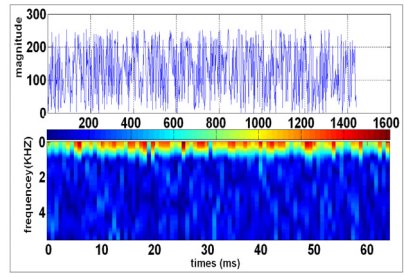

Figure 9: Encrypted LPC coefficients and their spectrogram.

Concerning this first set of results, we implemented the synthesis blocks. The quality of the synthesized signal was not so good but remained acceptable if we disregard the strict intelligibility of the speech signal. Figure 10 illustrates the synthesized speech signal and its spectrogram.

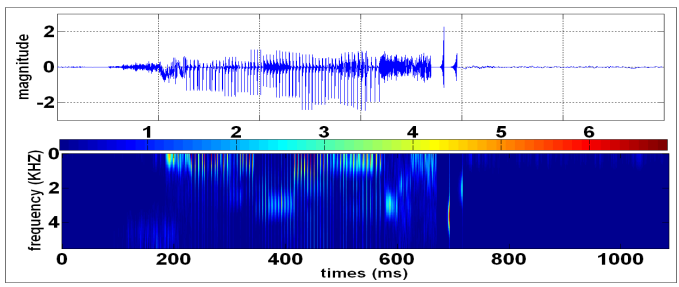

Figure 10: Synthesized speech signal and its spectrogram.

5.2 Tests on the encryption block

We applied the same tests to the encryption block and we got the results illustrated in Figures 11, 12 and 13 which display the spectrogram of the STFT of the speech signal. The speech segment corresponds to a selected pronounced statement of the Arabic language which was repeated many times by several males.

As we tested the blocks separately, we have been able to localize the cause of signal weakening at the synthesis step, which is caused mainly by the energy modification on the one hand and the coding errors on the other. The latter was

found to be logically explained as it is directly bounded to the compression stage whereas the encryption block does not introduce any kind of signal degradation or corruption.

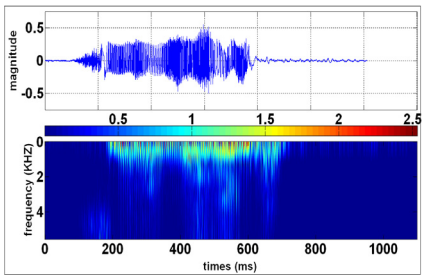

Figure 11: Original signal and its spectrogram.

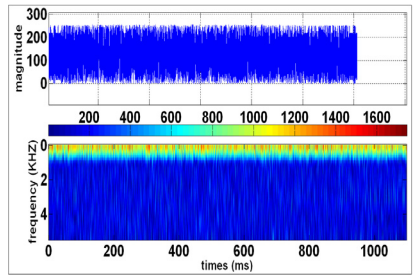

Figure 12: The encrypted signal and its spectrogram.

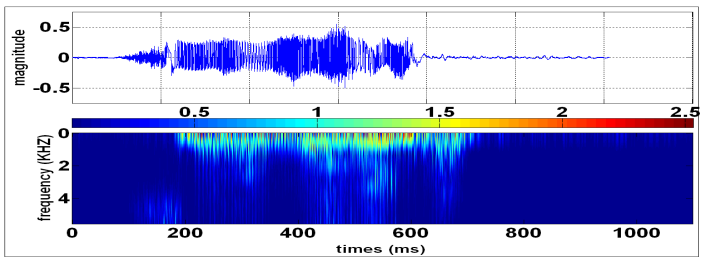

Figure 13: The synthesized signal and its spectrogram.

6 Conclusion

In this work, our main goal was to implement in real time two important chaining blocks of a digital speech coding transmission chain of a telecommunications system. At first, we focused on the simulation of the speech coding via the LPC analysis/synthesis blocks that insure a flow rate of 2.4 to 16 Kbits/second. As the LPC model is one of the basic models in the treatment of the speech signal, the quality of the synthesized speech signal was not very good but considering the compression ratios (13 LPC parameters instead of 256 samples) remains acceptable with a considerably reduced execution time. The flow rate of the source is about 4.3 Kbits/second obtained thanks to a uniform quantization and an 8 bit coding resolution for each parameter. The encryption block has insured the confidentiality of the transmitted signal and the results were examined by listening via headphones and found to be acceptable. Work is in progress to implement these blocks on an FPGA hardware based card.

References

[1] J. D. Markel and A. H. Gray Jr. “Linear Prediction of Speech” New York: Springer-Verlag, 1976.
[2] V. A. Chandrasetty and S. M. Aziz “FPGA Implementation of High Performance LDPC Decoder using Modified 2-bit Min-Sum Algorithm” 05/2010; In proceeding of: Computer Research and Development, 2010 Second International Conference. School of Electrical and Information Engineering University of South Australia Mawson Lakes, SA 5095, Australia, 2010.
[3] M. A. Raza, P. Akhtar “Implementation of voice excited linear predictive coding (velp) on TMS 320C6711 DSP kit” PNEC, National University of Science & Technology (NUST), Karachi, Pakistan.
[4] Nation Institute of Standards and Technology (NIST), Data Encryption Standard (DES), National Technical Information Service, Springfield, VA 22161, Oct. 1999.
[5] J. Nechvatal *et al.*, “Report on the development of Advanced Encryption Standard” NIST publication, Oct 2, 2000.
[6] Marko Mali, Franc Novak and Anton Biasizzo “Hardware Implementation of AES Algorithm” Journal of Electrical Engineering, Vol. 56, No. 9-10, 2005, 265-269.
[7] L. Thulasimani, “A Single Chip Design and Implementation of AES - 128/192/256 Encryption Algorithms” International Journal of Engineering Science and Technology, Vol. 2(5), 2010, pp. 1052-1059.
[8] T Good, M. Benaissa, “Very small FPGA application specific instruction processor for AES”, IEEE Trans. Circuit and System, vol. 53, no. 7, 2006.
[9] R. Viswanthan and J. Makhoul, “Quantization properties of transmission parameters in linear predictive system” IEEE Trans Acoustic Speech and Signal process. Vol. ASSP-23. No 3, June 1975.
[10] G. Fant, “Acoustic Theory of speech production”, Mounton and Co, Gravenhage, The Netherlands, 1960.
[11] S. Grassi, “Optimized Implementation of Speech Processing Algorithms”, PhD thesis, faculté des sciences de l’Université de Neuchâtel pour l’obtention du grade de docteur ès sciences, February, 1998.
[12] J. Bradbury, “Linear Predictive Coding”, December, 5, 2000.
[13] R. Boite et M. Kunt, “Traitement de la parole”, Presses Polytechniques Romandes. 1987.1 vol (280p).
[14] Federal Information Processing Standards Publication 197, “Announcing the Advanced Encryption Standard (AES)”. November 26, 2001.
[15] J. Daemen and V. Rijmen, The Design of Rijndael, AES, The Advanced Encryption Standard, Springer-Verlag 2002.