

## ***New Design Approach of an FIR Filters Based FPGA-Implementation for a Bio-inspired Medical Hearing Aid***

*Lotfi Bendaouia*

*Equipe de Traitement de l'Information et des Systèmes  
CNRS ENSEA UMR 8051  
Cergy, France  
[lotfi.bendaouia@ensea.fr](mailto:lotfi.bendaouia@ensea.fr)*

*Hassen Salhi*

*Département d'électronique et d'informatique  
Université Saad Dahleb  
Blida, Algérie  
[labset@yahoo.fr](mailto:labset@yahoo.fr)*

*SiMahmoud Karabernou, Lounis Kessal*

*Equipe de Traitement de l'Information et des Systèmes  
CNRS ENSEA UMR 8051  
Cergy, France  
[karabernou@ensea.fr](mailto:karabernou@ensea.fr), [lounis.kessal@ensea.fr](mailto:lounis.kessal@ensea.fr)*

*Fayçal Ykhlef*

*Architecture des Systèmes et Multimédias  
CDTA  
Baba Hassen, Algérie  
[fykhlef@cdta.dz](mailto:fykhlef@cdta.dz)*

**Abstract**—The work focuses a new design approach to a hardware implementation of a Bio-Inspired Medical Hearing Aid which has the specificity to be portable and hence needs reduced resources and low power. This paper describes how we appropriately applied hardware optimization in the aim to reach on computationally intensive DSP algorithms for use to improve performance and efficiency of a hearing aid device on FPGA. Deaf people suffer from their social disease, so a device which could correct their hearing loss is needed. Even the technology advances these embedded devices can still be optimized for low cost. Contributions, mainly focus area reduction and hence low power consumption and dissipation. We propose a new design approach to meet the specifications of this embedded system.

**Keywords**—Hearing aid; DWT; FIR; area; area; Latency; Power consumption.

### I. INTRODUCTION

Voice conversation is one of the most important tools for communication whereas some people don't have profit from this opportunity because of their hearing loss. Speech is badly detected by these impaired persons yielding to poor intelligibility. This happens mostly in noisy and reverberant environment [1][2].

Since many years, prosthesis and cochlear implants have been used but deaf persons are still feeling uncomfortable and suffer from this social handicap.

A Lot of researchers were held for speech enhancement leading to many contributions for algorithms development and circuits' design with less complexity and fast processing [3][4][5].

Knowing that it is difficult, say impossible, to reproduce the natural hearing for impaired persons, our contribution is to make the filtered signal more closely to the original one for better intelligibility and hearing comfort.

The DSP algorithms [6] were implemented using traditional DSP or general purpose microprocessors. It was released that they have limited capabilities for processing

high volume data efficiently at real time. The trends were shifted to specific processors such as Asics in order to meet the increased complexity and performance requirements of these algorithms but with high cost function.

The rapid growth in the industrial technologies has participated in the development of several and performed hardware digital signal processing application systems. The implementations of intensive computational DSP algorithms become for researchers a day to day application area for digital hardware platforms [7][8].

FPGA [9] has taken a large area of use because of some advantages over Asic technologies.

Several hardware platforms [10][11][12] were designed with different combinations and optimizations of filters structures. They were simulated and synthesized using Xilinx or Altera FPGA development kits.

The impact of these hardware optimization techniques on the overall DWT hardware system are analyzed and the tradeoffs between the pertinent hardware performance metrics particularly power consumption, latency, resource utilization and operating frequency are considered and investigated [13].

Today, FPGAs are highly preferred to the relatively high capacity and low cost, short design cycle and short time to market. They afford the capability of constant reconfiguration to meet application performances which are highly preferred.

Recent FPGA includes enhanced signal processing capabilities of high performance logic and inherent parallelism enabling FPGAs to have special Multiply-Accumulate (MAC) blocks within its hardware.

By using FPGA, the design can be simulated and then synthesized with low cost. The hardware design is then ready for fabrication and use.

Our work is focused in the implementation of an efficient multi level one dimension Discrete Wavelet Transform (DWT) on FPGA for medical hearing aid application.

The proposed architecture combines hardware optimization techniques to develop a flexible DWT architecture that has a high performance and is suitable for portability, high processing speed and power efficiency [9][10] in order to optimize the hardware, we have reduced FPGA resources by using some techniques and orientated the VHDL [14] program in order to have a synthesized IP-based on customized DSP slice resources of the FPGA.

The paper is divided into four main sections. An overview of deafness is presented in the next section. Section three illustrates how the DWT algorithm is modelled to the cochlea structure. In Section four, we present the implementation of the FIR filters and show the simulations, the synthesis and the appropriate approximations. Performance analysis of the whole system and the perspectives for future work are presented in the conclusion.

## II. OVERVIEW OF DEAFNESS

Impaired persons are affected by some perturbation phenomena, namely noise and echo, which reduces the intelligibility and hence their capability for understanding and communicating. Some studies were conducted to assess the status of noise and echo stripping.

### A. Noise Reduction

Noise overlaps speech signal and can be suppressed or reduced by a low pass filtering as shown by Figure 1. However, when applying such action on a signal, we sometimes rescue to the elimination of some singularities of the signal which can contain significant information.

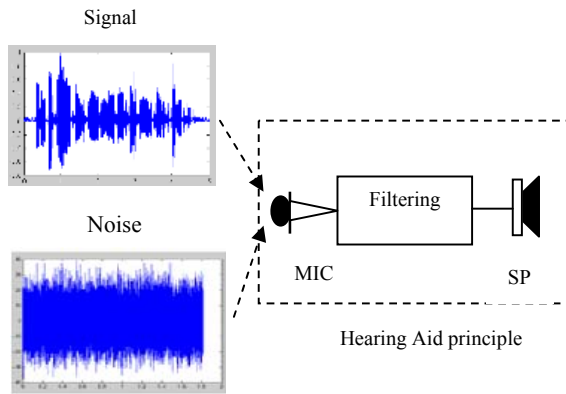


Figure 1. Speech signal versus noise separation .

In order to improve intelligibility, noise should be reduced but cannot be fully eliminated from speech because nothing is known about this latter and how far is it modeled with speech [15].

Care should be taken when applying denoising algorithms to avoid making any severe degradation of the resulting signal. The time-scale analysis is the perfect solution for speech processing as we will in Section III. Noise overlaps the speech signal in both time and

frequency; so, it is difficult to remove it completely. Hence, greatest attention was made in the development of noise reductions techniques.

### B. Echo Cancellation

The Acoustic feedback refers to the acoustical coupling between the loudspeaker and the microphone of the hearing aid. The amplified sound sent through the loudspeaker is sometimes fad back into the microphone as shown in Figure 2.

As a result, the original signal gets distorted and then poor intelligibility occurs. One direct solution is to reduce the gain, but this limitation generates low energy making signals falling below the hearing threshold and no compensation is made to the impaired persons.

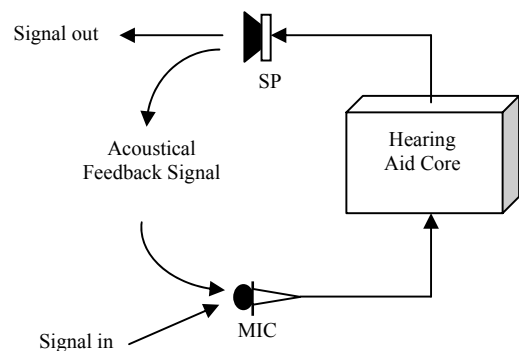


Figure 2. Acoustical feedback in hearing aids .

Acoustic feedback suppression techniques are suitable to increase the maximum gain of the system without making it unstable, as shown in Figure 3.

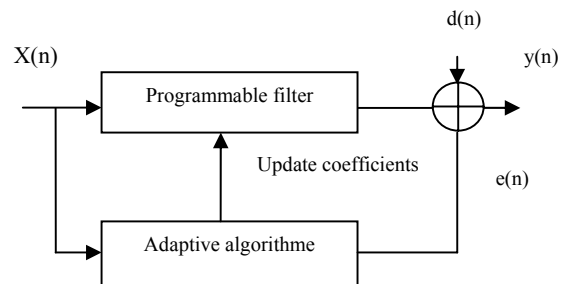


Figure 3. Adaptive filtering diagram .

Wiener adaptive filtering techniques [16] consider stationary signals and use, LMS [17], NLMS [18] or RLS [19] algorithms. For non-stationary signals, generalized techniques are used based on Kalman filtering [20].

The estimated error is given in [16] by:

$$e(n) = X(n) - \sum_{j=1}^m a_j X(n-j) \quad (1)$$

The objective is to find the optimal coefficients which make the error as small as possible. In order to do that, we should minimize the energy of the total error given by the equation:

$$E = \sum_{n=0}^{N-1} e(n)^2 = \sum_{n=0}^{N-1} [X(n) - \sum_{j=1}^m a_j \cdot X(n-j)]^2 \quad (2)$$

By making  $\partial E / \partial a_j = 0$ , we can obtain the coefficients  $a_j$  from the  $m$  generated equations.

### III. THE BIO INSPIRED ALGORITHMIC MODEL

Before making the algorithm model, we first make an overview on how the basilar membrane acts as a group of mimicked filters.

#### A. Basilar Membrane Modeling (BMM)

The cochlea is an opened spiral tube lying in the middle ear. The opening in its base makes possible the penetration of the sound signals. The closed end is called the apex. The sound is detected and coded according to its frequency but it is place coding on the basilar membrane. The sounds of high frequencies are detected at the base, whereas those of low frequencies are detected at the apex. The frequencies are distributed along the basilar membrane in a very precise manner as represented by Figure 4.

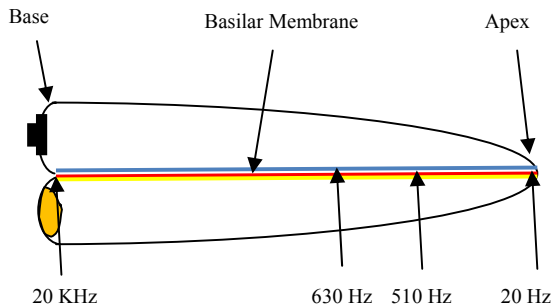


Figure 4. Longitudinal view of the cochlea .

The filters within the cochlea are distributed in bands along the basilar membrane and are responsible of the selectivity of the sound frequency in the ear. These filters can be modelled in a pseudo-logarithmic way. They are either triangular (Mels) or rectangular (ERB). The bands are linear up to 500 Hz and logarithmic beyond. Each bandwidth can be determined using the formula [16]:

$$\Delta f = 25 + 75 \left[ 1 + 1,4 \left( \frac{f}{1000} \right)^2 \right]^{0,69} \quad (3)$$

where  $f$  is the central frequency of the band.

#### B. Survey of the DWT

The Discrete Wavelet Transform [22] is an important approach for the analysis of a transient signal. The connection was made between the wavelet transform and multi-rate filter bank trees by Mallat since 1989 [13]. From signal processing point of view, the wavelet transform of a sequential signal is to recursively decompose a sampled sequence of a signal into two components in octave bands. A recursively asymmetric decomposition levels, leads to a similar bands distribution as the basilar membrane which make the DWT the adequate algorithm [23]. Three level wavelet decomposition is shown in Figure 5.

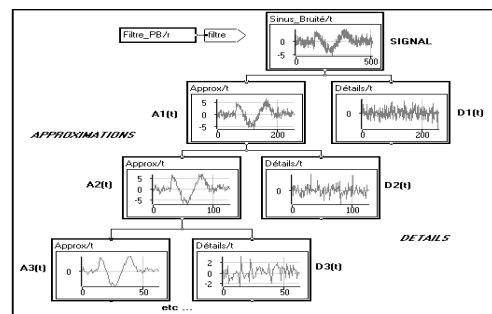


Figure 5. The level decomposition of Wavelet analysis .

The input signal is spread into two signals by the Low Pass Filter giving what we call the approximation of the signal and, by the High pass filter giving a detail of the signal at the first level. The process can be repeated on for other levels using a symmetric or asymmetric decomposition. The wavelet coefficients can be used to reconstruct the original signal without any distortions. The LPF and HPF are Finite Impulse Response (FIR) Filters. FIR filters are the basis of the DWT. For a dyadic representation, the basic analysis / synthesis structure of the DWT is represented by the Quadratic Mirror Filters (QMF) shown in Figure 6.

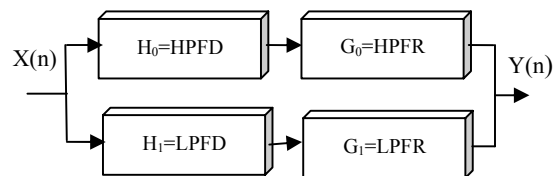


Figure 6. Quadratic Mirror filter structure .

where the  $H_x$  filters state for Decomposition (D) and  $G_x$  for Reconstruction (R). HPF means High Pass Filter and LPF is the Low Pass Filter. These filters are related by the following equations:

$$H_0(z) \cdot G_0(z) + H_1(z) \cdot G_1(z) = z^{-T} \quad (4)$$

$$H_0(-z) \cdot G_0(z) + H_1(-z) \cdot G_1(z) = 0 \quad (5)$$

By setting:  $G_0(z) = H_1(-z)$  and  $G_1(z) = -H_0(-z)$ , we satisfy:

- Perfect reconstruction with latency (T)
- No aliasing

The coefficients are obtained by Matlab using Wfilters for different wavelet. They are stored as real and then converted to fix point numbers so as they can be treated in the hardware design circuit.

The selection of the FIR filters is due to coefficient sensitivity, round off noise, stability and are suitable for high speed applications [24].

The FIR filters use a convolution principle of the input signal  $X(n)$  by the impulse response  $h(n)$ . The output  $Y(n)$  is given by the following mathematical expression:

$$Y(n) = \sum_{k=0}^{L-1} h(k) \cdot X(n - k) \quad (6)$$

The FIR filter is composed of multipliers, adders and delay units. Recent FPGA includes DSP48A1 elements making ideal to implement DSP functions.

The  $n$  input samples from the data set are presented at the input of each DSP48A1 slice.

Each slice can be used to multiply these samples with the corresponding coefficients within the DSP48A1. The outputs of the multipliers are combined in the cascaded adders. A basic DSP48 slice is shown in Figure 7.

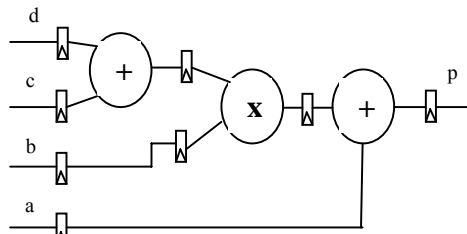


Figure 7. DSP Slice 48 A1 with Pre-Adder .

The sample delay logic is denoted by  $Z^{-1}$ , the (-1) represents a single clock delay. The delayed input samples are supplied to the one input of the multiplier. The coefficients represented by  $(h(0) \text{ to } h(N-1))$  are supplied to the other input of the multiplier through individual ROMs, RAMs, registers or constants. The output  $Y(n)$  is merely the summation of a set of input samples, and in time, multiplied by their respective coefficients. The DSP48A1 lying inside the FPGA is suitable for low power dissipation and high throughput based pipelining and parallel processing [25].

#### IV. ARCHITECTURE IMPLEMENTATION

The filter structure of FIR of length  $L$  (called order of the filter) is represented in Figure 8. This structure describes the relationship between the input and output sequences.

The input samples are delayed and multiplied by the suitable coefficients and then added to give the output at time  $n$ .

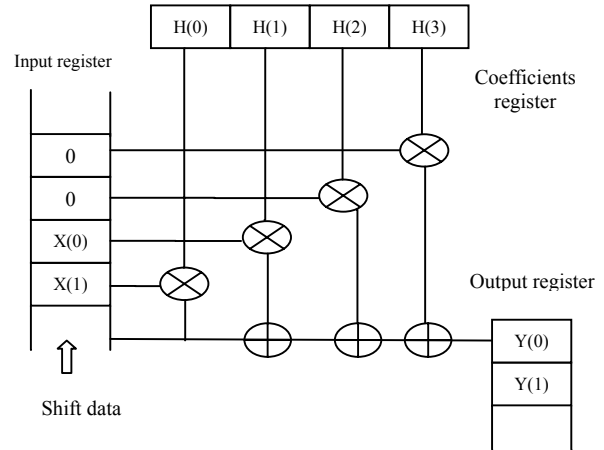


Figure 8. Convolution principle in FIR .

The architecture of each FIR block includes a FIFO register for data input, a register for the coefficients and the operators. The data output is stored in memory. The FIR block is presented in Figure 10.

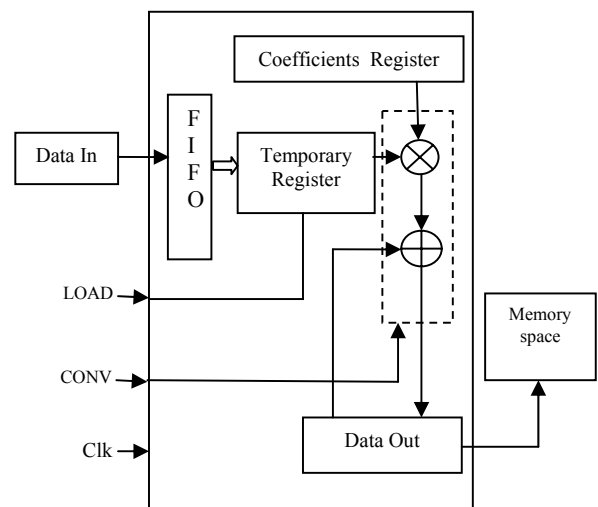


Figure 9. Basic FIR filters architecture Design .

The FIFO at the input is filled by the input data samples. An input register of length  $N$  is used to store the input sequence  $X(n)$  taken from the FIFO. These samples are convolved with the coefficients which have been already stored in a coefficients register. The output sequence is also stored in a FIFO register (memory).

V. EXPERIMENTAL RESULTS

The experiments were held with a real man speaking speech signal “the Discrete Fourier Transform of a real value signal is conjugate-symmetric”. The wave signal is sampled at a frequency  $F_s = 22050$  Hz. We take blocs of 10000 samples by Hamming windowing. For our experiments, Daubechies 4 was chosen. The filters coefficients are presented in Table I.

TABLE I. GENERATED FILTERS’ COEFFICIENTS FOR DB4

Coefficients	LPFD	HPFD	LPFR	HPFR
1	- 0.0106	- 0.2304	0.2304	- 0.0106
2	0.0329	0.7148	0.7148	- 0.0329
3	0.0308	- 0.6309	0.6309	0.0308
4	- 0.1870	- 0.0280	- 0.0280	0.1870
5	- 0.0280	- 0.1870	- 0.1870	- 0.0280
6	0.6309	0.0308	0.0308	- 0.6309
7	0.7148	- 0.0329	0.0329	0.7148
8	0.2304	- 0.0106	- 0.0106	- 0.2304

The input data samples and the coefficients were quantified and approximation tests were done for 8, 10, 12, 14, 16 bits. The better approximation was obtained for 12 bits and over.

The 12 bits quantization has been chosen for the rest of the experiment. The input samples and the coefficients have been converted to signed fix point data using a Q1.11 format. The output results after multiplications and additions are of format Q5.22. So, a truncation is done giving an output data of Q1.11 format with negligible errors. The samples are converted from floating to fixed point numbers, so they can be treated by the VHDL programs and then compared to those obtained by Matlab. A flow chart is given by Figure 11.

A. Simulations

In the VHDL simulations process, we generated the outputs of the design which are compared to the outputs generated at the algorithmic level using Matlab. We used for accuracy in metric estimation, the peak error and the Mean Square Error (MSE). A maximum performance is achieved with an error of less than 0.3 %.

In Figure 10, we make a comparison between the data obtained from the resulting output samples stored in the output files, using Matlab textread and plot functions.

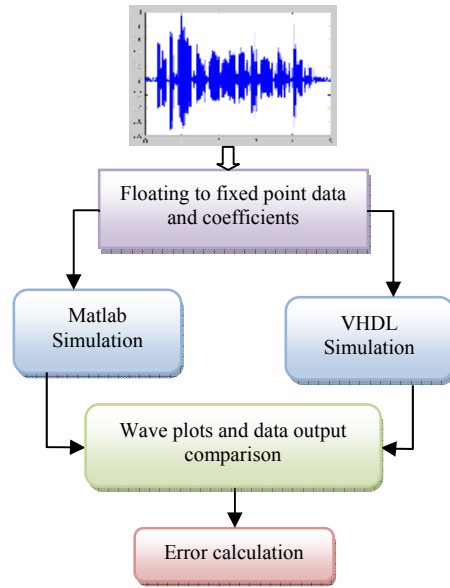


Figure 10. Matlab Versus VHDL simulation analysis .

B. Synthesis

For post-synthesis, we used the EP2C70F896C6 device of Cyclone II family at 100 Mhz clock. A VHDL Netlist containing Altera simulation primitives was generated and has been used again for correct compilation and simulation.

The design was synthesized using QuartusII giving the circuit at the Register Transfer Level (RTL) shown by Figure 11. The architecture needs registers of N bits if the data input is N bits but, our suggested architecture only needs L bits FIFO register and therefore we get an optimized structure for implementation.

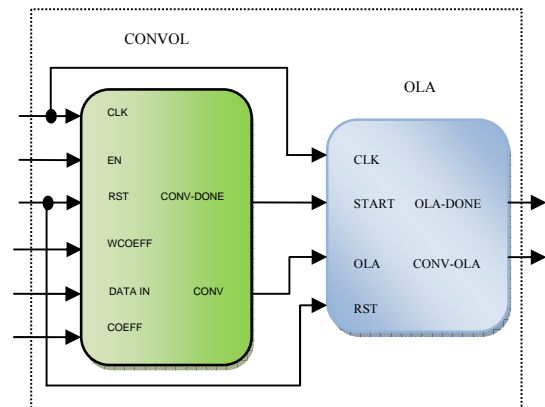


Figure 11. FIR block Diagram at RTL .

As shown by Figure 11, the basic FIR bloc is divided into two processing units. The first calculates the convolution coefficients with the L first bits and then with the L next bits with a save f the overlap data. Only a few bits are saved dealing to a gain of space memory.

The technological schematic is given by Figure 12. We can see that the circuit use combinatory components such as adders and multipliers. The system is controlled by a state machine which synchronizes the data acquisition and data treatment processes.

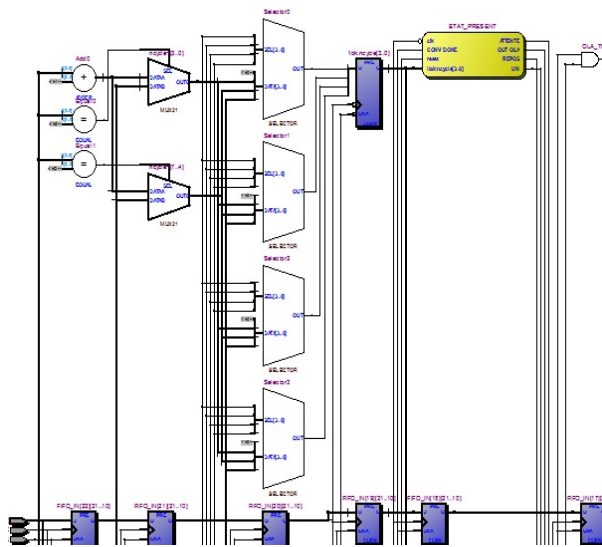


Figure 12. FIR Technological Block Diagram .

The performance metrics show the efficient contribution of FPGA resources in the implementation of the FIR filters which are the basic elements for the global system design (DWT). This hardware design has permitted the reduction of the used area; the used resources are less than 4% and the optimization of the latency and the power consumption. The result summary for resources utilization from synthesis is given by Table II.

TABLE II. EDA NETLIST FINAL REPORT

Logic elements	TF	DF
LUTs	1668 (2%)	6157 (9%)
Registers	1868 (3%)	3420 (5%)
Memory	240 (< 1%)	80640 (7%)
Multipliers	32 (21%)	96 (64%)

It is also compared to that obtained in previous work using direct form design. From the plotted graphs, we can see that the curves are totally overlapped, meaning of the very accurate results; see Figure 12.

VI. RELATED WORKS

The direct form (DF) FIR filters architecture was first implemented by Baganne et al. [26]. This architecture was modular and had low design complexity, low hardware

latency and could be easily expanded to further levels of decomposition. However, this architecture had a large critical path delay, needs more resources and high power dissipation.

We have first implemented this architecture in order to compare it with our new design approach [27]. For optimization, we apply the appropriate hardware pipeline and parallel techniques within the DF and use polyphase filters instead of decimation process. In this new architecture, we used the transpose form (TF) filters to reduce critical path delay. We also use an Over Lapp Add (OLA) computation method in order to overcome the memory space and the waiting time for filling the data in the input FIFO register.

VII. CONCLUSION AND FUTURE WORK

We have developed and tested a DSP system for hearing impaired persons. It incorporates wide bandwidth and a great deal of flexibility in adjusting the overall speech processing algorithm.

The DSP48A1 slices have been integrated into FPGA for Digital Signal Processing purpose. They are assembled in columns using less wiring which reduces internal connections and avoid critical paths and hence time latency. The use of DSP48A1 has allowed a high performance of the system and about 20 % of the power dissipation is gained.

The principle objective of this paper was to present the new design approach of a low cost and reconfigurable FPGA platform. It can be used to test the DWT algorithms with different parameters as to meet the specifications for different hearing pathologies.

We are currently pursuing research to improve our current algorithms and architecture design for further noise reduction and echo cancellation. Knowing that the perception of speech is highly subjective in nature, the system is subjected to be tested on human subjects in the real world environment. The patient’s response will determine the success or failure of this DSP system.

REFERENCES

- [1] N. Magotra and S Sirivara “Real time digital speech processing strategies for the hearing impaired” Tewas Instruments, Application Report, USA April 2000
- [2] J. Pang and S. Chauhan “FPGA design of speech compression by using discrete wavelet transform” Proceedings of the World Congress on Engineering and Computer Science WCECS08 Oct. 22-24, USA 2008.
- [3] Z. J. Mou and P. Duhamel “Fast FIR Filtering: Algorithms and Implementation” Signal Processing Vol. 13 N° 4 pp. 377-384 Dec. 1987.
- [4] L. Bendaouia, SM. Karabernou, L. Kessal, H. Salhi and F. Ykhlef “Fast DWT based FPGA implementation for medical application” IEEE International Conference on Phealth, Lyon, France June 2010.

[5] S. Powell and P. Chan “Reduced complexity programmable FIR Filters” IEEE Int. Symposium on Circuits and Systems pp. 561-564 May. 1992

[6] All programmable FPGAs <http://www.xilinx.com> , [Retrieved:April, 2012]

[7] F. J. Taledo Moreo, A Leraz Cano, J. J. Martinez Alvarez, J. Martinez Alajarin and R. Ruiz Merino, “Compression system for the phonographic signal” Journal of sociocybernetics Vol. 7 N°2 pp. 770-773, winter 2009.

[8] J. Chilo and T. Lindblad “Hardware of 1D wavelet transform on an FPGA for infrasound signal classification” IEEE Transaction on Nuclear Science Vol. 55 issue 1 pp. 9-13, 2008

[9] Tim Erjavec, “Introducing the Xilinx targetd design platform”, [www.eetimes.com](http://www.eetimes.com) [Retrieved , February 2, 2009].

[10] K. Parki, “VLSI Architecture for discrete wavelet transform” IEEE Transaction VLSI Systems Vol. 1 pp. 191-202 June 1993.

[11] K.R. Borisagar, “Speech processing using wavelet transform and implementation for digital hearing aids” International conference on engineering trend, Pune, Dec. 2008.

[12] N.A. Ghamry and S.E. Habib, “An efficient FPGA Implementation of a wavelet Coder/Decoder” International conference on Microelectronics ICM 2000, Tahrn, October, 2000.

[13] R. Hourani, W Alexander and T. Raithatha “Automated design space exploration for DSP applications” Journal of Signal Processing Systems, Springer, 2009

[14] S. Chan, W. Liu and K. Ho “Multiplier less perfect reconstruction modulated filter banks with sum of powers of two coefficients” IEEE Signal Processing Letters Vol. 8 N° 6 pp. 163-166 June 2001.

[15] B. Cope, P.Y.K. Cheung and L. Howes “Performance comparison of graphics processors to reconfigurable logic: a case study” IEEE Transactions on Computers Vol. 59 N°4 April 2010

[16] “2D adaptive noise removal filtering”. [www.mathworks.com](http://www.mathworks.com) [Retrieved:May, 2012]

[17] T. Fillon “Traitement numérique du signal pour une aide aux malentendants” Thesis ENST France April, 2005.

[18] K. R. Rekha, B. S. Nagabishan and N. R. Natary, “FPGA implementation of NLMS algorithm for the identification of unknown system”, International journal of engineering, science and technology, Vol. 2 (11) 2010.

[19] R.C.D dePaiva, W.P. Biscainho and S.L. Netto, “On the application of RLS adaptive filtering for voice pitch modification” Proc. Of the 10<sup>th</sup> International conference on digital audio effects , Bordeaux France, Sept. 10-15 2007.

[20] J. Flanagan and M. Saslow “Speech analysis, synthesis and perception” Springer, New York 2<sup>nd</sup> edition 1972.

[21] G. Kartik, M. Kumar and M. Rahman, “Speech enhancement using gradient based variable size adaptive filtering techniques” IEEE International Journal of Computer Science & Emerging Technologies Vol. 2, issue 1, Feb. 2011

[22] S. G. Mallat, “A theory for multiresolution signal decomposition : the wavelet representation” IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 11 pp. 674-693, July 1989

[23] C. Chakrabarti, M. Vishuanath and R.M. Owens, “Architecture for wavelet transforms” VLSI Signal Proc. VI IEEE Special Publications pp. 507-515, 1993

[24] X. Hu, L. DeBrunner and V. DeBrunner, “An efficient design for FIR filters with variable precision” Proceeding IEEE International Symposium on Circuits and Systems Vol. 4 pp. 365-368 May 2002

[25] FPGA, DSP48A1 user Guide, [www.datasheets.org.uk/250-DSP](http://www.datasheets.org.uk/250-DSP) ,, [Retrieved:August 13, 2009]

[26] A. Baganne, I. Bennour, M. Elmarzougui, R. Gaiech and E. Martin, “A multi level design flow for incorporating IP cores: case study of 1D wavelet IP integration” in Design, Automation and Test in Europe Conference and Exhibition, pp. 250-255, 2003

[27] L. Bendaouia, SM. Karabernou, L. Kessal, . Salhi and F. Ykhlef, “DWT based FPGA implementation of a reconfigurable platform for a bio-inspired medical hearing aid” International Conference on Systems, Modeling and Design, Istanbul, Turkey Feb. 3<sup>rd</sup>-5<sup>th</sup> 2012

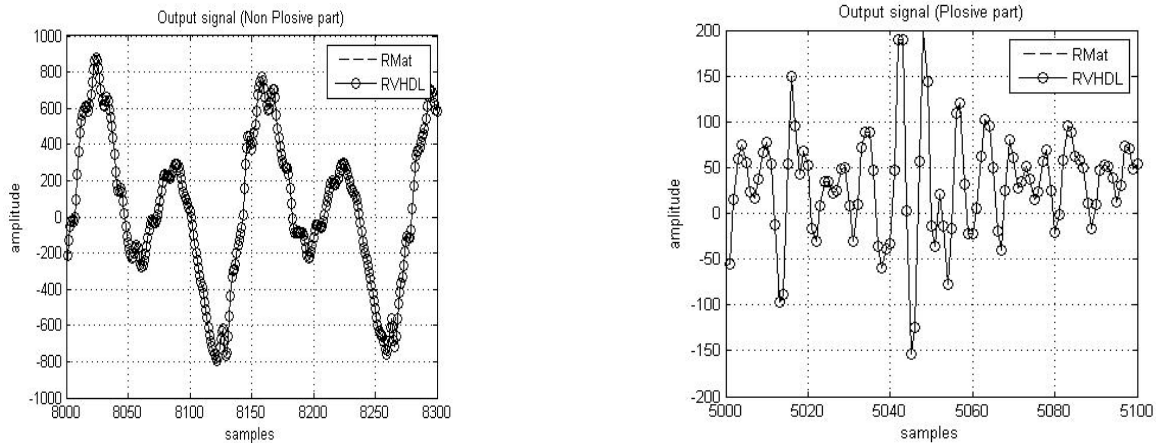


Figure 13. Matlab versus VHDL Output data analysis and comparison.