VLSI Implementation of Hybrid Algorithm Architecture for Speech Enhancement

Jigar Shah[1] and Satish Shah[2]

**[1] Electronics and Telecommunication Engineering Department, S.V.M. Institute of Technology
Bharuch, Gujarat 392001, India**

**[2] Electrical Engineering Department, Faculty of Technology and Engineering, The M.S. University of Baroda
Vadodara, Gujarat 390001, India**

Abstract

The speech enhancement techniques are required to improve the speech signal quality without causing any offshoot in many applications. Recently the growing use of cellular and mobile phones, hands free systems, VoIP phones, voice messaging service, call service centers etc. require efficient real time speech enhancement and detection strategies to make them superior over conventional speech communication systems. The speech enhancement algorithms are required to deal with additive noise and convolutive distortion that occur in any wireless communication system. Also the single channel (one microphone) signal is available in real environments. Hence a single channel hybrid algorithm is used which combines minimum mean square error-log spectral amplitude (MMSE-LSA) algorithm for additive noise removal and the relative spectral amplitude (RASTA) algorithm for reverberation cancellation. The real time and embedded implementation on directly available DSP platforms like TMS320C6713 shows some defects. Hence the VLSI implementation using semi-custom (e.g. FPGA) or full-custom approach is required. One such architecture is proposed in this paper.

***Keywords: Speech Enhancement, Additive Noise, MMSE Algorithm, RASTA Algorithm, Hybrid Algorithm*.**

1. Introduction

The main objective of the speech enhancement is to improve one or more perceptual aspects of the speech, such as overall quality, intelligibility, or extent of listener fatigue. Speech processing systems are usually designed for noise free environments, but in real world environment the presence of background noise is unavoidable. Speech enhancement is required not only for human listener but also it is required as a pre-processer in several speech processing systems such as speech coding, speech recognition, speaker recognition etc. Speech enhancement algorithms have been applied to problems of background noise removal and cancellation of reverberation in modern speech communication systems. Various speech enhancement methods have been proposed by researchers over the years. The limitations of these methods still pose a considerable challenge to the researchers in this area.

Most single channel speech enhancement techniques are based on transform domain approach. The short time discrete Fourier transform (STDFT) is used as transformation tool in most techniques used at present [1]. These methods are based on the analysis-modify-synthesis approach. They use fixed analysis window length (usually 20-25ms) and frame based processing. They are based on the fact that human speech perception is not sensitive to spectral phase but the clean spectral amplitude must be properly extracted from the noisy speech to have acceptable quality speech at output and hence they are called short time spectral amplitude or attenuation (STSA) based methods. The phase of noisy speech is preserved in the enhanced speech. The synthesis is mostly done using overlap-add method. They have been one of the well-known and well investigated techniques for additive noise reduction, have less computation complexity and easy implementations but always offer a musical and residual noise trade-off. The MMSE-LSA approach originally proposed by Ephrahim and Malah [2] now referred to as STSA-MMSE85 algorithm based on weighted noise estimation is employed in millions of 3G handsets as the one and only commercially available 3GPP-endorsed noise suppressor [3]. But still there are open questions like how the parameters of statistical models can be estimated in a robust fashion and what can be meaningful optimization criteria for speech enhancement; which will require further research.

The RelAtive SpecTral Amplitude (RASTA) processing algorithm originally proposed by Hermanskey and Morgan [4] to enhance speech for automatic speech recognition (ASR) in reverberant environment. The original RASTA approach was extended for speech enhancement by modifying the power spectral magnitude filter cutoff frequencies [5]. It is motivated by some perceptual properties of speech signal. Also the RASTA approach can be used to handle both additive noise and reverberation. The multiband RASTA algorithm was suggested in [7] for better speech enhancements. Research shows that the speech enhanced by RASTA has high residual noise during initial portion. Also the musical noise problem is not completely solved by using RASTA alone. The perceptual properties can be combined with MMSESTSA85 algorithm to improve the performance. One such algorithm is proposed in [6]. Also its real time and embedded implementation on hardware platform is performed and profile report for implementations are generated and compared in [8]. This paper first briefs the hybrid algorithm for speech enhancement and problem with its DSP implementation and then suggests the possible VLSI hardware implementation architecture.

2. The Hybrid Approach for Speech Enhancement

The hybrid approach proposed in [6] uses combination of MMSE STSA85 algorithm and multiband RASTA filter. The connection is not simple cascade but the blocks are interacting as shown in figure 1. Degraded speech signal $y(n)$ consists of clean speech signal $x(n)$ and additive noise $d(n)$ and it is modeled as

$$y(n)=x(n)+d(n) \tag{1}$$

In Frequency domain the model is given by equation

$$Y(K)=X(K)+D(K) \tag{2}$$

The noisy speech is presented simultaneously to both multiband RASTA and MMSE STSA85 algorithms. The VAD is required to estimate speech/silence segment for MMSE STSA85 algorithm. This block is responsible for malfunctioning of algorithm if the detection is false. The MMSE STSA85 algorithm is highly dependent of VAD false rate. So VAD is not directly getting the noisy speech for estimation but the output of multiband RASTA filter is given to VAD for estimation. Some speech distortion and musical and residual noise remain in enhanced speech by RASTA algorithm. However, the VAD can now better detect the speech/silence segment compared to direct detection from noisy speech. But the white noise after RASTA filtering gets converted into colored noise with sharp spectral peaks. Hence, the accuracy in noise estimation reduces; this causes the rise in musical noise. So the noise power is estimated for RASTA filtered as well as original noisy speech spectrum. The ratio (λ) of original noise power to the filtered noise power is calculated and it is used to calculate *a priori* SNR $\xi(K)$ from *a posteriori* SNR $\gamma(K)$ for frame *t*. A mild linear compression is required to avoid over suppression. The modified decision direct rule taking this factor into consideration is given by

$$\xi^{(t)}(K)=\eta\frac{\left|\hat{X}^{(t\text{-}1)}(K)\right|^2}{\left|\hat{D}^{(t)}(K)\right|^2/\lambda}+(1-\eta)\max(\bar{\gamma}^{(t)}(K)-1,0);$$

$$\bar{\gamma}(K)=\frac{\left|\bar{Y}(K)\right|^2}{\left|\bar{\hat{D}}(K)\right|^2/\lambda} \tag{3}$$

The smoothing parameter η controls the trade-off between speech distortion and residual noise and it is usually set to 0.98 for optimum performance [2].

3. SIMULINK and DSP Implementations

Since the complete implementation of the hybrid algorithm has a great computational complexity, it is necessary to test the possibility of implementing it in a real time and embedded environment. The real time implementation on PC using SIMULINK and on TMS320C6713 DSP using DSP Starter Kit DSK 6713 is reported in [8]. The results of the profiler report obtained for SIMULINK and DSK6713 implementation of the model are briefed in table 1. Looking at the results the hybrid algorithm block occupies only 12.4% of total execution time when running on PC; while 284% average time of the base sample time when running on DSK6713. This is the constraint for the DSK6713 implementation of the same model which has no problem at all when runs on PC. The algorithm needs some optimizations before its implementation on DSK6713.

Table 1: Profile results comparison

Function/Block	*PC Implementation*	*DSP Implementation*
CPU clock speed	2166MHz	225MHz
	Average Execution Time	
Input	0.3%	0.78%
Data buffering/windowing	0.8%	6.7%
Hybrid algorithm (Main loop)	12.4%	284%
Overlap-add	0.3%	7.04%
Output	0.3%	0.34%

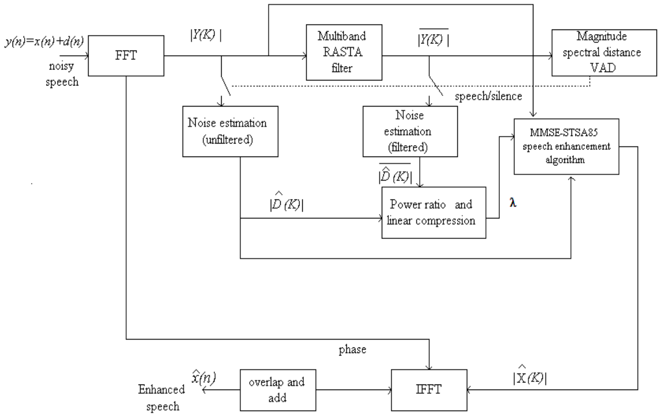
$y(n)=x(n)+d(n)$
noisy speech
FFT
$|Y(K)|$
Multiband RASTA filter
$\overline{|Y(K)|}$
Magnitude spectral distance VAD
speech/silence
Noise estimation (unfiltered)
Noise estimation (filtered)
MMSE-STSA85 speech enhancement algorithm
$\overline{|\hat{D}(K)|}$
$|\hat{D}(K)|$
Power ratio and linear compression
λ
phase
Enhanced speech
$\hat{x}(n)$
overlap and add
IFFT
$|\hat{X}(K)|$

Fig. 1 Block diagram of hybrid algorithm.

4. Proposed VLSI Implementation Architecture

Figure 2 shows the block diagram of hardware architecture for real time implementation of hybrid algorithm. The ASIC/FPGA block is programmed to implement the heart of the algorithm [9]. It is shown by means of various flow charts in figure 3 and figure 4 for different parts of the algorithm. The flow charts can be implemented by writing software code in VHDL/Verilog or IP core of the block can be used.

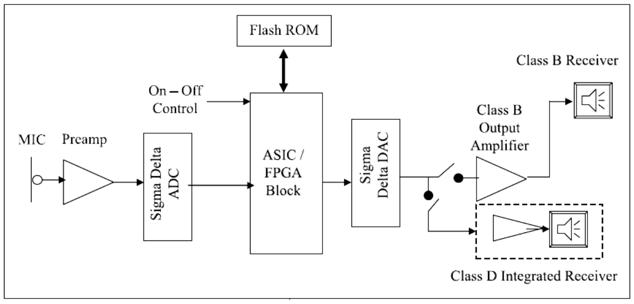

Fig. 2 Hardware architecture for real time implementation of hybrid algorithm.

Figure 3 presents a complete flow diagram to implement the algorithm on hardware platform. The voice activity detector (VAD) routine implementation is shown in figure 4.

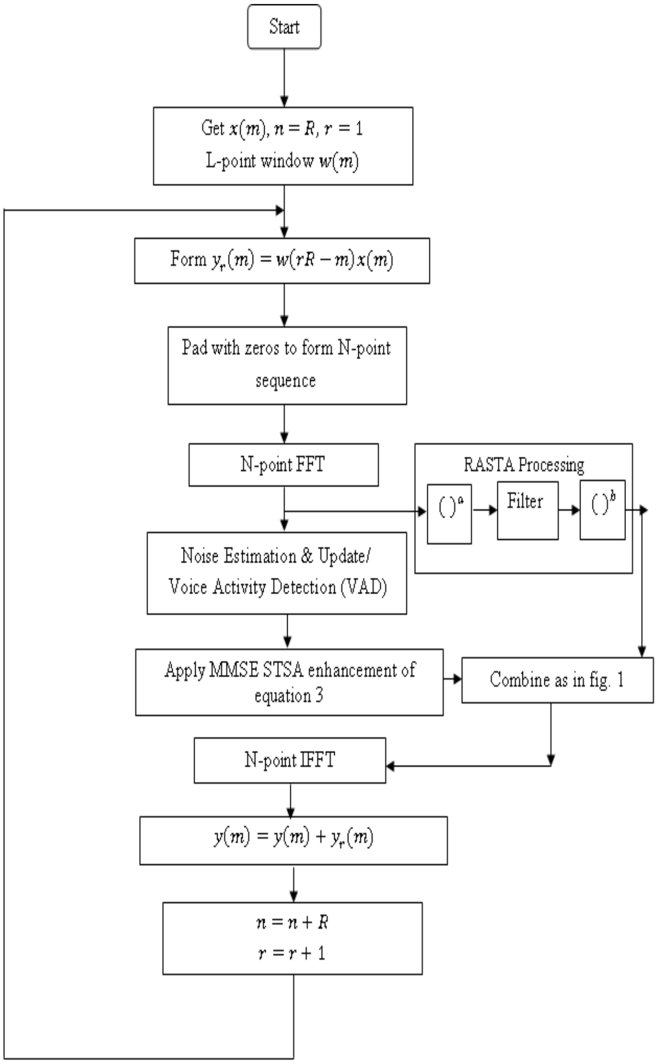
Start
Get $x(m)$, $n = R$, $r = 1$
L-point window $w(m)$
Form $y_r(m) = w(rR - m)x(m)$
Pad with zeros to form N-point sequence
N-point FFT
RASTA Processing
$()^a$
Filter
$()^b$
Noise Estimation & Update/ Voice Activity Detection (VAD)
Apply MMSE STSA enhancement of equation 3
Combine as in fig. 1
N-point IFFT
$y(m) = y(m) + y_r(m)$
$n = n + R$
$r = r + 1$

Fig. 3 A flow chart showing the implementation steps on FPGA/ASIC.

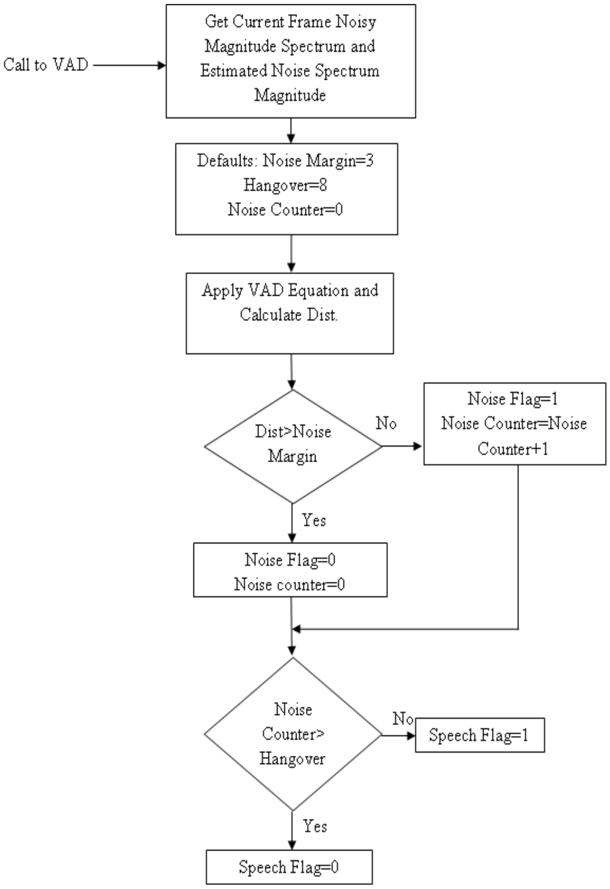

Fig. 4 A flow chart showing the implementation of VAD (part of flow chart in figure 3).

5. Conclusions and Future Scopes

The combination of STSA and RASTA approach is termed here as hybrid approach used to improve the performance at lower SNRs (0-5dB). The VLSI implementation architecture is suggested here. Future work can explore some of the research directions pointed out here.

- Further optimization of the code can be done through the algorithm tuning process [10].
- The hybrid approach can be optimized by merging MMSE and RASTA algorithm together. Also the RASTA filters can be redesigned with better specifications.
- Also, due to high complexity the algorithm can be implemented using soft computing techniques like fuzzy logic, neural network and genetic algorithms.

The investigation of these implications is a valuable topic for future research and might yield substantial improvements.

References

[1] Pavel Sinha, Speech Enhancement: Algorithm and Architecture, VDM Verlang Dr. Muller Aktiengesellschaft & Co. Germany, 2008.

[2] Y. Ephrahim and D.Malah, "Speech Enhancement using a Minimum Mean Square Error Log Spectral Amplitude Estimator," IEEE Trans. on Acoustics, Speech and Signal Processing, vol. ASSP-33, no. 2, pp. 443-445, April 1985.

[3] 3GPP2 specifications: http://www.3gpp2.org/Public_html/specs/index. cfm, 2007.

[4] Hynek Hermanskey and Nelson Morgan "RASTA Processing of Speech," IEEE Trans. on Acoustics, Speech and Signal Processing, vol.2, pp. 578-589, Oct.1994.

[5] H. Hermanskey, E.A. Wan and C.Avendano, "Noise Suppression in Cellular Communications," 2nd IEEE workshop on Interactive Voice Technology for Telecommunications Applications (IVTT 94), Kyoto, Japan, Sept. 1994.

[6] S.K.Shah, J.H.Shah, and N.N.Parmar, "Objective Evaluation of STSA Based Speech Enhancement Techniques for Speech Communication Systems with Proposed Modifications," Proc. International Conference on Advances in Communication, Network and Computing (CNC 2010), IEEE Computer Society, pp. 19-23, Oct. 2010, doi: 10.1109/CNC.2010.13.

[7] S.K.Shah, J.H.Shah, and N.N.Parmar, "Evaluation of RASTA Approach with Modified Parameters for Speech Enhancement in Communicaiton Systems," Proc. IEEE Symposium on Computers and Informatics (ISCI 2011), pp. 159-162, March 2011, doi: 10.1109/ISCI.2011.5958902.

[8] S.K.Shah and J.H.Shah, "Real Time and Embedded Implementation of Hybrid Algorithm for Speech Enhancement," Proc. IEEE World Congress on Information and Communication Technologies (WICT 2011), pp. 341-345, Dec. 2011, doi: 10.1109/WICT.2011.6141269.

[9] S.M. Kuo, B.H. Lee, and W. Tian, Real Time Digital Signal Processing: Implementations and Applications, 2nd Ed., John Wiley & Sons Ltd., West Susex, England, 2006.

[10]Jigar Shah and Kruti Dangarwala, "C Implementation & comparison of companding & silence audio compression techniques," International Journal of Computer Science Issues (IJCSI), pp. 26-30, Vol. 7, Issue 2, No. 3, March 2010.

Prof. Jigar H. Shah had obtained B.E. (Electronics) degree in 1997 and M.E. (Microprocessor Systems and Applications) in 2006. Presently he is pursuing Ph.D. degree from the M.S. University of Baroda, Vadodara under the guidance of Prof. Satish K. Shah. He is currently employed with SVMIT, Bharuch, Gujarat State, India as an associate professor. He has published two research papers in international journals, five research papers in various international conferences and five research papers in national conferences. He has also five book titles viz. Digital Signal Processing, Basic Electronics, Advance Electronics, Digital Logic Design and Engineering Physics. His current research interests are in the area of Speech Enhancement, Digital Signal Processing, and Embedded and VLSI systems. He is also life member of ISTE and IETE.

Prof. Satish K. Shah had obtained B.Sc. (Physics) degree in 1970 and M.E. (Automatic Control Engineering) in 1972. He is currently employed with Faculty of Technology and Engineering, The M.S. University of Baroda, Vadodara, Gujarat State, India as a professor and head. He is registered as a Ph.D. supervisor in several universities. He has published several research papers in international journals and in various international and national conferences. He has also two book titles viz. 8051 Microcontrollers: MCS Family and its variants, Embedded DSP System Design and Implementation using TI Digital Signal Processors. His current research interests are in the area of Control System Engineering, Digital Image & Signal Processing, Embedded Controllers, MIMO Systems, and Wireless Networking & Communication. He is also life member of ISTE, IETE, IE(I), IEEE and ISA.