Performance Enhancement on Voice using VAD Algorithm and Cepstral Analysis

[1]T. Ravichandran and [2]K. Durai Samy
[1]Velalar College of Engineering and Technology, Thindal, Erode, Tamilnadu, India
[2]K.S.R.C.T, Tiruchengode, Erode, Tamilnadu, India

Abstract: The Transport of Voice Over Network has been existing for a long time. Due to the bad quality of speech and absence of useful service, it has not yet been widely spread. The benefits of reduced cost and bandwidth savings of carrying voice over data network is associated with Quality of Service (QoS). Many different techniques and protocols are used to improve the quality of service in voice over data network. We review our data on the voice quality effects of background noise and the compression option of suppressing transmission during silence. Because the main problem of echo has emerged repeatedly in the VOIP environment, we review this issue in voice over data communication network is used to VAD algorithms and cepstral analysis of VOIP network performance.

Key words: Voice over IP, quality of service, voice activity detector, peak finding algorithm

INTRODUCTION

Companies and organizations around the world are willing to reduce rising communication costs. The consolidation of separate voice and data offers an opportunity for significant reduction in communication costs. Since, now days data traffic is growing much faster than telephone traffic, a need has been identified to transport voice over data networks, instead of the transmission of data over voice networks.

The voice over data network is brought about the rise for Voice Over IP (VOIP).VoIP has become especially attractive given the low-cost, flat rate pricing of the public Internet. Many components have to be designed to accommodate voice over data network, such as the access gateways that link the data and the telephonic network among others. Applications that offer Voice over IP services will have to include a comprehensive technology that reduces the impairments caused by sending voice over data network that were not designed to handle it. An important factor to be considered by network designers is the problem of Quality of Service (QoS). This is because IP is the best effort service but the existing system provides no guarantees on delivery and data integrity. Voice processing causes delays, jitter, noise and cancel echoes that will be introduced from the telephonic side. It also have to include an appropriate algorithm to mask the gaps (silence) caused by dropped packets due to congestion on the network.

Time domain VAD algorithms: Voice activity detection is important in many areas of voice signal processing such as voice coding, voice recognition, voice enhancement etc. An effective voice activity detection (VAD) algorithm is proposed for improving voice quality performance in noisy environments[1]. The VAD algorithm is trained for a small period by a prerecorded sample that contains only background noise and silence gap. The initial threshold level for various parameters is computed from these samples. The threshold value can be fixed on the different VAD algorithm technique used in the voice signal. A threshold is applied to the extracted parameter in order to divide the speech signal between voice and non-voice segments[2].

In the early VAD algorithms, short-time energy, zero crossing rate, formant shape, least-square periodicity measure and multiple statistical models are some of the recent ideas in VAD designs.

In this work, we consider two recently proposed VAD algorithms. These include the VAD used in the Eigen Value Based Detector (EVD), the VAD used in the enhanced Median Value Based Detector (MVD).

In the Eigen Value Based Detector (EVD) and Median Value Based Detector (MVD), an adaptive noise suppressor filter is used to filter the input signal frame. The coefficients of the filter are computed during noise only periods. The energy of the filtered signal is compared to a noise dependent threshold. As both the filter coefficients and the threshold are computed during noise only frames, special measures are taken to identify noise frames. These include both signal stationarity and periodicity tests.

To improve the performance of the Eigen Value Based Detector (EVD) and Median Value Based Detector (MVD) VAD for both stationary and non stationary noise, ravichandran and duraiswamy[3] proposed several new features to the basic VAD design. These include several VAD algorithm results are compared and analysis of improvement to performance in quality of speech.

Corresponding Author: T. Ravichandran, Research scholar, Velalar College of Engineering and Technology, Thindal, Erode, Tamilnadu, India

We have implemented the Eigen Value Based Detector (EVD) and Median Value Based Detector (MVD) VAD used to tested their performance for different voice, and noise environment. MATLAB was used to test these algorithm developed on various samples. The test templates used varied in loudness, speech continuity and background noise. Both male and female voice have been used.

Comparisons of the aforementioned VAD algorithms of Energy Based Detector (EBD) Zero Crossing Detector (ZCD) and new proposal algorithm Eigen Value Based Detector (EVD) and Median Value Based Detector (MVD) performances are presented. The performance of algorithms was studied on the basis of the parameters in 1.speech quality (Mean Opinion Score - MOS) 2.Suppression ratio 3. % Misdetection.
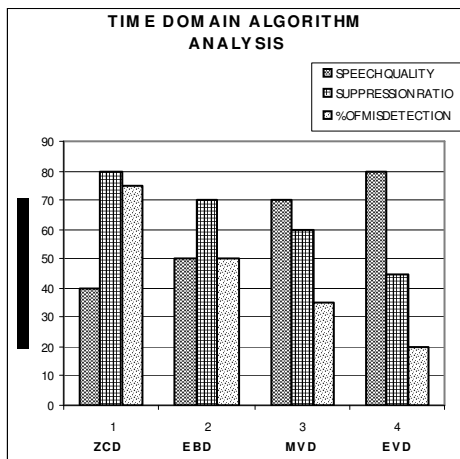

Fig. 1:   Comparison of performance analysis

Results of the comparison charts are shown in Fig. 1. We have concluded to performance of all aspects in the testing simulation for our algorithm of EVD and MVD is more suitable compare with existing in real time applications.

Frequency domain VAD algorithms: This algorithm takes its decisions based on energy comparisons of the signal frame with a reference energy threshold in the frequency domain. Srinivasan and Gersho[4] proposed several new features to the frequency VAD design. These include a multi band (4 bands) energy comparison, spectral flatness measurement and using the fraction of the energy of the low frequency band. This same method is follow R.V.prasad and A.sangwan[1] and small change in the algorithm modification used to implemented real time speech transmission on the internet. In the inclusive VAD Algorithm is used to filter the noise and suppression in silence for the input signal frame, the basic idea of VAD algorithm in this inclusive VAD algorithm is combination or earlier median value based Detector (MVD) and eigen value based Detector(EVD).

This VAD algorithm is capable of removing white noise as well as frequency selective nose and maintaining a good quality of speech. Inclusive VAD algorithm works with multiband energy comparison and passes median value based Detector and eigen value based Detector to test computation. The results are taken in each band energy calculated and It filter the noise double time the silence or suppression ratio.
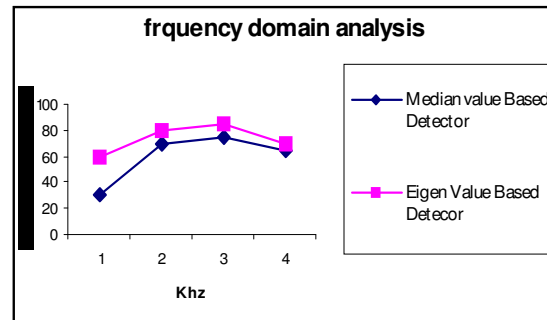

Fig. 2:   Performance of each band quality

MATLAB was used to analysis or each band energy in EVD and MVD results are shown in Fig. 2. Although the quality of speech is better compared to all other previous algorithms, it's performance is noise ratio and high computation complexity.

Cepstral based pack finding algorithm: Cepstral analysis is a mature fool developed for speech analysis and recognition. Xinli[2] investigated cepstral analysis to consist of three consecutive steps (refer to flow chart) 1.FFT(Fast Fourier Transform),2.Take logarithm and 3. IFFT (Inverse Fast Fourier Transform).
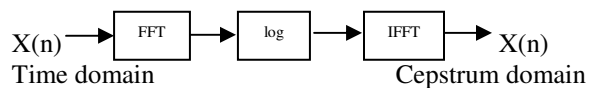
It is easy to see that those three operations are all leaner and we can exactly recover the original signal in time domain from its cepstrum domain representation by taking correspondent inverse operations. The large cepstrum coefficients around the center contain important envelope information and small coefficients on both sides consist of fine detail pitch information. In this condition, We proposed to finding the peak value of the cepstrum is calculated and set through the threshold. The peak values to find each frame to measure the concept of Euclidean distance to developed in Peak Finding Algorithm[PFA].

To improve the quality of speech to finding peak values are activated and others are inactivated for this concepts of VAD result is shown Fig. 3. This method is more suitable for real time application.
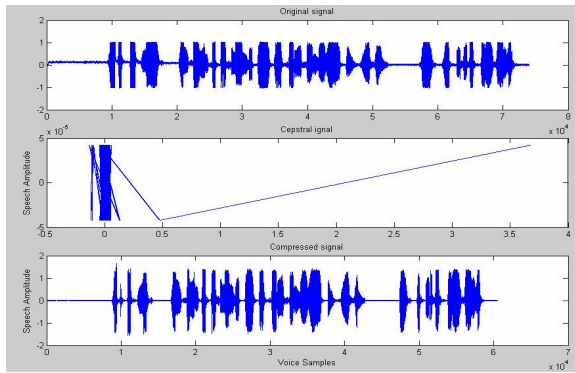
Fig. 3:   Quality speech performance

Peak LMS algorithms: Acoustic echoes cause great discomfort to the users since their own speech (delayed version) is heard. To solve the problem of acoustic echoes, an acoustic echo canceller is proposed. The echo canceller estimates the impulse response of the echo path and generates a replica of the echo. Following that, the estimated echo is subtracted from the received signal. The objective is to eliminate the sound (through loudspeaker) from the far end speaker being transmitted again to him or her through the microphone, hence the term echo cancellation[1]. The obvious choice to implement such canceller is by employing adaptive filter which operates satisfactorily in an unknown environment and has the ability to track time variations of input statistics.

Time domain adaptive filtering: In general various algorithms are used in practice to train the adaptive filters. Essentially the performance of an echo canceller depends on the selection of the adaptive filter structure and algorithm for the adaptation[5]. In other words, the adaptive filter structure and the algorithm used determine the accuracy in estimating the echo path and the speed to adapt to its variation. The adaptive algorithm on the other hand, adjusts the weight coefficients in the filter to minimize the error e (n).
The various time-domain algorithms used in this paper are,
1.   Least Mean Square (LMS)
2.   Peak Finding LMS (PLMS)
     The best known algorithm for adaptive filtering in time domain is the Least Mean Square (LMS) algorithm and peak finding LMS algorithm. The main properties of echo is sited in top of the speech signal. In this condition our proposal to removal of echo is used to finding peak value. The top portion of echo peak is identify to remove and use through LMS algorithm. Due to its simple structure and low computational complexity compared to other time domain algorithms, it has become very popular. Its convergence behavior however, depends strongly on the input signal and correlation.
     By normalization of the update equation of LMS we obtain the Peak Finding LMS (PLMS) algorithm.

The PLMS convergence behavior no longer depends on the input signal variance. Both the LMS and PLMS algorithm make one update every sample interval.

Frequency domain adaptive filtering: The various frequency-domain algorithms used in this paper are,
1.   Fast Peak Block LMS (FPBLMS).
2.   Partitioned Peak Block FLMS (PPBFLMS).
     In the time domain LMS algorithm, the tap weights of a finite-duration impulse (FIR) filter are adapted by taking every sample. Recognizing that the *Fourier transforms* maps time-domain signals into the frequency domain and that the inverse Fourier transform provides the inverse mapping back into the time domain, it is equally feasible to perform the adaptation of filter parameters in the frequency domain. In such a case, the adaptive filter is referred as frequency-*domain adaptive filtering (FDAF)*[1].
     In Frequency domain adaptive filters the signal is processed by splitting it into blocks such that each block contains a definite number of samples. By doing this the complexity of the algorithm reduces and the convergence rate increases. In a peak block-adaptive filter, depicted in Fig. 4, the incoming data sequence $u(n)$ is sectioned into L-point blocks by means of a serial-to-parallel converter and the blocks of input data so produced are applied to an FIR filter of length M, one block at a time. The tap weights of the filter are updated after the collection of each block of data samples, so that adaptation of the filter proceeds on a block-by-block basis rather than on a sample-by-sample basis as in t he conventional LMS algorithm.
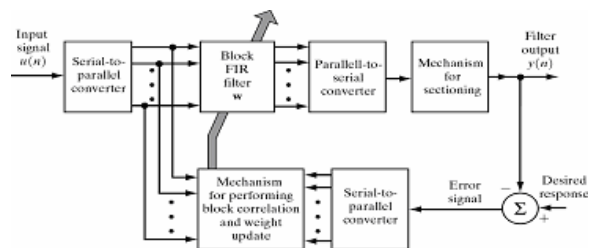

Fig. 4: Peak block-adaptive filter

     The two main operations to implement a frequency domain adaptive algorithm are a (linear) convolution, to perform the filtering of the input signal with the adaptive weights and a (linear) correlation, to calculate an estimate of the gradient that is needed for the update of adaptive weights. Overlap-save is a well known technique to convolve an infinite length input sequence with a finite length impulse response.
     In Partitioned peak Block Frequency Domain LMS algorithm the filter coefficients are partitioned into blocks with block length equal to the filter length and then processed in the same way as the FPBLMS algorithm. This particular algorithm overcomes the disadvantages of FBLMS algorithm.

Table 1:   Comparison table for different Adaptive filtering algorithms

| Parameters	Time domain adaptive filters		Frequency domain adaptive filters	
	LMS	PLMS	FPBLMS	PPBFLMS
Complexity (no. of multiplications required per input sample)	2049	2051	126	347
Convergence rate	Slow	Slow for speech input	Fast	Fast
Maximum ERLE obtained (dB)	30	34	29	39

Desirable properties for the algorithms: The performance of the different adaptive algorithms is primarily measured by the following properties:

* Misadjustment: The misadjustment of the algorithm concerns how much the parameters differ from their true values.
* Computational requirements: This concern the number of arithmetic operations required to update each filter parameter. Mostly divisions and multiplications are counted since these take much longer time to perform than additions or subtractions.
* Convergence rate: This concerns how fast the algorithm will change the filter parameters to their final values.
* Tracking: Tracking concerns how the algorithm will respond to changes in the true parameters.
* Robustness: Small disturbances such as noise should not result in large modifications of the filter parameters.
* Numerical properties: The algorithms should be numerically stable in the sense that they should not be sensitive to quantization errors. These quantization errors occur when the algorithms are implemented on digital computers which always have a finite word-length.
* Stability: An algorithm is said to be stable if the mean-squared error converges to a final (finite) value.

`The different time-domain and frequency-domain algorithms are simulated using MATLAB and the results are shown in Table 1. The algorithms are tested by taking a telephone conversation which contains both Near End Speech (NES) and a Far End Speech (FES) samples as input. This signal is referred to as microphone signal. The FES is filtered and it is given as input to adaptive filter which after several iterations outputs the estimate of the far end speech. This output is then subtracted from the microphone signal so that the desirable NES alone reaches the receiver thereby suppressing the echo.
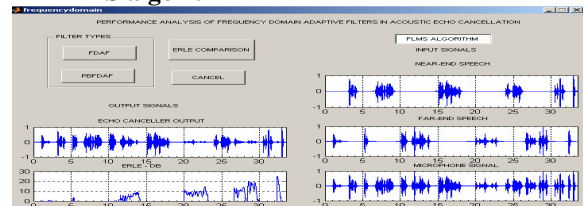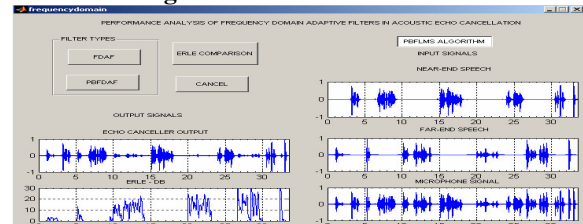
Simulation of time domain algorithms
LMS slgorithm
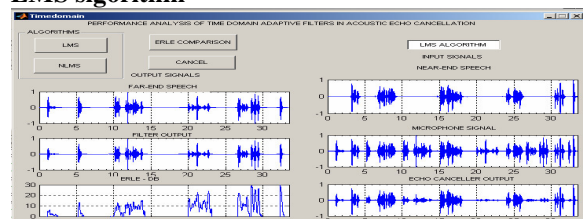
PLMS algorithm
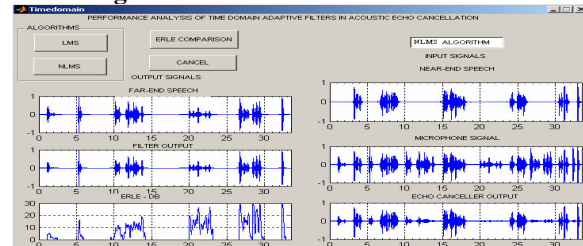
Simulation of frequency domain algorithms
FPBLMS algorithm

PPBFLMS algorithm

Security for voice communications: Basically STEGANOGRAPHY is the method of information or data hiding into another voice file (generally called as COVER) without any notable change or damage to the original content of the cover file[4]. Steganography provides a secured and secret mean of communication. In our proposal of implement the stegnography concept the voice file should change the cepstral form used in the container file for the secret messages transmission. Thus the secret messages can be hidden into the used cepstral voice container file developed through Even Bit Swapping Algorithm[EBSA].

This main application area of current copyright marking proposals, lies in digital representations of analogue objects such as audio, still pictures, video and multimedia generally. Here there is considerable scope for embedding data by introducing various kinds of error. Many writers have proposed embedding the data in the least significant bits[6]. An obviously better technique, which has occurred independently to many writees, is to embed the data into the least significant

bits of pseudo-randomly chosen pixels or sound samples. In this way, the key for the pseudo-random sequence generator becomes the stego-key for the system and Kerckhoffs' principle is observed. In this condition our proposal for voice container file converted into cepstral form is used to EBS Algorithm in message hiding.

More care should be taken while appending the data into the voice file, since it should not affect the data part of the voice file. The Encrypted character from the first stage is given as an X-ORing the Digital key (i.e example: Digital key Expression: a+(b*c)) provided with this input. The output of this stage is obtained. Replacing the 8th bit read from input voice file to the encrypted message bit. The bit pattern of message file is compared with the bit pattern of key code generated using digital key. If the two bit pattern matches with each other, bit 0 is generated else if it doesn't match bit 1 is generated.

Our proposal is hiding information bits of even values bit inter change this modification form use in EBSA. This bit pattern is stegged in the voice file at the LSB of each character in the voice file. The encrypted character of the message file is converted to binary and each bit is placed in the LSB of a character. So encrypted character will be stored in the LSB's of 8 successive characters of the voice file. This modified binary pattern of the voice file is stored in the stegged file.

The decrypted message file is a bit stream (string) which is X-ORed with the digital key. The bit stream from the above process is again decrypted using the same Even Bit Swapping algorithm. The final bit stream is then converted into decimal to obtain the decrypted character, which is stored in the output file.
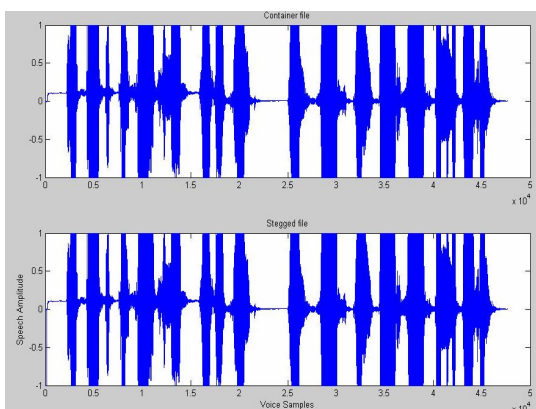

Fig. 5: Comparison of information voice container file and voice container file

This process is mainly used for data security in networks. Since this method is using voice file, the hackers motive is less as he may not be able to predict that the voice file contains secret message. Since it is highly secure, it is more suitable for protected voice signal communication

CONCLUSION

VOIP has become a reality now. It has become very much essential as the existing system does not provide satisfactory results or provides reliability. A practical solution could be to use the efficient VAD schemes for voice communication. Time domain and Frequency Domain VAD algorithms are found to be computationally less complex. With these schemes, good speech detection and silence were observed. MATLAB was used to test the algorithms developed on various sample signals and the Quality of the samples is rated on MOS(Mean Opinion Score) a scale of 1 (poorest) to 5 (best) Where 4 represents acceptable grade quality. The input signal has been taken to have speech quality 5. The speech samples after suppression are played to independent juries in a random manner for an unbiased decision. The number of frames, which has speech content, but has been classified as ACTIVE and number of frames without speech content but classified as INACTIVE, are counted. The ratio of this count to the total number of frames in the sample is explored as a percentage of MISDETECTION.

The thesis also asses the behaviors of the various VAD algorithm's accuracy for the VAD decisions. Time and Frequency domain VAD algorithm methods are used to improve voice quality, Silence suppression ratio of the voice network through 50 - 60 % of band width savings and also the improvement of bandwidth saving in the communication network by delay reduction. This thesis is implemented for algorithm to be develop VAD algorithm concepts for median value based detector and eigan value based detector using time & frequency domain methods are suitable in real time application and verified through comparesion chart in our results.

Cepstral analysis has been optimized in the context of voice recognition and provides an excellent model of the slow but meaningful variations in voice. The performance of cepstral based Peak Finding Algorithm is used in VAD algorithm concepts to find the peak value of the each sample. The detected lower values of peak value are eliminated with high degree of independence to levels of background noise and successful voice can be achieved via threshold cepstral analysis. This method is used to improve elimination of background noise in the acceptable voice quality.

In the hidden information using Even Bit Swapping Algorithm, the most commonly used technique is the removal of least significant bit of the original data of the voice container file. The concept of digital key function method is used with the help of the LSB technique for the double encryption and decryption. This provides greater security for the contents of the message file. Since it is using voice file, the hacking attempts are reduced. This process is mainly used for data security in voice network improvement on voice communication system.

The echo is an important factor of the quality of communication. The removal of echo is used to calculate the peak value of the LMS algorithm. The echo signals are sitting at the every top of the voice signals. In our algorithm to detect the peak value of echo's voice signal to calculated average value of the peak taken in the threshold. This threshold value is used to detect the upper in threshold values are eliminated and lower value of the threshold out is reduced echo. This method is effectively experimented to improve the voice quality in voice communication.

The results are the outcome of the existing results, the voice communication system is adapted to this improvement in real time communication system. These methods are new and progressing very fast. Hence, many other voice quality problems are open for further research.

REFERENCES

1. Sangwan, A., M.C. Chiranth, R. Shah, V. Gaurav and R.V. Prasad, 2002. Comparison voice activity detection algorithms for VOIP. IEEE Proc. Seventh Intl. Symp. Computers and Communications.

2. Li, X. and H.H. Yu, 2000. Transparent and robust audio data hiding in cepstrum domain. IEEE Intl. Conf. Multimedia and Expo, pp: 397-400.

3. Markopoulou, A.P., F.A. Tobagi and M.J. Karam, 2002. Assessment of VOIP quality over internet backbones. Proc. Conf. Computer Communications (IEEE INFOCOM), New York.

4. Kida, Y. and T. Kawahara, 2004. Voice activity detection based on optimally weighted combination of multiple features. Kyoto 606-8501.

5. Ravichandran, T. and K. Duraiswamy, 2005. Implementation of quality of service to a voice over internet protocol networks. Intl. Conf. ICIS, Malaysia.

6. Srinivasan, K. and A. Gersho, 1993. Voice activity detection for cellular network. Proc. of the IEEE Speech Coding Workshop, pp: 85-86.