A Novel Method to Modify VAD used in ITU-T G.729B for Low SNRs

H. Farsi, M.A. Mozaffarian, and H. Rahmani

***Abstract*—**Voice Activity Detection (VAD) Systems have various standards. One of the famous and practical standards is ITU-U G.729B. In G.729B, VAD decision, compare with other standards and new methods, has poor performance for voice frames especially in low Signal-to-Noise Ratios (SNRs). However, since this standard has good properties that we try to modify its low performance in adverse environments. One point that we focus on this standard is trying to have minimum changes. Therefore we use some new methods without major modification on basic structure. In this methodology we use some parameters in the first part and connect it to the main block.

We compare the proposed method, using objective parameters, with basic version of G.729B standard and ETSI AMR option 1 and 2. This comparison study in various types of noises like Gaussian, Vehicle and Babble noise.

***Keywords*—** Geometrically Adaptive Energy Threshold, LPFing method, Periodicity Estimation, True Envelope LPC, Voice Activity Detection.

I. Introduction

As is well-known, voice activity detection (VAD) achieves silence compression, which is important in both fixed and mobile telecommunication systems [1]. In communication systems based on variable bit rate speech coders, it represents the most important block, reducing the average bit rate; in a cellular radio system using the discontinuous transmission (DTX) mode, a VAD is able to increase the number of users and power consumption in portable equipment. Unfortunately, a VAD is far from efficient, especially when it is operating in adverse acoustic conditions.

While previous work in the area of speech analysis, such as detection, voicing classification or pitch estimation, have attempted to exploit some of the observed features of the statistics of speech signals, little has been done in providing an analytical framework for using these cumulates: in [15], a voiced/unvoiced detector using the bispectrum is developed and based on the observation that unvoiced phonemes are produced by a Gaussian-like excitation and thus result in a small bispectrum whereas the same is not true for voiced phonemes.

In [16] a method based on Gaussian tests for the bispectrum and the triple correlation is used to discriminate voiced and unvoiced segments.

The method exploits the Gaussian blindness of statistics but not the peculiarities of the statistics of voiced speech to better classify the segments.

In [17], the normalized skewness and kurtosis of short-term speech segments are used to detect transitional speech events (termed innovation), based on the observation that these two statistics take on nonzero values at the boundaries of speech segments, but no analytical ground is given to support the results.

In [18] a pitch estimation method based on the periodicity of the diagonal slice of the third-order cumulant is described and yields more reliable pitch estimates than the autocorrelation, but the claim of the third-order Cumulant slice having similar periodicity as the underlying speech is not clearly demonstrated.

Earlier algorithms are based on the Itakura LPC distance measure [l9], energy levels, timing, pitch and zero crossings rates [20], and periodicity measure [21].

Haigh [22] developed an algorithm using cepstral features, and Yoma, McInnes and Jack used adaptive noise modeling where they assumed the noise to be reasonably stationary and correlated [23].

In parallel, those algorithms are tested on specific applications like the Pan-European digital cellular mobile telephone service [24], cellular networks [25], digital cordless telephone systems [26], and structured noise environments [27]. Most recently, El-Maleh and Kabal compared various detection algorithms for wireless personal communications systems [28].

Unfortunately, the present speech detection algorithms have problems in low SNRs and also in nonstationary noise environments.

Consistent accuracy cannot be achieved since most algorithms rely on a threshold level for comparison. This threshold level is assumed to be fixed [29] or calculated in the non-speech intervals.

For example, in the autoregressive analysis with the LMS algorithm, non-speech intervals are required to train the FIR filters used [30]. Similarly, third order statistics-based VAD initially requires noise-only frames [31]. When the background noise is nonstationary and when the speech signal is dominated by voiced segments, the optimum threshold value should be monitored in each time frame to achieve reliable detection. To obtain that, methods which can effectively adapt to the changing background noise are required.

In order to evaluate the impact of background noise on recent voice activity detectors, this paper presents a performance evaluation and comparison of recent ITU-T and ETSI VAD algorithms. The latest ITU-T VAD standard is Rec.

G.729 Annex B [2], developed for fixed telephony and multimedia communications. This method provides a poor performance, in noisy environments especially for non-stationary noise or low Signal-to-Noise Ratios (SNRs), compare with other standards (e.g., ETSI AMR option 1 and 2 [3]) and new methods (e.g., statistical or estimation methods [10], [11]).

In this paper, we modify G.729 method by minimum change in its structure and compare its performance with other standards by objective measures.

II. G.729B VAD Standard

As an extension to the G.729 speech coder, ITU-T SG16 released G.729 Annex B in order to support DTX by means of VAD, CNI, and CNG. G.729B conducts a VAD decision every frame of 10 ms, using four different parameters:

1) differential power in the 0–1 kHz band (ΔE_l);
2) differential power over the whole band (ΔE_f);
3) differential zero crossing rate (ΔZC);
4) spectral distortion (ΔLSF).

where E_f, E_l, LSF_i, and ZC are the full-band energy, low-band energy, i^{th} line spectral frequency, and zero-crossing rate of the input signal. E_f, E_l, LSF_i, and ZC are the noise characterizing parameters updated using the background noise.

The block diagram of G.729B VAD is shown in Fig. 1. The input parameters for the VAD can be obtained from the input signal or from the intermediate values of the speech encoder. Subsequently, the difference parameters, ΔE_f, ΔE_l, ΔLSF, and ΔZC, are computed from the input and noise parameters. A decision of voice activity is conducted over a four dimensional hyper-space, based on a region classification technique, followed by a hangover scheme.

The noise parameters are updated based on a first order autoregressive (AR) scheme, if the full-band energy difference is less than a certain fixed threshold. Unfortunately these parameters do not provide a good performance in various environments, that caused by parameters features. As an example the zero crossing rate has problems at low SNRs, especially in the presence of noise and speech with high zero crossing rates and the energy threshold method has problems in non-stationary noise and low SNRs. Therefore, in this paper, by changing some blocks in G.729 diagram, we try to increase the system performance in various environments also in low SNRs.

III. Parameters Used for Minimization

As described in section II, the G.729B standard needs to have a set of modifications to provide the better performance. Before description of this modification, we indicate some new methods:

A. TE-LPC

The true envelope estimator and then using the band limit envelope to derive an all pole envelope model named TE-LPC.

This proposition to improve the spectral envelope estimation is based on the *true envelope* estimator. The resulting estimation can be interpreted as a band limited interpolation of the observed sub-sampled spectral envelope [5].

Related to the speech signal, the resulting predictor is not optimal in the sense of the MSE criteria but it is supposed to fit closer the spectral envelope. A comparison between LPC and TE-LPC is shown in Fig.1.

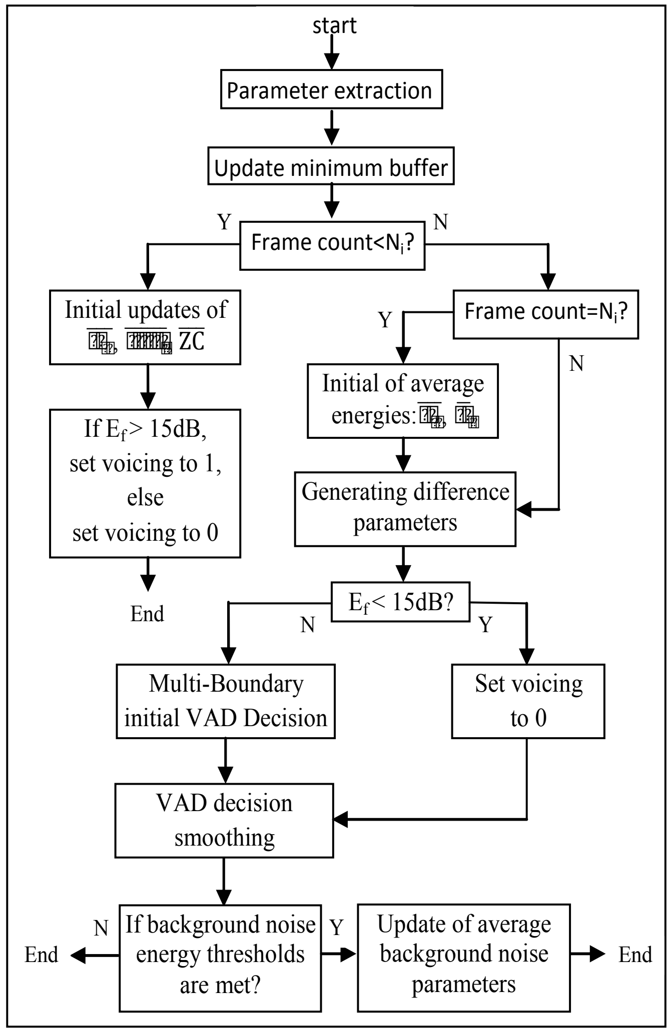


Fig. 1 Block diagram of ITU-T G.729B VAD

The results show that TE-LPC performs better Spectral-Peak Flatness Measure (SPFM) maximization in all the cases we measured. While the improvements are rather small if measured over the whole spectrum or the high frequency band, they are significant in the low frequency band which is more perceptually important.

The improvement is bigger for high-pitched signals and is not very sensitive to the model order for the selected order values.

Improvements found for the unvoiced cases could be due to voiced and mixed parts remaining in the unvoiced segments [6].

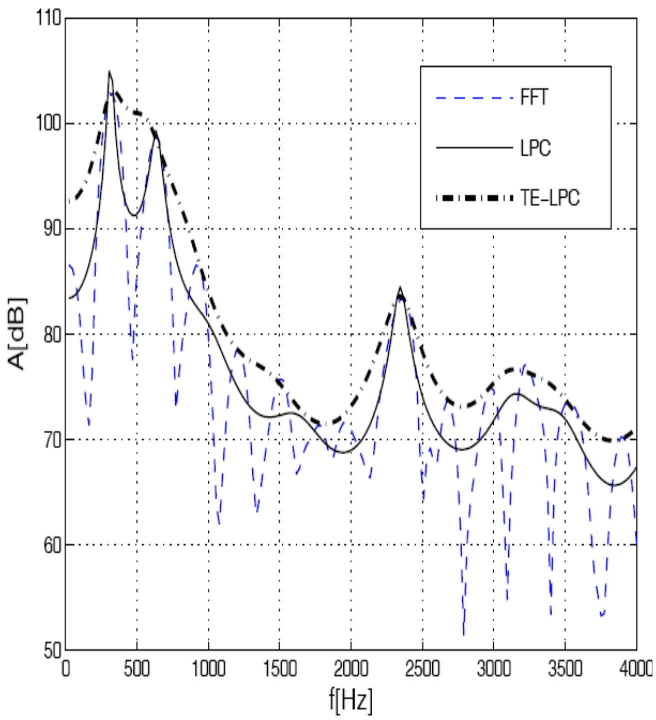


Fig. 2 Example of LPC and TE-LPC spectral fitting

B. Low Pass Filtering

A Low Pass Filtering (LPFing) operation is introduced as a pre-processing stage prior to decimation in order to overcome the problems of spectral aliasing present in the LSF calculation of the classic methods.

An LPFing experiment was performed to show the effect such an operation has on the resultant smoothed LSF vectors. The setup was as follows:

- First, LSF vectors $\boldsymbol{f}$ were extracted from 15 Hz bandwidth-expanded LPC parameters $\boldsymbol{A}=\alpha_1, \alpha_2, \alpha_3,\ldots, \alpha_p$ calculated every sample for Hamming windowed speech data of size 200 samples at 8 kHz sampling rate.
- LSF tracks $\boldsymbol{f_i}$ were then produced from the LSF vectors $\boldsymbol{f}$.
- Finally, filtering was performed in the frequency domain (using a rectangular window) for each LSF track $\boldsymbol{f_i}$ separately with a cut-off frequency as given by $f_c=1/(2\tau)$ (where τ is the system's LSF vector transmission rate according to sampling and decimation theories). This generated a second LSF vector set $\boldsymbol{g}=g_1, g_2, g_3, \ldots, g_p$.

A tenth-order LPC filter (i.e. p=10) was used and we therefore have ten LSF tracks. In Fig. 3a, it can be seen that all LSF tracks spectra have a substantial amount of their energy in the low frequency band (below 100 Hz). Fig. 3b shows the region of interest, from Fig. 3a, for 20, 10 and 5 ms LSF vector transmission rates.

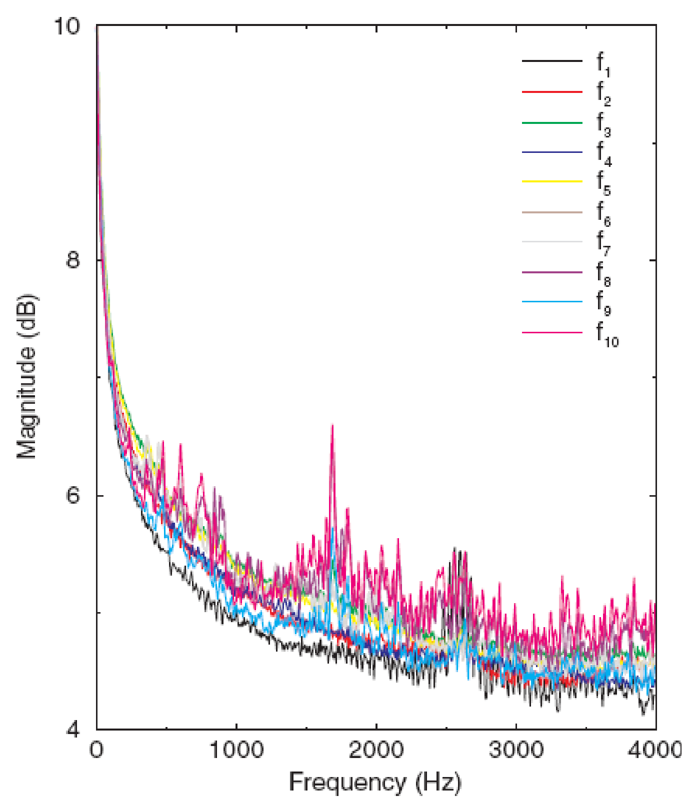
Each LSF track was generated from a speech file (for either a male or a female speaker, uttering two sentences of 4 s each). The FFT size used was sufficiently large (N = 65536) for the side lobe effects of the rectangular window to be ignored.

Figs. 4a and 4b show a section of the variations of certain LSF tracks for both classic $\boldsymbol{f_i}$ and LPFed $\boldsymbol{g_i}$ methods (for f_c= 25 Hz). It is evident in Fig. 4 that significant variations exist in the LSF tracks produced by the classic method because of the weak stationarity assumption within the analysis window, especially at transitions from the voiced speech segment to the unvoiced speech segment (offset) and vice versa (onset). The LPFed method, however, produces smoother and slowly evolving LSF tracks [32].

Knagenhjelm and Kleijn [33], show that using a perceptually smoothed (from an interpolation point of view) power spectral envelope leads to a significant increase in subjective performance.

Additionally, Eriksson et al. [34] show that low-rate quantization is possible through smoothing (from an LPFing perspective) the LSF parameter evolution. (Note that LPFing an over-sampled signal will follow the changes better than what is achieved through interpolation).

An informal listening test was conducted for synthesized speech of both male and female speakers, generated using both the classic, $\boldsymbol{f}$, and LPFed, $\boldsymbol{g}$, LSF vectors. The LPFed tracks $\boldsymbol{g_i}$ was decimated to an LSF transmission rate of 20 ms so that it could be used in the 2.4 kbps SB-LPC speech codec [35]. SB-LPC is a sinusoidal-based speech codec, which at the encoder extracts, quantizes and transmits speech model parameters [i.e. LPC (LSFs), pitch, energy, spectral amplitudes (excitation) and voicing]. At the decoder, these



a

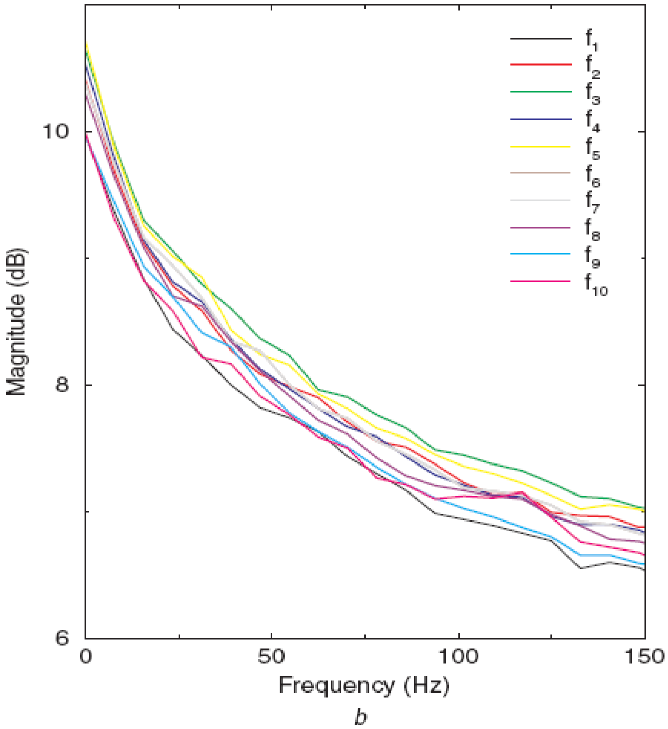


Fig. 3 Logarithmic magnitude spectrum of LSF tracks

parameters are dequantized to produce the excitation, which along with the LPC parameters, produces synthesized speech. Excitation in SB-LPC is discrete Fourier transform (DFT)-based, using a technique similar to that of the MELP coder [36]. All model parameters (in the listening tests) were quantized as in the original system [35]. Only LSF vectors were unquantized to measure whether LPFing removes important spectral details. Six expert listeners participated in the test in which they listened to 16 (4s) speech segments and then specified their preferences as A, B or similar. Ninety-five percent thought they were similar, whereas 5% interchangeably preferred one over the other. This similarity in the synthesized speech quality of both original and LPFed LSFs led us to believe that selecting the cut-off frequency of the LPF according to the final LSF vector transmission rate only removes the non-important information while keeping the quality and important information intact. Additionally, because the LPFed LSF vector tracks possess smoother and slower varying tracks, advantages are expected in terms of easier quantization and of gains through bit saving and better quality synthesized speech [32].

A. GAET

Recently Özer and Tanyer developed a new technique to estimate the optimum threshold for noise in the presence of speech accurately by using the amplitude probability distributions. The geometrically adaptive energy threshold (GAET) method is developed to set the threshold level adaptively without the need of voice inactive segments by using the amplitude probability distributions of the speech signal [7]. The GAET method is robust to non-stationary noise but false triggering is often observed when noise has short burst.

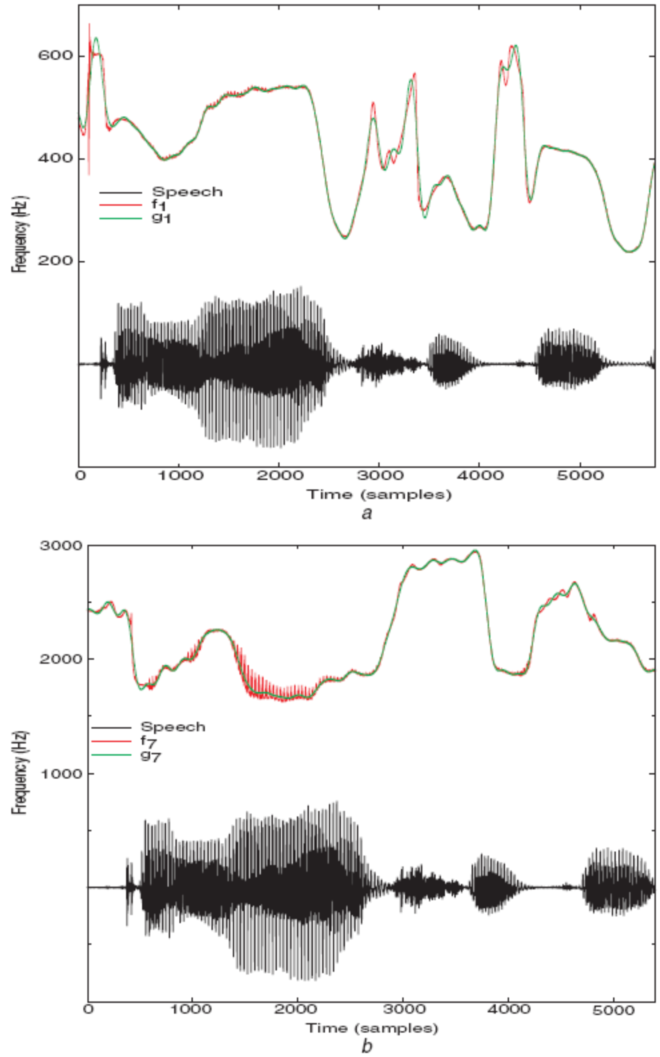


Fig. 4 Typical LSF track variations over time
a LSF tracks f_1 and g_1
b LSF tracks f_7 and g_7

B. LSPE

Tucker designed a VAD based on periodicity [8] named *Least-Square Periodicity Estimator* (LSPE). The major difficulty in designing a VAD based on periodicity is its sensitivity to any periodic signal which may well be interference or a background signal. Great care should be taken to avoid false triggering on non-speech periodic signals. If the speech signal contains non-periodic components, inaccurate values for endpoints of the voice-active segments could be obtained. Tucker used a preprocessor to detect and if possible remove, most of the expected types of interference. Different environments will have different interference, so the exact nature of the preprocessor will depend on the expected type of interference.

IV. Modifying G.729B VAD

In Fig. 5 we introduce a modified version of G.729 Annex B standard, based on four differential parameter that describe in section II.

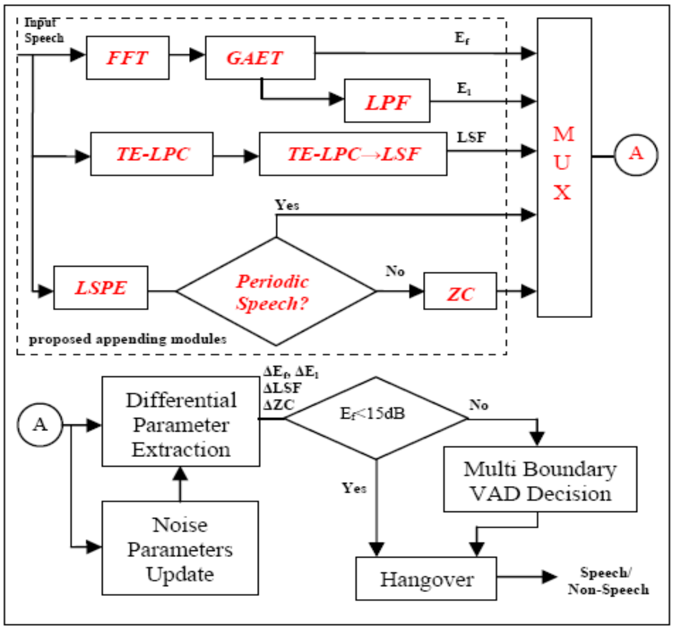


Fig. 5 Block diagram of Modified ITU-T G.729B VAD

The energy threshold methods (include of Low and Full-Band energy) have some problems in nonstationary and low values of SNR.

As we seen in new methods to modify the performance of VAD systems in the case of energy threshold [10]-[11], we used an adaptive threshold. In this way, we use one of the best methods that work rapid and have a good performance especially in Low-SNR. It is the GAET method.

Zero-Crossing (ZC) method has a better performance. However, this method also has some problems in Low- SNRs especially in presence of periodic noise and speech with high Zero-Crossing Ratios. To solve the problem of periodic speech we use the LSPE method. By using this method, the ZC eliminate in periodic speech that cause some improvements in rate of the system in the case of Low-SNR.

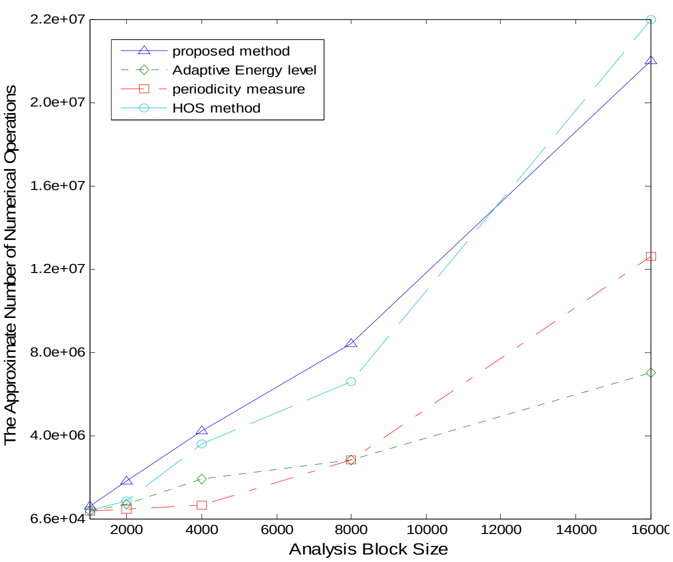


Fig. 6 Approximate numbers of numerical operations by the algorithm as a function of analysis blocks size.

To repair the LPC coefficients and estimate better spectrum shape of the speech signal, we use TE-LPC method. It has a good performance also its rate is two times better than LPC method.

We compare the complexity of commonly used VAD algorithms with our proposed method in Fig. 6. This method used four differential parameters and the complexity of the system is equal to sum of two methods; Adaptive energy method and periodicity measure. In the case of low SNRs, that the blocks size comes great, this method is better that higher order statistics method in rate.

V. PARAMETERS USED FOR COMPARISON

Using the implemented system outlined in Section 4, the effectiveness of the proposed algorithm was evaluated. Surveying literature indicates two distinct schools pertaining to VAD evaluation, namely subjective and objective evaluation. In general, subjective evaluation methods attempt to determine the effect of erroneous VAD decisions on human perception [9]. Tests such as the ABC [9] however does not take into consideration the effect of false alarms and as such are inappropriate for a thorough evaluation of VAD performance. Therefore, in order to evaluate the performance of the proposed scheme objective evaluation was used. In order to evaluate the amount of clipping and how often noise is detected as speech; the VAD output is compared with that of an ideal VAD, i.e., one obtained by manual marking of the database. The performance of a VAD is evaluated on the basis of the following four traditional parameters:

- *Front End Clipping* (FEC): Clipping introduced in passing from noise to speech activity.
- *Mid Speech Clipping* (MSC): Clipping due to speech misclassified as noise.
- OVER: Noise interpreted as speech due to the VAD flag remaining active in passing from speech activity to noise.
- *Noise Detected as Speech* (NDS): Noise interpreted as speech within a silence period.
- *Correct VAD decision* (Correct)*:* Correct decisions made by the VAD.

The FEC and MSC parameters give the amount of clipping introduced, whereas OVER and NDS give the increment in the activity factor. Fig. 7 shows the objective parameters for performance evaluation.

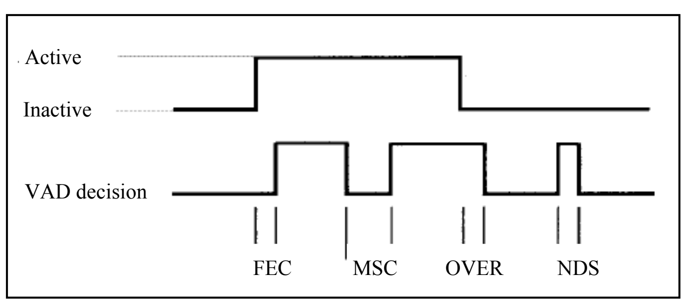
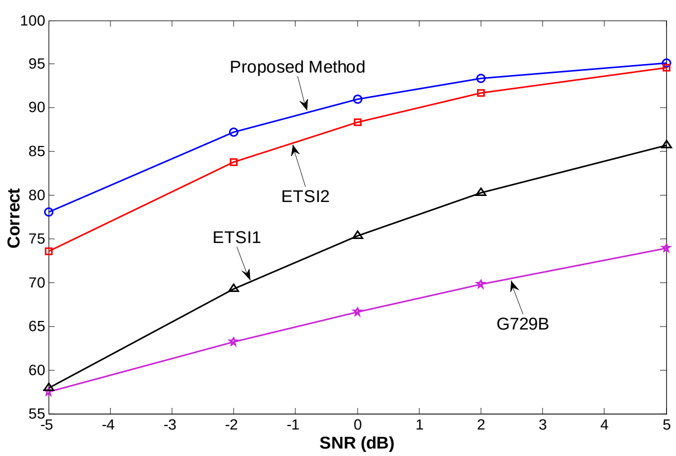


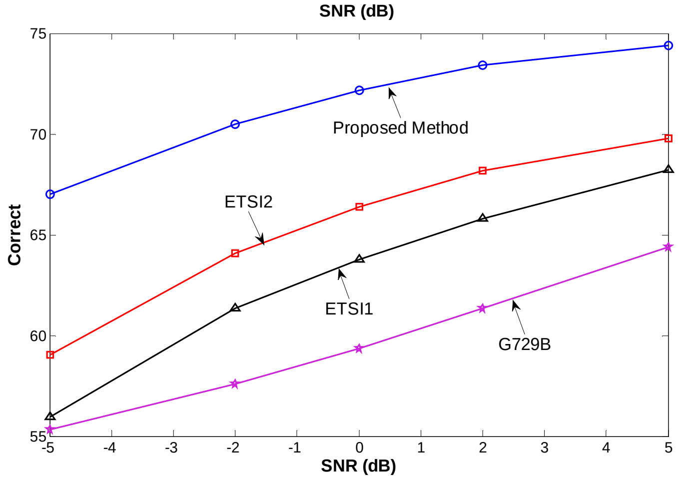
Fig. 7 objective parameters for performance evaluation

VI. Results

VAD performance comparison is complicated and time consuming process. It should be considered carefully. Ideally, a VAD should maximize the correct value, and minimize all errors. However failing this, the affect different types of errors have on the discontinuous speech signal (speech signal with non-speech periods removed and comfort noise inserted) should be considered. The purpose of a VAD in the context of a telephone conversation is to enable data savings by not transmitting non-speech periods, while maintaining speech quality. Speech quality should be of utmost importance. Therefore, it is important to note the affect that each of the different errors have on speech quality.

In contrast, insertion errors such as NDS and OVER do not have any effect on speech quality. They do however result in reduced effectiveness of the VAD scheme. Here we will use the broad notion that clipping errors are less desirable than insertion errors. In Fig. 8 we show a comparison of correct parameter for three standards, ITU-T G.729 Annex B, ETSI AMR option 1 and 2 with our proposed method. In the case of G.729B, it exhibits the worst average results over the correct test. In Fig. 6, the first part shows the signal in the gaussian noise. It is observed that, AMR2 and proposed method have same performance in high SNRs, but the modified G.729B has a better performance in low SNRs. In the case of babble noise that illustrate in the second part, we have a good modification and in the third part, the proposed scheme has a good performance but lower than ETSI AMR1 for vehicle noise.





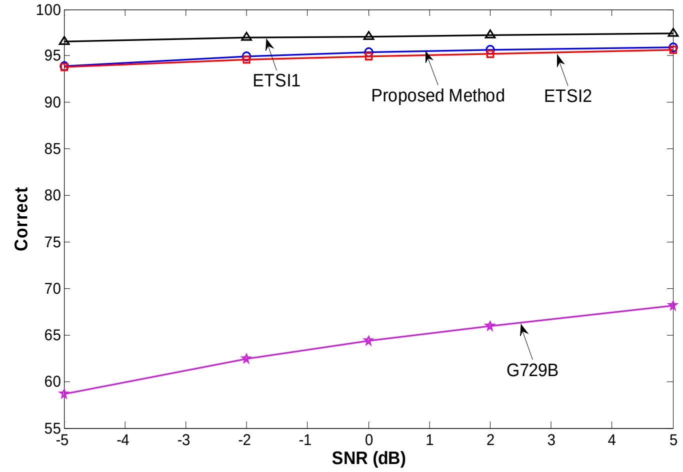
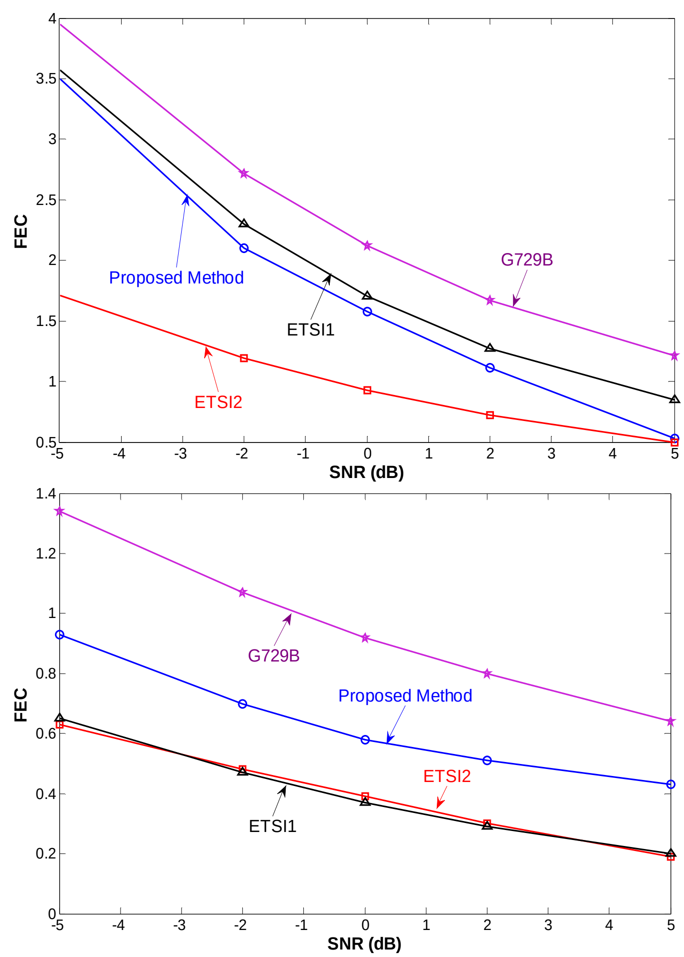


Fig. 8 Comparison of Correct Parameter for G.729B, ETSI AMR1, 2 and Modified version of G.729B in (a) Gaussian, (b) Babble and (c) Vehicle Noise

One of the other parameters that exhibit in Fig. 7 is FEC parameter for three standards, and our proposed method. As you know, we try to minimize Front End Clipping (FEC) parameter. In the case of G.729B, it exhibits the worst average results over the FEC test. In Fig. 9, the first part shows the signal in the gaussian noise. It is observed that, AMR option 2 and proposed method have same performance in high SNRs, but AMR option 2 has a better performance in low SNRs.



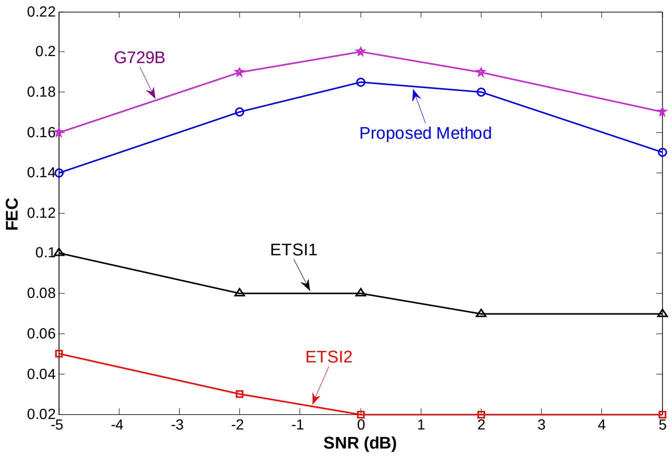


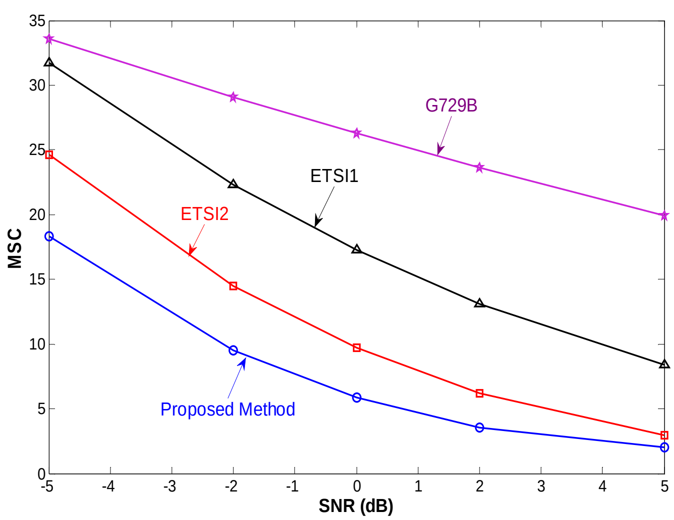
Fig. 9 Comparison of FEC Parameter for G.729B, ETSI AMR1, 2 and Modified version of G.729B in (a) Gaussian, (b) Babble and (c) Vehicle Noise

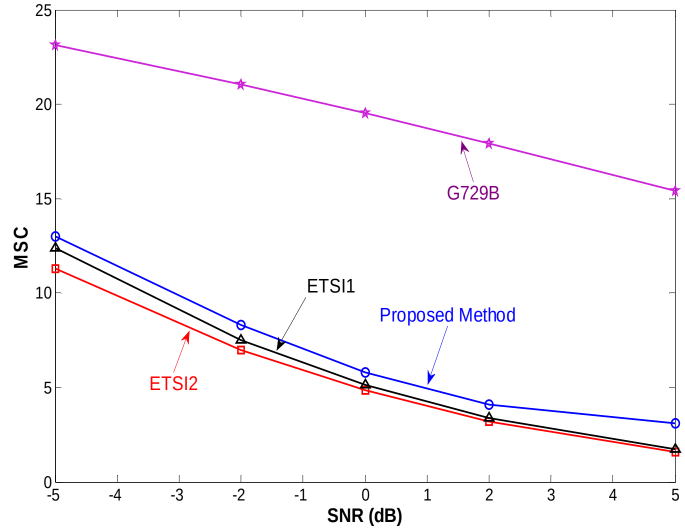
In the case of babble noise that illustrate in the second part, we has a good modification but ESTI AMR1 and 2 have better performance and in the third part, the modified version of G.729B has some modification in respect of G.729B. Gaussian noise is complicated respected in babble and vehicle noise.

Another parameter that we exhibit in Fig. 7 is MSC parameter for three standards, ITU-T G.729 Annex B, ETSI AMR option 1 and 2 and our proposed method.

Mid Speech Clipping (MSC) parameter, like FEC parameter, must be minimized. In the case of G.729B, it exhibits the worst average results over the MSC test. In Fig. 8, the first part shows the signal in the gaussian noise. It could be observed that, AMR option 2 and proposed method have same performance in high SNRs, but our proposed method has a better performance in low SNRs.

In the case of babble noise that illustrate in the second part, we have a good modification and the performance of proposed method and ESTI AMR1 and 2 are same; and in the third part, the modified version of G.729B has some modification in respect of G.729B but ETSI AMR1 and 2 are better. In summary this parameter, after correct parameter is very important.





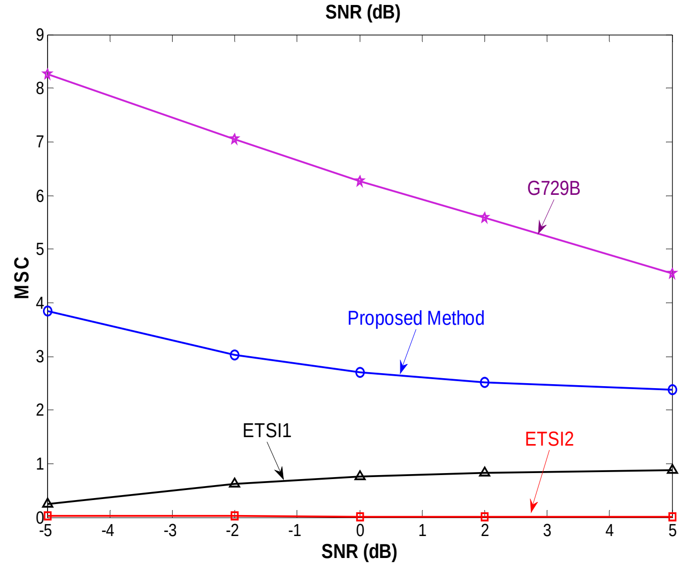


Fig. 10 Comparison of MSC Parameter for G.729B, ETSI AMR1, 2 and Modified version of G.729B in (a) Gaussian, (b) Babble and (c) Vehicle Noise

Next parameter that occur in the silence parts of speech signal and shows in Fig. 11 is NDS parameter for three standards, ITU-T G.729 Annex B, ETSI AMR option 1 and 2 and our proposed method. This parameter is greater when are in noisy environments and low SNRs. Noise Detected as speech (NDS) parameter, like FEC and MSC parameters, must be minimized.

In the case of G.729B, it exhibits the worst average results over the vehicle noise, but in the case of babble and gaussian noise, ETSI AMR1 is worst. In Fig. 9, the first part shows the signal in the gaussian noise.

It is observed that, ETSI AMR option 2 has the better performance and proposed method is next; although all of this standards and this method have good performance in respect of babble and vehicle noise.

In the case of babble noise that illustrate in the second part, the G.729B standard has a good performance and we have a good modification. The performance of ESTI AMR2 is excellent; and in the third part, the modified version of G.729B has some modification in respect of G.729B and same to ETSI AMR1 and 2. An important care should be give in the case of babble noise.

It could be observed that in the second figure, previous standards have not good performance in the babble noise but modified version of G.729B easily extracts the speech signal in a loud and noisy environments.

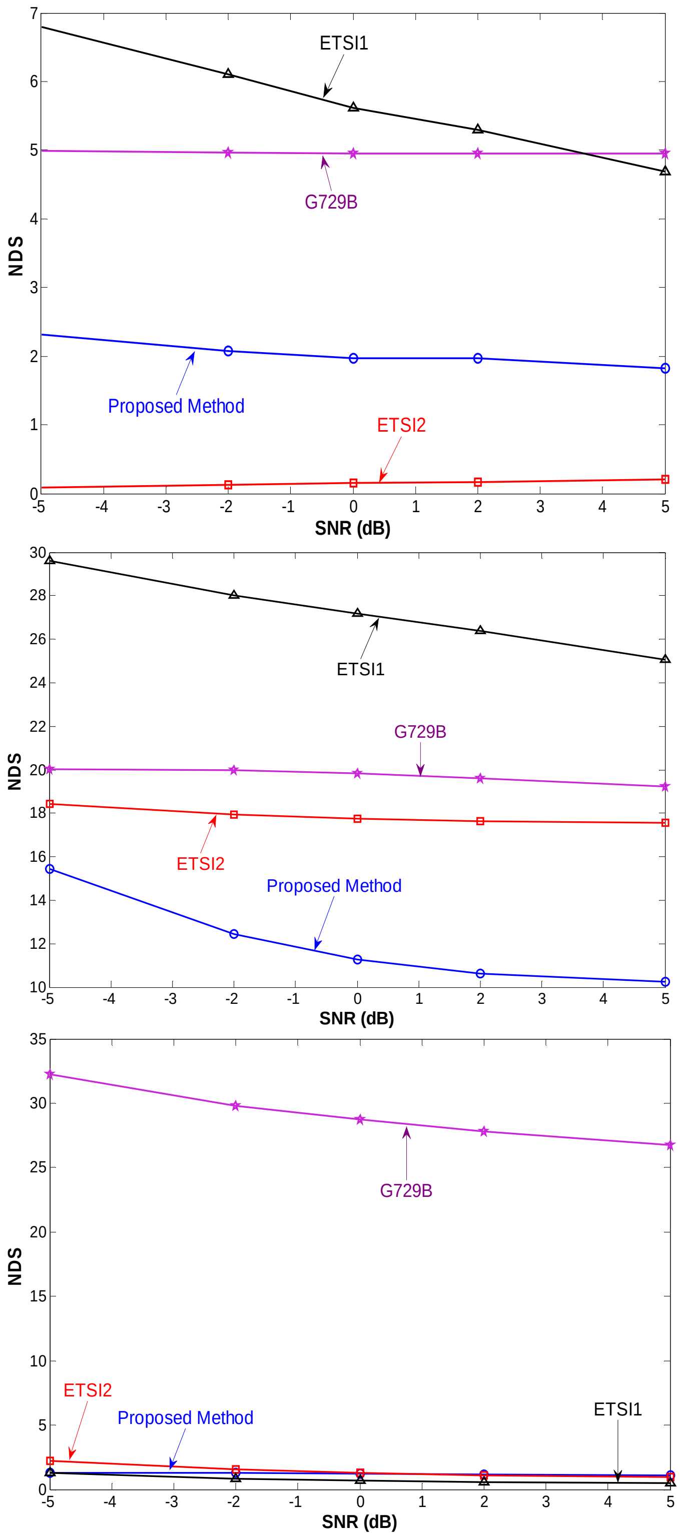


Fig. 11 Comparison of NDS Parameter for G.729B, ETSI AMR1, 2 and Modified version of G.729B in (a) Gaussian, (b) Babble and (c) Vehicle Noise

Finally, the last parameter that could be observed in Fig. 7 occurs when noise interpreted as speech due to the voice activity detection flag remaining active in passing from speech activity to noise. It's the opposite parameter of the FEC that we want it to minimize.

As you know, in the most processors of speech signal, we have periodic properties and consider that noise is a random signal; but in the some parts of speech signal it may be a random, also the noise has some periodic properties. In this case OVER parameter may occur.

Fig. 12 shows the OVER parameter in three standards, ETSI AMR option 1, 2, G729B and the modified version of G.729B. In the same way, we compare this method by other modern standards in the various noisy environments.

In previous parameters ETSI standards, especially ETSI AMR2, have good performance but in this parameter, G.729B is better and we can simply modify it.

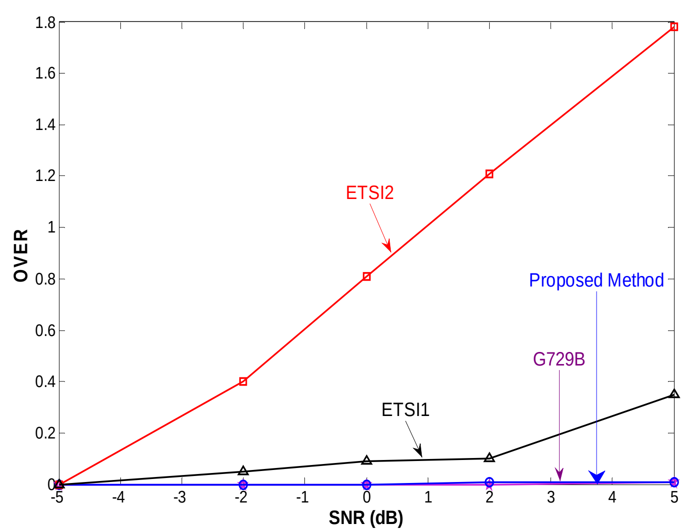
In the case of ETSI AMR2, it exhibits the worst average results the OVER test. In Fig. 12, the first part shows the signal in the gaussian noise.

It is observed that, G.729 and the modified version have same and better performances and ETSI AMR is good for low SNRs; although all of this standards and this method have good performance in respect of babble noise.

In the case of babble noise that illustrate in the second part, the G.729B standard has a good performance and we have a good modification. The performance of ESTI AMR1 is better than AMR2.

In the third part, for vehicle noise, the modified version of G.729B has some modification in respect of G.729B and better than ETSI AMR1 and 2. An important care should be give in the case of babble noise because it's heave for programming systems.

This parameter is very important in the packet networks, because we don't want to transmit empty packets and joint them to the speech signal. OVER parameter and FEC parameter are not great but important. This is one of the main targets of voice activity detection systems.



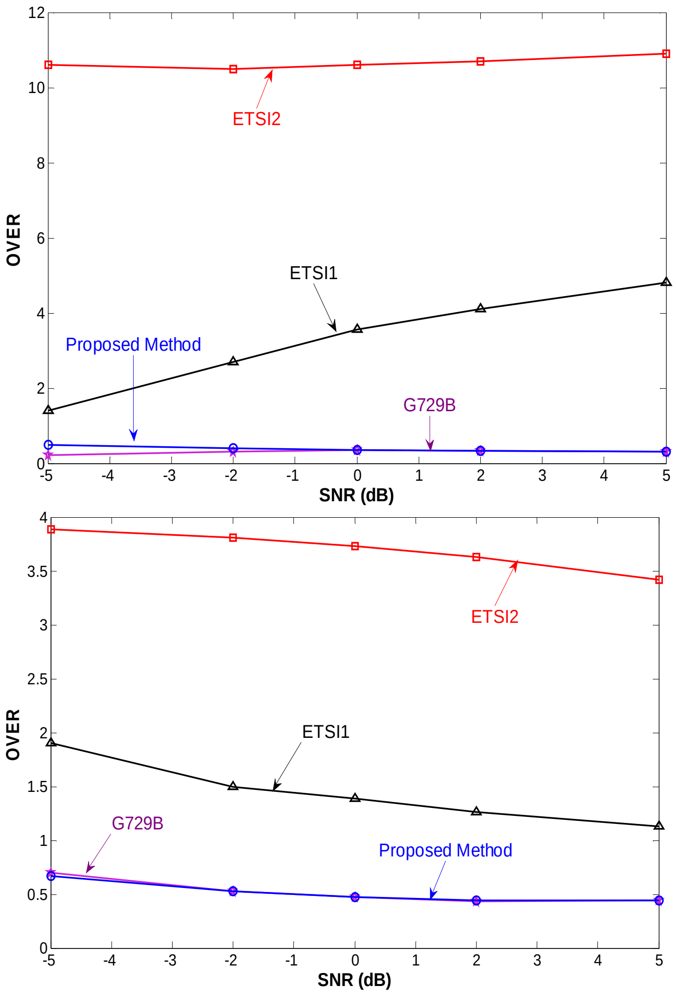


Fig. 12 Comparison of OVER Parameter for G.729B, ETSI AMR1, 2 and Modified version of G.729B in (a) Gaussian, (b) Babble and (c) Vehicle Noise

In summary, the proposed scheme presents a better alternative compare with standardized algorithms. It exhibits a consistent correct rate over a variety of noise environments and conditions. It has lower average NDS than all standardized algorithms over the test set and has low FEC and MSC while maintaining a low OVER rate.

These characteristics make it a simple and reliable choice for many VAD applications. Further, the scheme requires only low computation time and memory.

The simplicity of the proposed VAD coupled with the encouraging results, mathematical tractability and high detection consistency make it a good alternative to current schemes. The behavior of the VAD is easily altered by changing one meaningful parameter, and as such makes the VAD well suited to varying applications.

VII. Conclusion

There are a lot of works in the case of Voice Activity Detection Systems. But many of these methods are not practical for telecommunication systems, because it must be matched with modern standards. G.729B standard uses four parameter for VAD system. Unfortunately it has a poor performance in low SNRs. In this paper, whereas the rate and complexity of this standard are better than spectral shape (i.e. GSM-FR) and sub-band energy standards (i.e. IS-95, AMR 1 and 2), we try to modify G729B standard with minimum changes.

References

[1] R. V. Cox and P. Kroon, "Low bit-rate speech coders for multimedia communications," *IEEE Commun. Mag.*, vol. 34, pp. 34–41, Dec. 1996.

[2] A. Benyassine, E. Shlomot, and H.-Y. Su, "ITU-T recommendation G.729 annex B: A silence compression scheme for use with G.729 optimized for V.70 digital simultaneous voice and data application," *IEEE Commun. Mag.*, vol. 35, pp. 64–73, Sept. 1997.

[3] GSM 06.94. (1999, Feb.) Digital cellular telecommunication system (Phase 2+); voice activity detector VAD for adaptive multi rate (AMR) speech traffic channels; general description. ETSI, Tech. Rep. V.7.0.0. [Online]. Available: http://www.etsi.org .

[4] A. Benyassine, E. Shlomot, and H.-Y. Su, "ITU-T recommendation G.729 annex B: A silence compression scheme for use with G.729 optimized.

[5] O. Cappé and E. Moulines, "Regularization techniques for discrete cepstrum estimation," *IEEE Signal Processing Letters*, vol. 3, no. 4, pp. 100–102, 1996. For V.70 digital simultaneous voice and data application," *IEEE Commun. Mag.*, vol. 35, pp. 64–73, Sept. 1997.

[6] F. Villavicencio, A. Röbel and X. Rodet, "Improving LPC Spectral Envelope Extraction of Voiced Speech by True-Envelope Estimation," *ICASSP*, pp. 869–872, Sept. 2006.

[7] H. Özer and S. G. Tanyer, "A geometric algorithm for voice activity detection in nonstationary Gaussian noise," in *Proc. EUSIPCO'98*, Rhodes, Greece, Sept. 1998.

[8] R. Tucker, "Voice activity detection using a periodicity measure," *Proc. Inst. Elect. Eng.*, vol. 139, pp. 377–380, Aug. 1992..

[9] F. Beritelli, S. Casale, and G. Ruggeri, "A physcoacoustic auditory model to evaluate the performance of a voice activity detector," in *Proc. Int. Conf. Signal Processing*, vol. 2, Beijing, China, 2000, pp. 69–72. S. Chen, B. Mulgrew, and P. M. Grant, "A clustering technique for digital communications channel equalization using radial basis function networks," *IEEE Trans. Neural Networks*, vol. 4, pp. 570–578, July 1993.

[10] A. Davis, S. Nordholm and R. Togneri, "Statistical Voice Activity Detection Using Low-Variance Spectrum Estimation and an Adaptive Threshold" *IEEE Trasns. On audio, speech and language processing, vol. 14, NO. 2, March 2006,* pp. 412-424.

[11] J. Sohn and W. Sung, "A voice activity detector employing soft decision based noise spectrum adaptation," in *Proc. IEEE ICASSP'98*, vol. 1, Seattle, WA, 1998, pp. 365–368.

[12] Kun-Ching Wang, Chiun-Li Chin: 'An Efficient Voice Activity Detection in Realistic Environments' WSEAS TRANSACTIONS on SYSTEMS, Issue 7, Volume 6, July 2007.

[13] Jun Li, Zheng Tang: 'Enhancing the Human-Computer Interaction on Camera-Equipped Mobile Phones' WSEAS TRANSACTIONS on SYSTEMS, Issue 7, Volume 6, July 2007.

[14] H. Farsi,: 'A Novel Post filtering Technique Using Adaptive Spectral Decomposition for Quality Enhancement of Coded Speech' WSEAS TRANSACTIONS on Signal Processing, Issue 5, Volume 4, May 2008, p289-299.

[15] K. Bullington and J. M. Fraser, "Engineering aspects of TASI," *Bell Syst. Tech. J.*, pp. 353–364, Mar. 1959.

[16] O. Tanrikulu, B. Baykal, A. G. Constantinides, and J. A. Chambers, "Critically sampled sub-band acoustic echo cancellers based on IIR and FIR filter banks," *IEEE Trans. Signal Processing*, vol. 45, pp. 901–912, Apr. 1997.

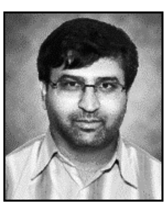
[17] L. R. Rabiner and M. R. Sambur, "Voiced-unvoiced-silence detection using the Itakura LPC distance measure," in *Proc. Int. Conf. Acoust., Speech, Signal Processing*, May 1977, pp. 323–326.

[18] J. C. Junqua, B. Reaves, and B. Mak, "A study of endpoint detection algorithms in adverse conditions: Incidence on a DTW and HMM recognize," in *Proc. Eurospeech'91*, 1991, pp. 1371–1374.

[19] J. A. Haigh and J. S. Mason, "Robust voice activity detection using cepstral features," in *Proc. IEEE TENCON*, China, 1993, pp. 321–324.

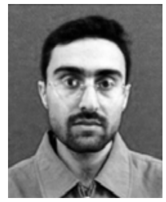
[20] N. B. Yoma, F. McInnes, and M. Jack, "Robust speech pulse-detection using adaptive noise modeling," *Electron. Lett.*, vol. 32, July 1996.

[21] R. Tucker, “Voice activity detection using a periodicity measure,” *Proc. Inst. Elect. Eng.*, vol. 139, pp. 377–380, Aug. 1992.
[22] D. K. Freeman, G. Cosier, C. B. Southcott, and I. Boyd, “The voice activity detector for the Pan-European digital cellular mobile telephone service,” in *Proc. Int. Conf. Acoust., Speech, Signal Processing*, Glasgow, U.K., May 1989, pp. 369–372.
[23] K. Srinivasan and A. Gersho, “Voice activity detection for cellular networks,” in *Proc. IEEE Speech Coding Workshop*, Oct. 1993, pp. 85 86.
[24] S. Sasaki and I. Matsumoto, “Voice activity detection and transmission error control for digital cordless telephone system,” *IEICE Trans. Commun.*, vol. E77B, no. 7, pp. 948–955, 1994.
[25] D. R. Halverson, “Robust estimation and signal detection with dependent nonstationary data,” *Circuits Syst. Signal Process,* vol. 14, no. 4, pp. 465–472, 1995.
[26] M. J. Irwin, “Periodicity estimation in the presence of noise,” in *Proc. Acoust. Conf.’79*, Windemere, U.K., 1980.
[27] F. Horner, *Advance in Radio Research*, New York: Academic, 1964, vol. 2, pp. 121–204.
[28] A. D. Spaulding,W. H. Ahlbeck, and L. R. Espeland, “Urban residential man-made radio analysis and predictions,” U.S. Govt. Printing Office, Washington, DC, OT Telecommun. Res. Eng. Rep., 1971.
[29] T. Nakai and H. Ohba, “On the graphical method of drawing APD’s for atmospheric radio noise,” *IEEE Trans. Electromagn. Compat.*, vol. EMC-26, pp. 71–78, May 1984.
[30] H. Özer, “Signal detection and estimation in nonstationary background,” M.S. thesis, Dept. Elect. Electron. Eng., Bas¸kent Univ., Ankara, Turkey, Aug. 1998.
[31] K. Srinivasan and A. Gersho, “Voice activity detection for cellular networks,” in *Proc. of the IEEE Speech Coding Workshop.,* pp- 85-86, October 1993.
[32] K. Al-Naimi, S. Villette, A. Kondoz and A.P. Heikkinen, “Improved line spectral frequency estimation through anti-aliasing filtering” IEE Proc.-Vis. Image Signal Process., Vol. 153, No. 5, October 2006.
[33] Knagenhjelm, H.P., and Kleijn, W.B.: “Spectral dynamics is more important than spectral distortion”. IEEE Proc. Int. Conf. on Acoustics, Speech, and Signal Processing, New York, NY, USA, 1995, vol. 1, pp. 732–735.
[34] Eriksson, T., Kang, H.-G., and Hedelin, P.: ‘Low-rate quantization of spectrum parameters’. Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing, Istanbul, Turkey, 2000, vol. 3, pp. 1447–1450.
[35] Villette, S.: ‘Sinusoidal speech coding for low and very low bit rate application’. Thesis, CCSR, University of Surrey, 2001.
[36] McCree, A., Truong, K., George, E.B., Barnwell, T.P., and Viswanathan, V.: ‘A 2.4 Kbits/s MELP coder candidate for the new U.S. Federal Standard’. Int. Conf. on Acoustics, Speech, and Signal Processing, 1996, vol. 1, pp. 200–203.



Hassan Farsi received the B.Sc and M.Sc degrees from Sharif University of technology, Tehran, Iran, in 1992 and 1995, respectively.
From 1992 to 1993, he worked in research communication centre, Tehran, Iran on waveform speech coders.
In 1995, he was employed in the University of Birjand, Birjand, Iran, as a lecturer in the department of Electronics and communication engineering. Since 2000, he started his PhD in the centre of communication systems research (CCSR), University of Surrey, Guildford, UK, in field of speech quality improvement provided by low bit rate speech coders and received the PhD degree in 2004. His research has also led to improving speech quality coded by 2.4 Kbits/s standard MELP coder. Beside speech processing, he is also interested in image and digital signal processing.

Mohammad Amin Mozaffarian received the B.Sc degree from Azad University of Karaj, Karaj, Iran, in 2006.
From 2006 to 2007, he worked in international providence Lab, Tehran, Iran on wireless and control systems.
He started his M.Sc in 2007 University of Birjand, Birjand, Iran. He is working in Azad University of Birjand as a lecturer. His research has also led to improving speech quality coded by 8 Kbits/s standard ACELP coder. Beside speech processing, he is also interested in digital signal processing and wireless communication systems.



Hossein Rahmani Darenjani received the B.Sc degree from Emam Hossein University, Tehran, Iran, in 2005.
He started his M.Sc in 2007 University of Birjand, Birjand, Iran. He has also led to wireless communication systems.
Beside speech processing, he is also interested in MC-CDMA.