Adapting Correction Factors in Probability Distribution Function for VAD Improvement

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

***Abstract*—**One of the new methods that used in Voice activity detection (VAD) systems is estimating the Probability Distribution Function (PDF) of the speech signal. This estimation becomes hard in noisy environments especially low value of Signal-to-Noise Ratios (SNR). This paper studies on three types of PDFs and selects one of them to modify and approximate the original signal. Then we compare the results of this PDF before and after modification.

***Keywords*—**Voice Activity Detection, Statistical Model, Correction Factors, Probability Distribution Function, Low Signal-to-Noise Ratio.

I. INTRODUCTION

THE process of separating conversational speech and silence is called the *voice activity detection* (VAD) [1]. In communications 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 [8-10].

In early VAD algorithms, short-term energy, zero-crossing rate and LPC coefficients were among the common features used for speech detection [11]. Cepstral features [12], formant shape [13], and least-square periodicity measure [14] are some of the more recent metrics used in VAD designs. In the recently proposed G.729B VAD [15], a set of metrics including line spectral frequencies (LSF), low band energy, zero-crossing rate and full-band energy is used along with heuristically determined regions and boundaries to make a VAD decision for each 10 ms frame.

When In order to improve the detection accuracy in low SNR case, especially when the noise is nonstationary, some robust VAD algorithms have been proposed. In [16] and [17], a voice detection algorithm based on a pattern recognition approach in which the matching phase is performed by a set of six fuzzy rules, namely, Fuzzy VAD, is introduced. In [18], a fusion algorithm which combines the geometrically adaptive energy threshold (GAET) method [19] and the Least-square periodicity estimator method [20] was proposed. The GAET method keeps track of the nonstationary background noise while the LSPE analyzes the periodical content of the incoming signal. Both of the two algorithms are shown to operate reliably down to very low SNR cases, say, 0 dB or even -5 dB.

Recently, attempts have been made to develop a statistical model-based VAD [21], [22]. These schemes adopt the model proposed by Ephraim and Malah. The model assumes Fourier coefficients are statistically independent Gaussian random variables [4] and is motivated by the central limit theorem. Using this model a likelihood ratio is developed and a statistical hypothesis test conducted.

The formulation of the hypothesis test presents some problem. It indicates two key parameters need to be determined, namely the *a priori* and *a posteriori* signal-to-noise ratios [4]. The problem of determining the *a priori* signal-to-noise ratio is addressed by estimating MMSE speech spectral amplitudes. This estimation however is undesirable, introducing complexity and a computational burden. The *a posteriori* signal-to-noise ratio is estimated using a scaled periodogram. Both ratios further depend on an estimate of the variance of the Fourier coefficients during periods of noise. This variance is either determined *a priori* or estimated using an exponential average of a scaled periodogram [5].

Further, the issue of determining the threshold for the hypothesis test is ignored. Bayesian hypothesis testing indicates a threshold should be determined on the basis of a cost or risk function [23]. This however requires *a priori* knowledge of the probabilities of occurrence of each hypothesis, which in this case makes determining a threshold in this manner impractical. Cho *et al.* addressed this and indicated a region for the threshold, but gave no specific analysis [22]. In general the threshold is set by some heuristic rule.

The schemes were reported to produce good results in both babble and vehicle noise. Sohn *et al.* also evaluated the scheme in Gaussian noise; however, results indicated a declining performance below 15 dB [21]. This is due, at least in part, to the method of estimation of the key parameters outlined earlier, namely a scaled periodogram. The periodogram is well known to be an inconsistent spectral estimator [24]. Typically it is shown that the variance is approximately the same size as the square of the power spectrum that is being estimated, and does not decrease with increasing data length. This high variance contributes to the reduced performance in white noise.

Another statistical scheme has been developed by McKinley and Whipple [25]. This scheme, in contrast to other statistical methods compares second-order statistics of the signal to models.

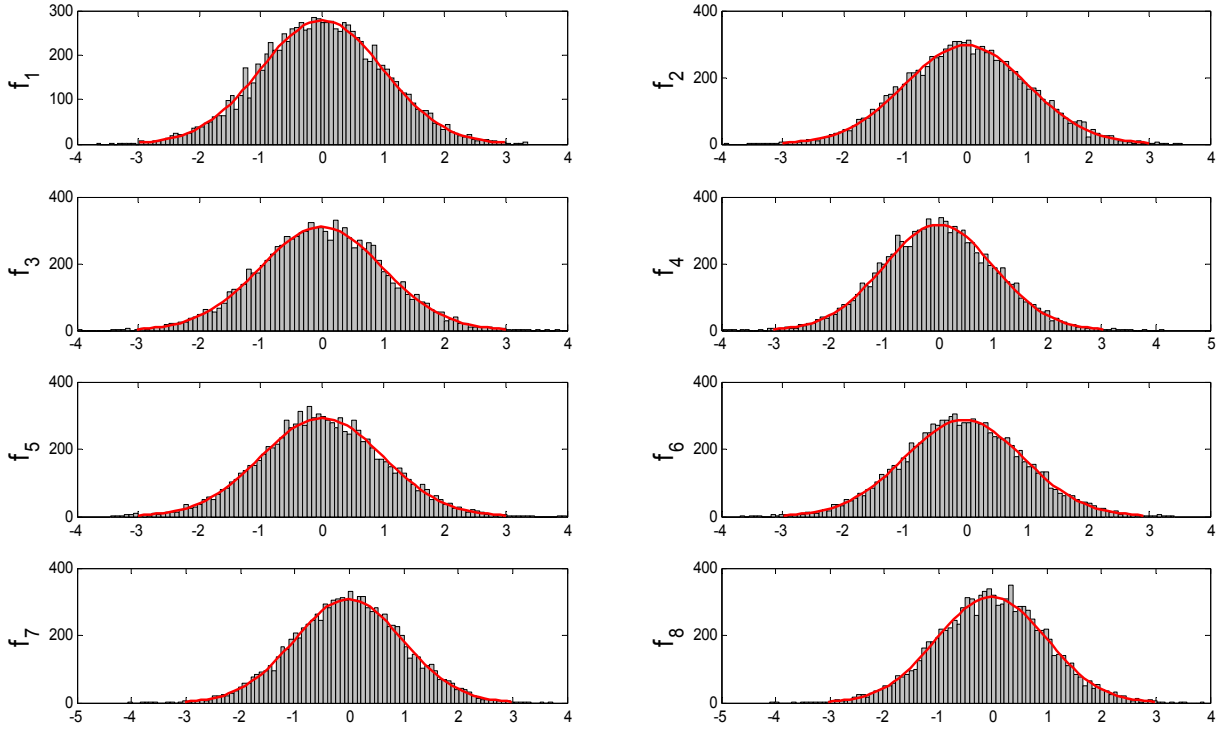


Fig. 1 Histogram of SNR measure in Gaussian noise environment for each frequency bin f_n

Speech models are estimated from a large speech set developed off-line, and noise models are estimated during an initial silence period. The scheme was reported to produce good results in a range of environments; however, only a small test set was used. Further, the scheme is computationally expensive and complex.

In this paper, the candidate models are the distribution of the spectral components under various noisy conditions. Not only the traditional Gaussian PDF but also the complex Laplacian and Gamma PDFs are applied to represent the distribution of each Discrete Fourier Transform (DFT) coefficients. We extend the PDFs described in [2] in low SNRs. At the end, we consider a set of correction coefficients to improve the performance of the estimation.

II. Statistical Model for Noisy Speech

In order to validate this, the PDF of speech signal was experimentally calculated in a range of noise environments. The measure was found to follow closely a Gaussian distribution in stationary noise environments such as Gaussian noise, pink noise, and HF channel noise as taken from the NOISEX-92 database. In highly variable environments such as babble and vehicle noise, the assumption is violated. Normalized histograms along with a Gaussian fit with zero mean can be seen in Figs. 1–3. The figures represent the Gaussian, vehicle, and babble noise environments as taken from the NOISEX-92 database. The effect of the Gaussian assumption failing is reduced performance in the affected environment. This is generally manifested as false alarms, due to the long tails on the probability density function (pdf). This phenomenon can be seen in the evaluation where the babble noise environments is considered.

In this section we extend the hypothesis of Gaussian PDF and study the best performance of Gamma, Laplacian and Gaussian distributions especially in low SNR. Then we evaluate the error function of the best distribution to modify the false alarm.

The first distribution is Gaussian PDF. We assume that the noise signal *n(t)* is added to the speech signal *x(t)*, with their sum being denoted by *y(t)* in time domain. *y(t)* is transformed by the Discrete Fourier Transform (DFT) as follows:

$$Y(t) = X(t) = N(t) \tag{1}$$

where:

$$\mathrm{Y(t)} = [\mathrm{Y_1(t)}, \mathrm{Y_2(t)}, \ldots, \mathrm{Y_m(t)}],$$
$$\mathrm{X(t)} = [\mathrm{X_1(t)}, \mathrm{X_2(t)}, \ldots, \mathrm{X_m(t)}],$$
$$\mathrm{N(t)} = [\mathrm{N_1(t)}, \mathrm{N_2(t)}, \ldots, \mathrm{N_m(t)}]$$

denote the DFT factors of the noisy speech signal, clean speech, and the added noise. Given two classes, H_0 and H_1 which, respectively, indicate speech presence and absence, it is assumed that:

$$\mathrm{H_0: speech\ absent\ : Y_K(t) = N_K(t)}$$
$$\mathrm{H_1: speech\ present\ : Y_K(t) = X_K(t) + N_K(t)}$$

With the Gaussian PDF assumption, the distributions of the noisy spectral components conditioned on both hypotheses are given by:

$$\mathrm{p_G\,(Y_k \mid H_0)} = \frac{1}{\pi\lambda_{n,k}} \exp\{-\frac{|\mathrm{Y_k}|^2}{\lambda_{n,k}}\} \tag{2}$$

$$\mathrm{p_G\,(Y_k \mid H_1)} = \frac{1}{\pi[\lambda_{n,k}, \lambda_{x,k}]} \exp\{-\frac{|\mathrm{Y_k}|^2}{\lambda_{n,k} + \lambda_{x,k}}\} \tag{3}$$

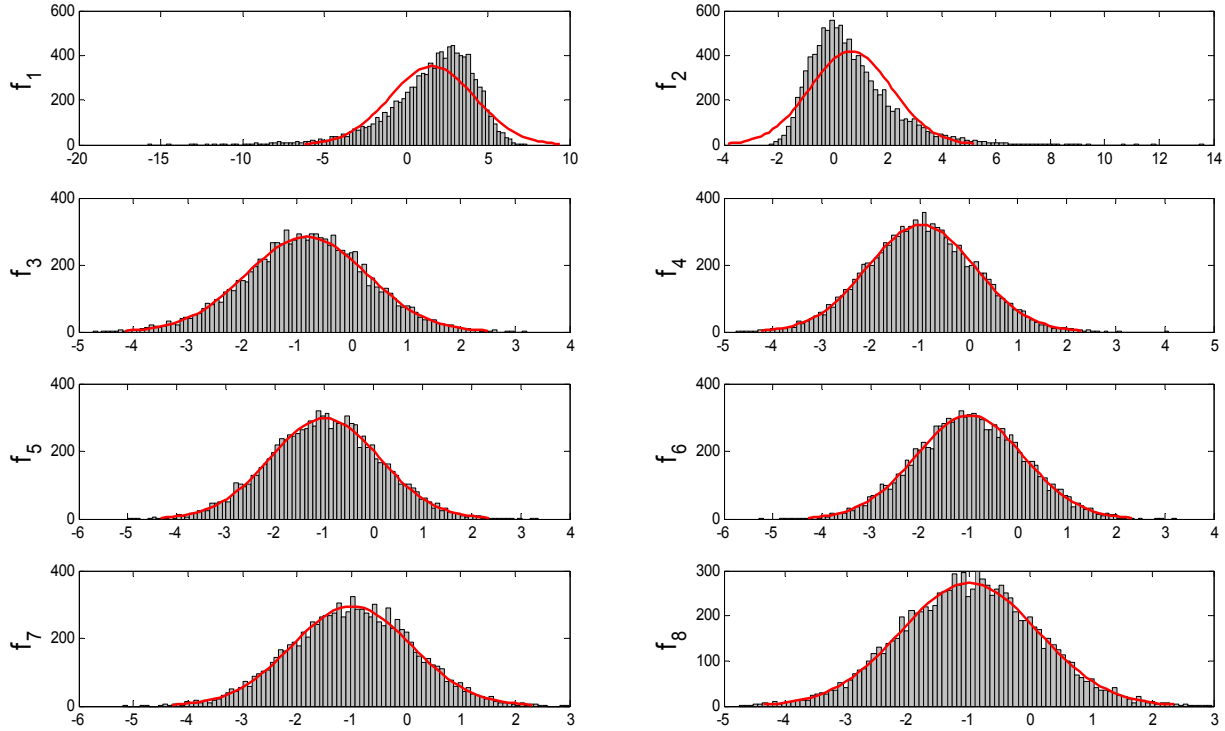


Fig. 2 Histogram of SNR measure in Vehicle noise environment for each frequency bin f_n

where $\lambda_{x,k}$ and $\lambda_{n,k}$ indicate the variances of noise and speech for the individual frequency band, respectively.

The second distribution is the complex Laplacian PDF. The real and imaginary parts of each DFT coefficients are assumed to be distributed according to a real Laplacian PDF. Let $X_{k(R)}$ and $X_{k(I)}$ denote the real and imaginary parts, respectively, of the DFT coefficients X_k.

If both the real and imaginary parts have the same variances and assume to be independent [3], the distribution $p(X_k)$ of X_k turns out to be:

$$p_L(X_k) = p_L(X_{k(R)}).p_L(X_{k(I)}) = \frac{1}{\sigma_x^2}\exp\{\frac{-2(|X_{k(R)}|+|X_{k(I)}|)}{\sigma_x}\} \quad (4)$$

From this equation, the distributions of the DFT coefficients under the respective hypotheses are given by [4]:

$$p_G(Y_k|H_0) = \frac{1}{\lambda_{n,k}} \times \exp\{-\frac{-2(|X_{k(R)}|+|X_{k(I)}|)}{\sqrt{\lambda_{n,k}}}\} \quad (5)$$

$$p_G(Y_k|H_1) = \frac{1}{\lambda_{n,k},\lambda_{x,k}} \times \exp\{-\frac{-2(|X_{k(R)}|+|X_{k(I)}|)}{\sqrt{\lambda_{n,k}+\lambda_{x,k}}}\} \quad (6)$$

The last statistical model is described in terms of the complex Gamma PDF.

If the real and imaginary parts assumed to be independent of each other as in the Laplacian case, the distribution of a DFT coefficient X_k is then given by:

$$p_M(X_k) = \left(\frac{\sqrt{6}}{8\pi\sigma_x |X_{k(R)}|^{0.5} + |X_{k(I)}|^{0.5}}\right) \times \exp\left\{-\frac{\sqrt{3}(|X_{k(R)}|+|X_{k(I)}|)}{\sqrt{2}\sigma_x}\right\} \quad (7)$$

Applying this equation in two hypotheses H_0 and H_1, which described it above, we have the distributions of the DFT coefficients as follows:

$$p_M(X_k|H_0) = \left(\frac{\sqrt{6}}{8\pi\sqrt{\lambda_{n,k}}|X_{k(R)}|^{0.5}|X_{k(I)}|^{0.5}}\right) \times \exp\left\{-\frac{\sqrt{3}(|X_{k(R)}|+|X_{k(I)}|)}{\sqrt{2\lambda_{n,k}}}\right\} \quad (8)$$

$$p_M(X_k|H_1) = \left(\frac{\sqrt{6}}{8\pi\sqrt{\lambda_{n,k}+\lambda_{x,k}}|X_{k(R)}|^{0.5}|X_{k(I)}|^{0.5}}\right) \times \exp\left\{-\frac{\sqrt{3}(|X_{k(R)}|+|X_{k(I)}|)}{\sqrt{2(\lambda_{n,k}+\lambda_{x,k})}}\right\} \quad (9)$$

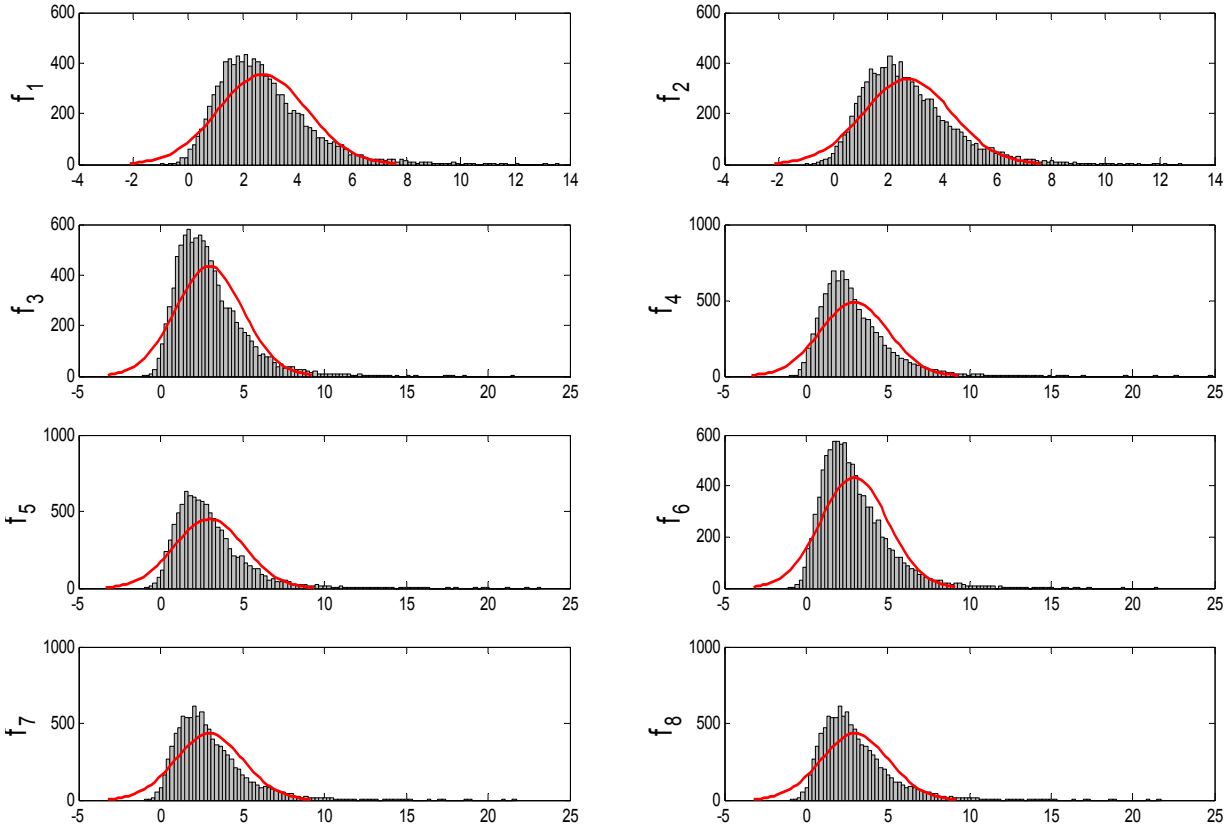


Fig. 3 Histogram of SNR measure in Babble noise environment for each frequency bin f_n

III. MODIFYING STATISTICAL MODEL FOR NONSTATIONARY NOISE

In [4] it has been shown a good performance in high SNR, because the threshold depends only on the background noise statistics. The lower variance in a particular spectral bin requires the lower threshold [26].

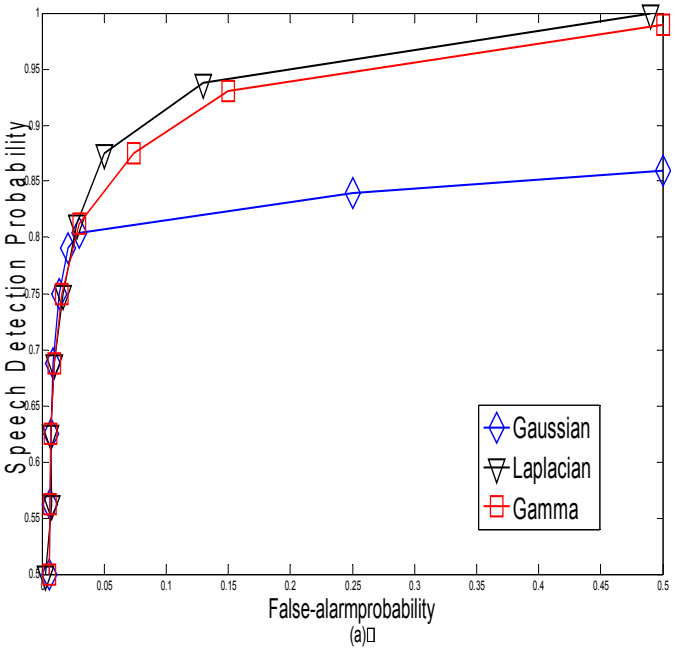
The system performance will be better under the less time variable background noise. However, as the SNR becomes lower, the fundamental assumption in which there will be a significant shift in mean during periods of speech becomes weaker.

The hangover scheme in [4] caused a lower performance and time consumption in non-stationary noise. We present the overall performance of the proposed statistical model-based VAD.

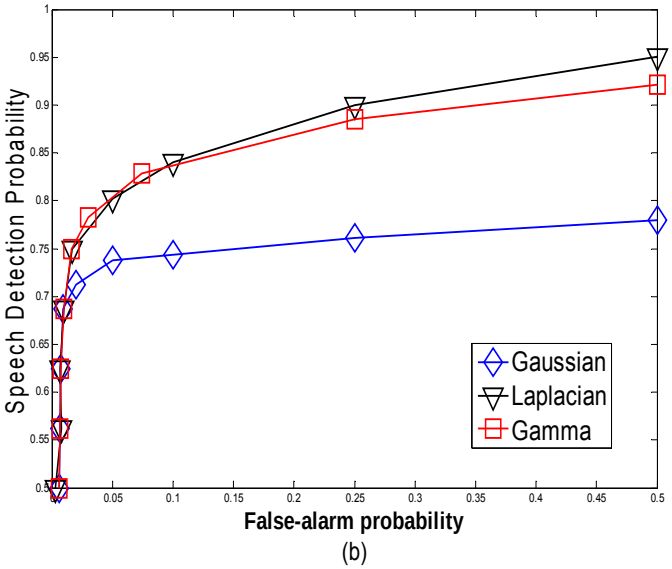
The values of speech detection probability (P_d) for these three models has been shown in Figs. 4-7 where the VAD algorithms were applied to the speech data corrupted by the aforementioned noises at a variety of SNRs (-10, -5, 0 and 5 dB).

The choice of the value of parameters k_L and k_M for Laplacian and Gamma models, respectively - that describe above - in 5dB SNR shows that the best choices for these parameters are k_L=0.9 and k_M=0.9 for Laplacian and Gamma model respectively [4]. From the obtaining results, we could obtain the following observations:

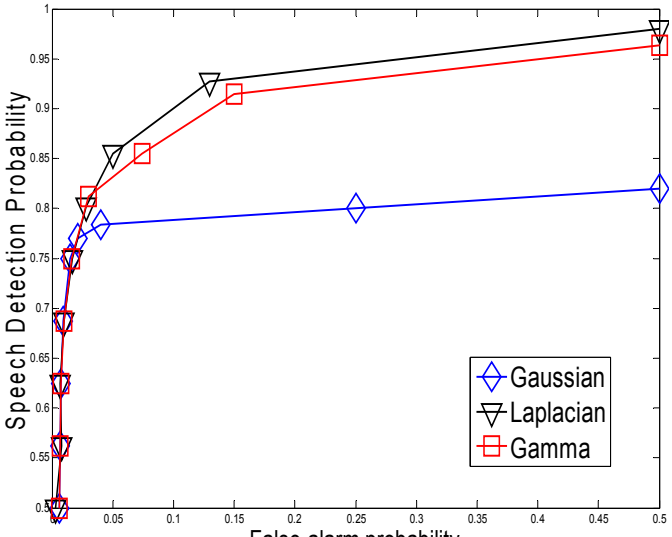
- In the case of the white noise, the Laplacian model-based VAD algorithm outperformed the other approaches. Also, the Gamma model-based resulted in a better performance than that of the Gaussian model in the most tested conditions. For example Fig. 4a shows the speech detection probability in 5dB SNR and we have a good probability for Laplacian and Gamma PDFs.



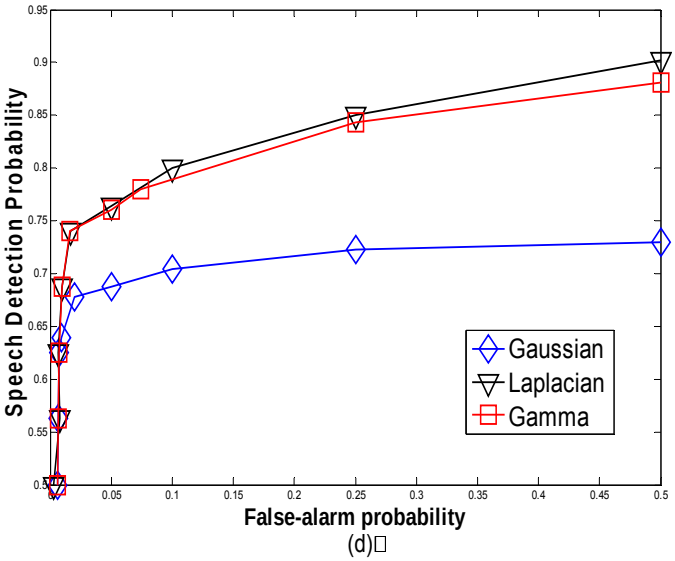
(a)



(b)



(c)

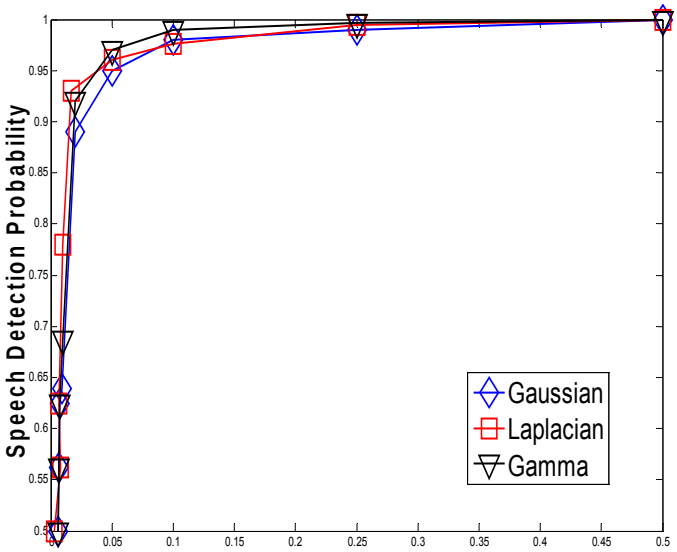


(d)

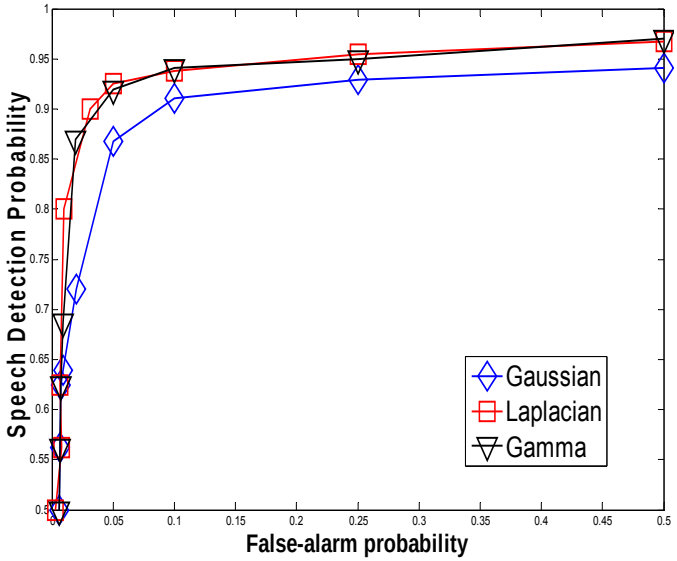
Fig. 4 Speech detection probability in white noise in (a) 5 dB, (b) 0 dB, (c) -5 dB and (d) -10 dB

It is observed in the Figs. 1(a-d), for different SNR values, Laplacian and Gamma PDFs have same probability but Laplacian is better.

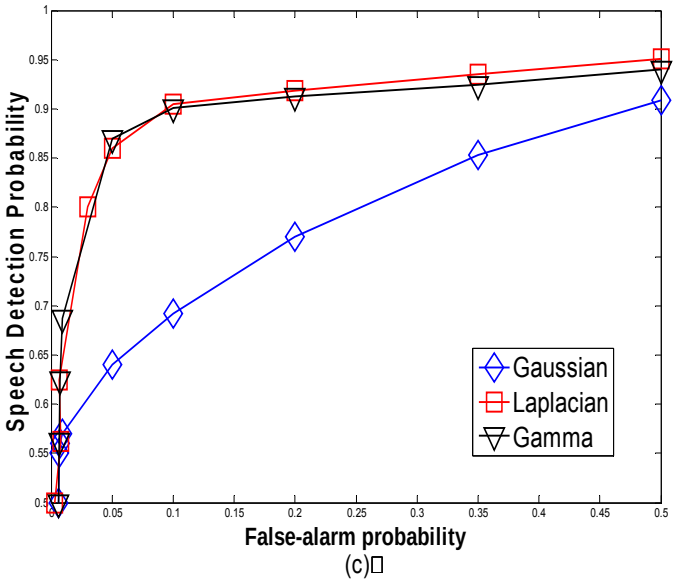
- In contrast, from the P_d's shown in Figs. 2(a-d) looks relatively close to each other in the case of the vehicular noise. However, it is observed that the Gamma model-based VAD algorithm demonstrates a slightly better performance than the other models. In the case of the vehicle noise, the Laplacian model-based VAD algorithm has a same performance with Gamma algorithm. Also, both of these model-based resulted in a better performance than that of the Gaussian model in the most tested conditions. For high SNRs, thus shown in the Fig. 5a and Fig. 5b we have good and same performance in all three PDFs. Fig. 5c and Fig. 5d show the speech detection probability in low SNRs and we have a same probability for Laplacian and Gamma PDFs. But the Gaussian PDF is worst.

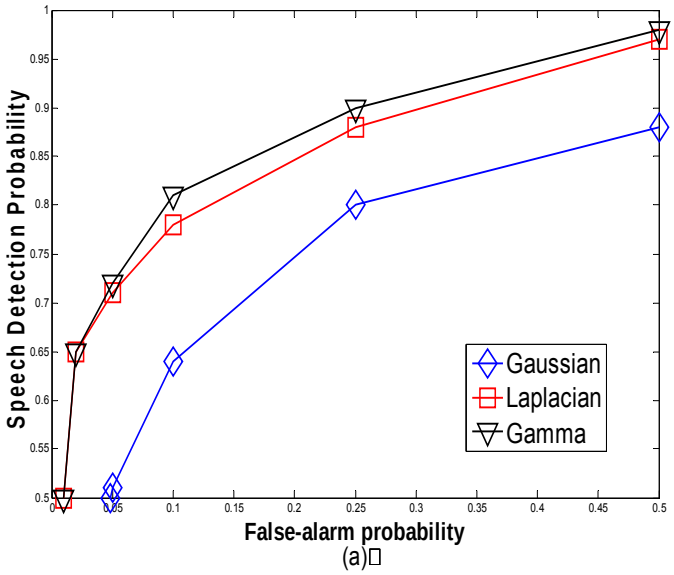


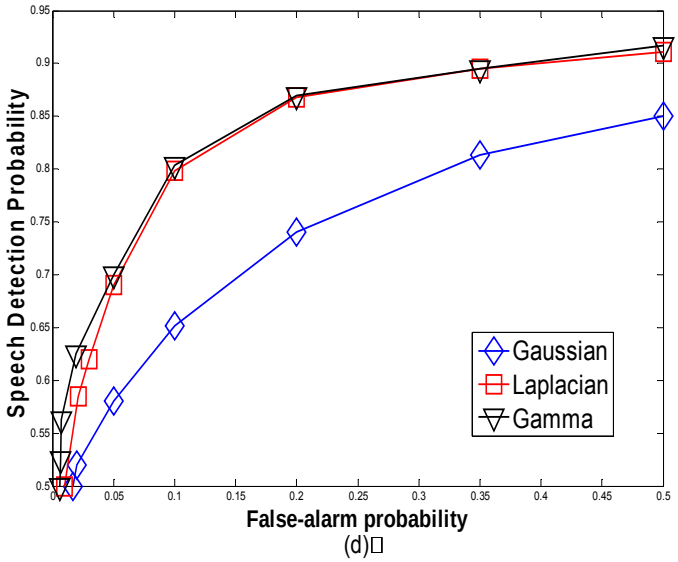
(a)

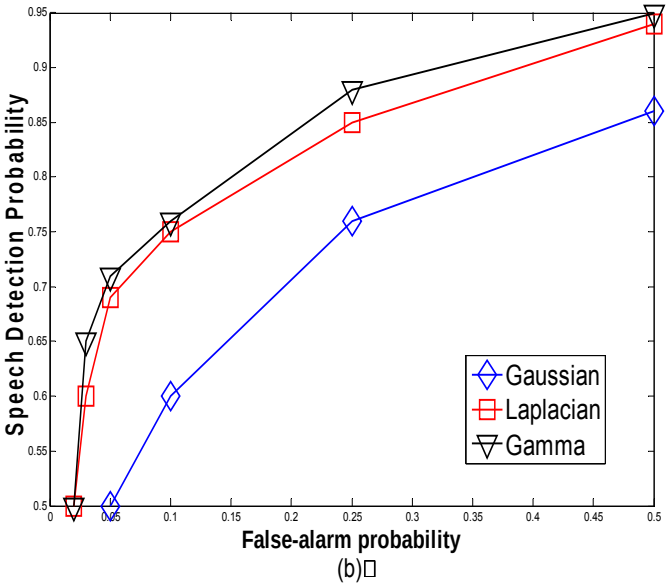


(b)









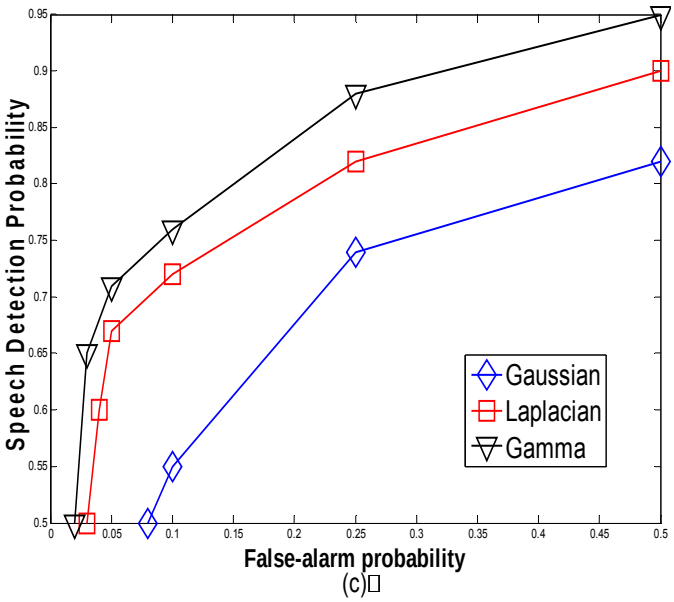
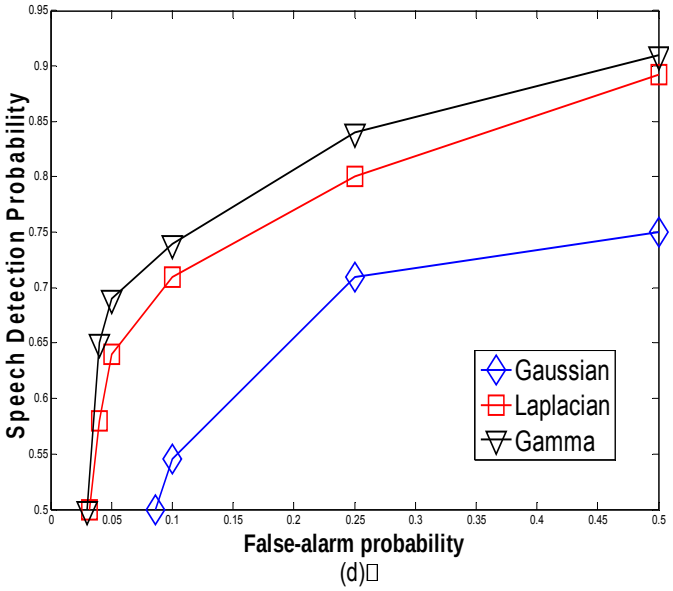


Fig. 5 Speech detection probability in vehicle noise in (a) 5 dB, (b) 0 dB, (c) -5 dB and (d) -10 dB

In the above Figures, it could be resulted in that the performance of all PDFs type in respect of white noise decreased. But in general, it seems good for Laplacian and Gamma model-based VAD.

- As shown in Figs. 6(a-d), for the babble noise, the VAD algorithm incorporating the Laplacian model yielded a performance superior to both the Gamma and Gaussian PDFs. Moreover, the performance difference became larger as the SNR decreased. Also, both of these model-based resulted in a better performance than that of the Gaussian model in the most tested conditions. For high SNRs, as shown in the Fig. 6a and Fig. 6b we have good and same performance in all three PDFs. Fig. 6c and Fig. 6d show the speech detection probability in low SNRs and we have a same probability for Laplacian and Gamma PDFs. But the Gaussian PDF is worst.

Babble noise is one of the hardest noises for separating and detecting the active parts of speech signal. This type of noise



(d)

Fig. 6 Speech detection probability in babble noise in (a) 5 dB, (b) 0 dB, (c) -5 dB and (d) -10 dB

has periodicity properties that make it hard for detection and the simulation is heavy and time consuming process.

In these Figures, we have the worst performance in respect of other noises type and performance decreasing as SNR increased [26].

IV. CORRECTION FACTORS

In this section, we propose a technique to adopt various factors for the Likelihood ratios (LRs) such as $c_k \log \Lambda_k$, which shows below, as we believe that incorporation of the different contributions of the LRs will increase the performance of the VAD [26]:

$$\Lambda_k \equiv \frac{p(Y_k \mid H_1)}{p(Y_k \mid H_0)} = \frac{1}{1+\xi_k} \exp\left\{\frac{\gamma_k \xi_k}{1+\xi_k}\right\} \tag{10}$$

where $\xi_k = \lambda_{x,k} / \lambda_{n,k}$ and $\gamma_k = Y_k / \lambda_{n,k}$ denote the *a priori* signal-to-noise ratio (SNR) and the *a posteriori* SNR, respectively [4]. The *a posteriori* SNR γ_k is estimated using $\lambda_{n,k}$, and the *a priori* SNR is estimated by the well-known Direct Decision (DD) method as follows [5]:

$$\hat{\xi}_k = \alpha \frac{|\hat{X}_k(t-1)|^2}{\lambda_{n,k}(t-1)} + (1+\alpha) P[\gamma_k(t)-1] \tag{11}$$

where $\hat{X}_k(t-1)$ the speech spectral amplitude estimate of the previous frame is obtained using the minimum mean-square error (MMSE) estimator [3]. Also, α is a weight determined in the range (0.95, 0.99) [1]. The function *P[x]=x* if $x \geq 0$ and *P[x]=0* otherwise. The final decision in the conventional statistical model-based VADs has been established from the geometric mean of the LRs computed for the individual frequency bins [7] and is obtained by:

$$\log \Lambda = \frac{1}{M} \sum_{k=1}^{M} \log \Lambda_k(t) \underset{<_{H_2}}{\overset{>^{H_1}}{}} \eta \tag{12}$$

where an input frame is classified as speech presence if the geometric mean of the LRs is greater than a certain threshold value η and speech absent otherwise. The factors c_k, is needed to satisfy the following conditions:

$$\sum_{k=1}^{M} c_k = 1, \quad c_k \geq 0 \tag{13}$$

Let $\Lambda_k = \frac{1}{M} \sum_{k=1}^{M} c_k \log \Lambda_k$ give the threshold value where c_k is:

$$c_k = \frac{\exp(\tilde{c}_k)}{\sum_{i=1}^{M} \exp(\tilde{c}_k)} \tag{14}$$

We therefore adopt the following parameter transformation, which is inversely transformed to c_k :

$$\tilde{c}_k = \log c_k \tag{15}$$

Let denote the set of estimations for the transformed factors at time. Then, it is updated based on the steepest descent algorithms as follows:

$$\tilde{c}_k(t+1) = \tilde{c}_k(t) - \varepsilon \frac{\partial L(t)}{\partial \tilde{c}_k} |_{\tilde{c}_k = \tilde{c}_k(t)} \tag{16}$$

The GPD approach approximates the empirical classification error by a smooth objective function, which is the 0-1 step loss function defined by:

$$L(t) = \begin{cases} \dfrac{1}{1+\exp[-2\gamma(\theta-\Lambda_c)]} & \text{if } H_1 \text{ occured} \\ \dfrac{1}{1+\exp[2\gamma(\theta-\Lambda_c)]} & \text{if } H_0 \text{ occured} \end{cases} \tag{17}$$

where γ denotes the gradient of the sigmoid function. Therefore the reformed system performed by:

$$\Lambda_c = \frac{1}{M} \sum_{k=1}^{M} c_k \log \Lambda_k \underset{<_{H_2}}{\overset{>^{H_1}}{}} \eta \tag{18}$$

V. EXPERIMENTAL RESULTS

In order to evaluate the performance of the proposed algorithm, we added the white, vehicular, and babble noises from the NOISEX-92 database [7] to the clean speech with varying SNR. The VAD test was carried out for each 10ms frame in length.

The parameters used for defining the objective function *L* were selected such that $\gamma = 1$ and the step size for parameter update was set to $\varepsilon = 1-(t)/4000$. In practice, a threshold

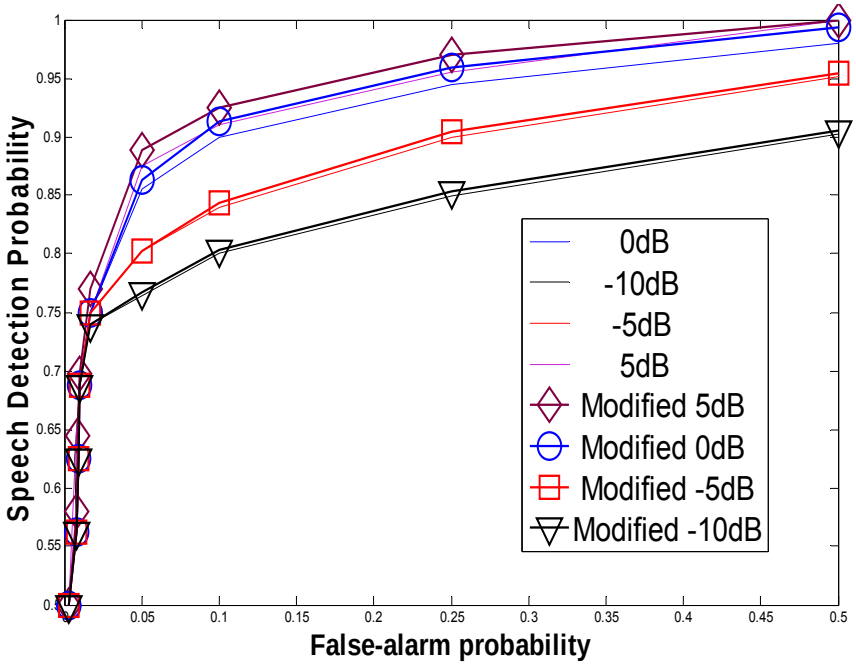


Fig. 7 ROC curves and modified ROC curves for white noise in 5 dB, 0 dB, -5 dB, and -10 dB.
—— the basic form of ROCs
– – – Modified form of ROCs

value of the combined score was set to 0 as the experimentally chosen boundary in the middle of Λ_S stemming from speech and Λ_N stemming from noise.

As a result, Figs. 7, 8 and 9 show the ROC curves for the proposed algorithm compared with the conventional Laplacian methods using white, vehicle and babble noises, respectively, in 5dB, 0dB, -5dB and -10dB.

Finally, among the different sets of the factors, we selected only a single set of the factors as a representative case which is obtained based on an observation that the weights under each training condition seem to be quite similar.

Fig. 7 represents the ROC curves for white noise in 5, 0, -5 and -10dB SNR. In 5dB we have a good modification by

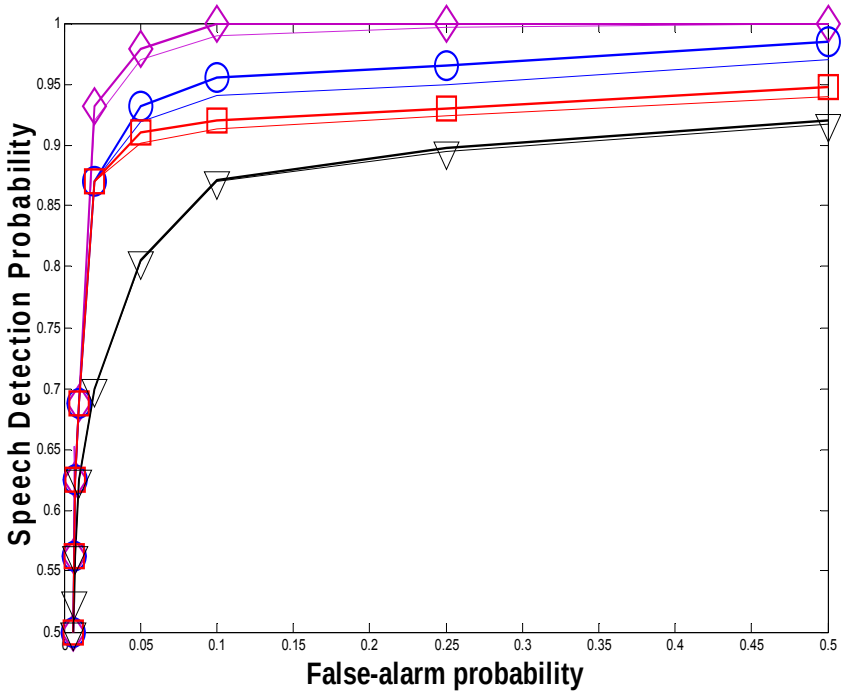


Fig. 8 ROC curves and modified ROC curves for vehicle noise in 5 dB, 0 dB, -5 dB, and -10 dB.
—— the basic form of ROCs
– – – Modified form of ROCs

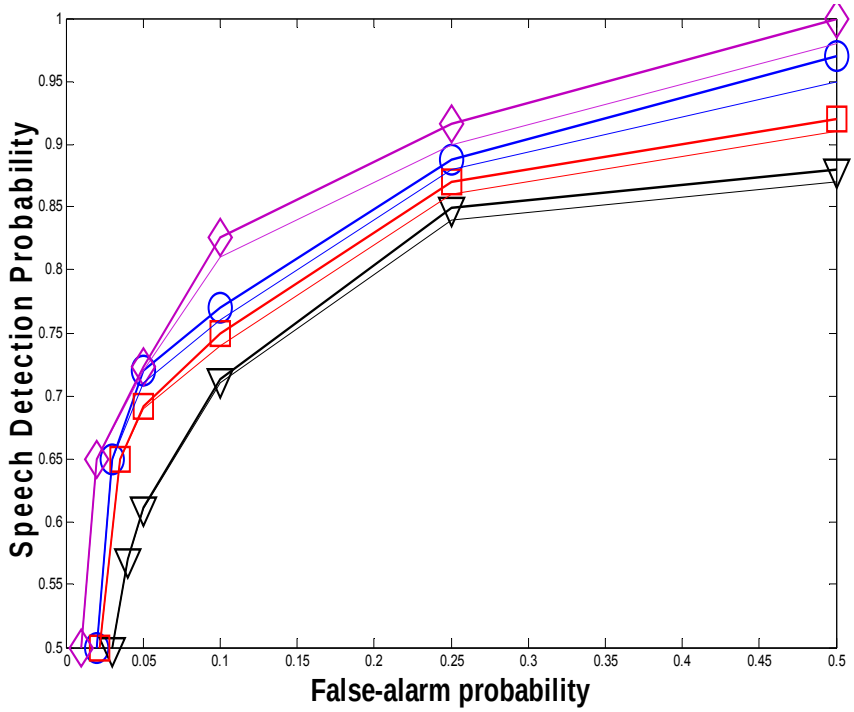


Fig. 9 ROC curves and modified ROC curves for babble noise in 5 dB, 0 dB, -5 dB, and -10 dB.
—— the basic form of ROCs
– – – Modified form of ROCs

means of correction factors and in the results in 0dB show that the modified version of 0dB is better that 5dB before modification and it has a good result. In low SNRs, -5 and -10dB, it can slightly modify the Laplacian PDF. Most types

Fig. 8 represents the ROC curves for vehicle noise in 5, 0, -5 and -10dB SNR.

It is as same as the white noise. Also the total performance of correct detection in this type is generally better than white noise by means of correction factors and the results in 0dB also good for Laplacian-based VAD systems. In low SNRs, -5 and -10dB, it can slightly modify the Laplacian PDF but it's better in -5dB. The modification of speech signal in the vehicle noise seems good and we can use this method for various types of VAD systems.

In all the three noise types, we have an increment in Speech Detection Probability when False-alarm Probability decreased.

Fig. 9 represents the ROC curves for babble noise in 5, 0, -5 and -10dB SNR.

In this case, the overall performance, in respect of white and vehicle noise, decreased. For 5dB and 0dB SNR we have a good modification for 0.1-0.5 False-alarm Probability and slight for low False-alarms. In low SNRs, for this test -5dB and -10dB, we can see some modification in 0.05-0.5 False-alarm for -5dB and 0.2-0.5 for -10dB.

Babble noise is one of the hardest noises for separating and detecting the active parts of speech signal. This type of noise has periodicity properties that make it hard for detection and the simulation is heavy and time consumption.

VI. Conclusion

An extensive study and experiments on the statistical models under Low SNR conditions have made it possible to understand that the complex Laplacian and Gamma PDFs could be strong candidates for a parametric representation of the noisy speech spectra distribution. We select the Laplacian model and trying to improve its PDF. The results show a good modification especially in case of Gaussian and babble noise.

In the future works we want to use the product of correction factors for modifying G.729 Annex B standard. This is the Voice Activity Detector (VAD) part of ITU-T standard. This standard has a lot of benefits but low performance in low SNRs.

References

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

[2] J-H. Chang, N-S. Kim, and S.K.MITRA, "Voice Activity Detection Based on Multiple Statistical Models" IEEE TRANSACTIONS ON SIGNAL PROCESSING, VOL. 54, NO. 6, JUNE 2006.

[3] R.J. Mcaulary, and M.L. Malpass, "Speech enhancement using a soft decision noise suppression filter" *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-28, pp. 137–145, Apr. 1980.

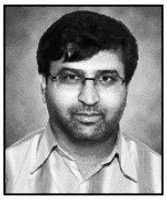
[4] Y. Ephraim, and D. Malah. :'Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator' IEEE Trans. Acoust., Speech, Signal Process., vol. ASSP-32, no. 6, pp. 1190–1121, Dec. 1984.

[5] J. Sohn, and W. Sung, "A voice activity detector employing soft decision based noise spectrum adaptation" in *Proc. Int. Conf. Acoustics, Speech, and Signal Processing*, May 1998, vol. 1, pp. 365–368.

[6] J. Ramirez, J.M. Gorriz, J.C.SEGURA, C.G. PUNTONET, and A. J. RUBIO, "Speech/non-speech discrimination based on contextual information integrated bispectrum LRT" *IEEE Signal Process. Lett,* vol. 13, no. 8, pp. 497–500, Aug. 2006.

[7] A. Varga, and H.J.M. Steeneken, "Assessment for automatic speech recognition, II—NOISEX-92: A database and an experiment to study the effect of additive noise on speech recognition systems" *Speech Commun.*, vol. 12, pp. 247–251, Jul. 1993.

[8] Kun-Ching Wang, and Chiun-Li Chin, "An Efficient Voice Activity Detection in Realistic Environments' WSEAS TRANSACTIONS on SYSTEMS", Issue 7, Volume 6, July 2007.
[9] Jun Li, and Zheng Tang, **"**Enhancing the Human-Computer Interaction on Camera-Equipped Mobile Phones" WSEAS TRANSACTIONS on SYSTEMS, Issue 7, Volume 6, July 2007.
[10] 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.
[11] L. R. Rabiner and M. R. Sambur, "Voiced–unvoiced-silence detection using the Itakura LPC distance measure," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing*, May 1977, pp. 323–326.
[12] J. A. Haigh and J. S. Mason, "Robust voice activity detection using cepstral features," in *IEEE TEN-CON*, 1993, pp. 321–324.
[13] J. D. Hoyt and H. Wechsler, "Detection of human speech in structured noise," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing*, May 1994, pp. 237–240.
[14] R. Tucker, "Voice activity detection using a periodicity measure," *Proc. Inst. Elect. Eng.*, vol. 139, no. 4, pp. 377–380, Aug. 1992.
[15] A. Benyassine, E. Shlomot, and H. 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 supplications," *IEEE Commun. Mag.*, pp. 64–72, Sept. 1997.
[16] F. Beritelli, S. Casale, and A. Cavallaro, "A robust voice activity detector for wireless communications using soft computing," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 9, pp. 1818–1829, Dec. 1998.
[17] F. Beritelli and S. Casale *et al.*, "Performance evaluation and comparison of G.729/AMR/fuzzy voice activity detectors," *IEEE Signal Process. Lett,* vol. 9, no. 3, pp. 85–88, Mar. 2002.
[18] S. G. Tanyer and H. Ozer, "Voice activity detection in nonstationary noise," *IEEE Trans. Speech Audio Process.*, vol. 8, no. 4, pp. 478–482, Jul. 2000.
[19] S. G. Tanyer and H. Ozer, "A geometric algorithm for voice activity detection in nonstationary gaussian noise," in *Proc. EUSIPCO*, Rhodes, Greece, Sep. 1998.
[20] R. Tucker, "Voice activity detection using a periodicity measure," *Proc. IEEE*, vol. 139, pp. 377–380, Aug. 1992.
[21] J. Sohn, N. S. Kim, and W. Sung, "A statistical model-based voice activity detection," *IEEE Signal Process. Lett,* vol. 6, no. 1, pp. 1–3, Jan. 1999.
[22] Y. D. Cho, K. Al-Naimi, and A. Kondoz, "Improved statistical voice activity detection based on a smoothed statistical likelihood ratio," in *Proc. IEEE ICASSP'01*, vol. 2, Salt Lake City, UT, 2001, pp. 737–740.
[23] S. M. Kay, *Fundamentals of Statistical Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1998.
[24] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*, 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, 1999.
[25] B. L. McKinley and G. H. Whipple, "Model based speech pause detection," in *Proc. IEEE ICASSP'97*, vol. 2, 1997, pp. 1179–1182.
[26] H. Farsi, M.A. Mozaffarian, H. Rahmani, "Modifying Voice Activity Detection in Low SNR by correction factors" in *Proc. of 3rd WSEAS conference in communication, signal and system (CISST'09)*, vol. 2, Ningbo City, Zhejiang Wanli UT, 10-12 Jan 2009, pp. 16-20.



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.